UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

FACTORS INFLUENCING ECOLOGICAL DYNAMICS OF THE HUMAN MICROBIOME

A DISSERTATION
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
Degree of
DOCTOR OF PHILOSOPHY

By

DAVID K. JACOBSON
Norman, Oklahoma
2020

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/325633

FACTORS INFLUENCING ECOLOGICAL DYNAMICS OF THE HUMAN MICROBIOME

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF ANTHROPOLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Cecil M. Lewis, Jr., Chair

Dr. Jeffrey Kelly

Dr. Tassie Hirschfeld

Dr. Brian Kemp

Dr. Krithivasan Sankaranarayanan

© Copyright by DAVID K. JACOBSON 2020
All Rights Reserved

Acknowledgements

I would like to thank everyone who provided academic and personal guidance, advice, and support throughout my time in graduate school. This doctoral degree would not be possible without the valuable input and council of a wide range of people. First, I would like to thank my chair, Dr. Cecil Lewis, for his continuing optimism in my abilities, as well as his encouragement in allowing me to try out new ideas/techniques, which took my down paths I never would have reached on my own. I am truly grateful. I would also like to thank Dr. Krithi Sankaranarayanan for his tremendous insights in experimental design, microbiology, and bioinformatics, all of which facilitated my progress in the wet and data analysis. I greatly appreciate the diversity of ideas and suggestions that Dr. Tassie Hirschfeld, Dr. Brian Kemp, and Dr. Jeff Kelly provided whenever we met, as well as in the drafts of this dissertation. I am deeply thankful to Dr. Christina Warinner, Dr. Jiawu Xu, Dr. Stephanie Schnorr, Dr. Tanvi Honap, and Dr. Sharmily Khanam for their assistance in brainstorming different projects and their patience in training me on a variety of molecular biology, microbiology, and analytical protocols/approaches.

I would like to extend a special thank you to Dr. Thérèse Kagoné and her team at Centre Muraz, including Issé Roumba, Bachirou Tinto, Alidou Zongo, and Dr. Amadou Dicko. Their patience with the language barrier and willingness to take time out of their work to assist in my dissertation research was incredibly generous and I cannot thank them enough. Thank you so much to all members of LMAMR (Justin, Rita, Nisha, Allie, Jacob, Sterling, Kristen, Sam, Christine, Abby, Robin, Annie, Lizi, Cara, and everyone else) who kept me sane and dealt with my venting during my time in LMAMR.

I cannot say enough about the support that my fiancée Nicole gave through thick and thin. I really would not be here without her and Stout; their positivity was a much-needed balance to my pessimism. To my parents, Kirk and Cindy, thank you so much for the encouragement and support in undergrad and graduate school and for helping me find my path. Thank you to my sisters, Sarah, Emily, and Molly for keeping me grounded and being understanding. Thank you to all co-authors in the different manuscripts included in this dissertation, as well as forthcoming manuscripts related to my work at LMAMR. Your suggestions greatly strengthened each segment of said manuscripts. Finally, thank you to all individuals who donated samples that were used in my research. I cannot express my gratitude enough and I am forever indebted to your generosity and willingness to participate in research.

Table of Contents

Acknowledgements ---- **iv**

Table of Contents ---- **v**

List of Tables ---- **viii**

List of Figures ---- **ix**

Abstract ---- **xi**

Chapter 1 – Introduction ---- **1**

1.1 MICROBIOME BACKGROUND ---- *1*

1.2 MICROBIOMES IN ANTHROPOLOGY ---- *2*

1.3 MICROBIOME ECOLOGY ---- *4*

1.3 DISSERTATION STRUCTURE ---- *10*

Chapter 2 – Functional Diversity of Microbial Ecologies Estimated from Ancient Human Coprolites and Dental Calculus, ---- **12**

2.1 ABSTRACT ---- *12*

2.2 INTRODUCTION ---- *13*

2.3 METHODS ---- *18*

Archaeological Context of Novel Data ---- 18

Shotgun-sequencing of ancient dental calculus samples ---- 20

Data processing ---- 20

Assessing preservation of ancient microbiome signatures ---- 21

Authenticating ancient DNA ---- 21

Generating taxonomic and functional profiles ---- 22

Network Analysis ---- 22

Functional Redundancy and Response Diversity Analysis ---- 24

Sample size analysis ---- 25

Statistical Tests ---- 25

2.4 RESULTS ---- *25*

Coprolites ---- 25

Dental Calculus ---- 30

Comparison to Modern Microbiomes ---- 35

Sample Size Simulation ---- 37

2.5 DISCUSSION ---- *38*

Chapter 3 – Non-Industrial Gut Microbiomes Provide a More Resilient Ecology for Short-Chain Fatty Acid Production' ---- 46

3.1 ABSTRACT ---- 46

3.2 INTRODUCTION ---- 47

3.3 MATERIALS AND METHODS ---- 51

Study Design ---- 51

Statistical Analysis ---- 53

3.4 RESULTS ---- 56

3.5 DISCUSSION ---- 62

Chapter 4 – Shifts in Gut and Vaginal Microbiomes Associated with Platinum-Free Interval Length in Women with Ovarian Cancer' ---- 69

4.1 ABSTRACT ---- 69

4.2 INTRODUCTION ---- 70

4.3 MATERIALS & METHODS ---- 72

Study Population ---- 72

Sample Collection ---- 74

Laboratory Methods ---- 75

Bioinformatic Methods ---- 75

Statistical Methods ---- 76

4.4 RESULTS ---- 77

Vaginal microbiome ---- 77

Gut microbiome ---- 81

4.5 DISCUSSION ---- 84

4.6 CONCLUSIONS ---- 88

Chapter 5 – Conclusions ---- 90

Bibliography ---- 94

Supplementary Material A ---- 116

Author List and Affiliations ---- 117

Authors' Contributions ---- 117

Microbiome Network Analysis ---- 117

Supplementary Figures A: 1-9 ---- 122

Supplementary Tables A: 1-8 ---- 130

Supplementary Material B ---- 157

Author List and Affiliations----------*157*

Authors' Contributions and Acknowledgements ----------*158*

Author contributions----------158

Acknowledgements:----------158

Supplementary Figures B: 1-3 ----------*159*

Supplementary Tables B: 1-5 ----------*162*

Supplementary Material C ----------187

Author List and Affiliations----------*187*

Authors' Contributions and Acknowledgements ----------*187*

Contributions ----------187

Acknowledgements ----------188

Escherichia Origin ----------*188*

Supplementary Figures C ----------*189*

Supplementary Tables C ----------*193*

List of Tables

Chapter 1

No Tables

Chapter 2

Table 2-1: Network Properties of ancient microbiome ecology datasets. 26
Table 2-2: Keystone taxa identified from ancient microbiome datasets. 27
Table 2-3: Network properties of ancient and modern microbiome networks. 31

Chapter 3

No Tables

Chapter 4

Table 4 - 1: Demographic and clinical data for individuals in this study 73

Chapter 5

No Tables

Supplementary Material A

Supplementary Table A - 1: Archaeological and anatomical context for Nuragic and Maya samples 130
Supplementary Table A - 2: Metagenome samples used in this study that were downloaded from NCBI 132
Supplementary Table A - 3A-D: Top 50 genes in keystone taxa - Rio Zape Coprolites 140
Supplementary Table A - 4A-B: Top 50 genes in keystone taxa - Nuragic dental calculus 146
Supplementary Table A - 5A-C: Top 50 genes in keystone taxa - Maya dental calculus 149
Supplementary Table A - 6A-B: Top 50 genes in keystone taxa - Radcliffe museum dental calculus 153
Supplementary Table A - 7: Network properties change with sample size 155
Supplementary Table A - 8: Keystone identification falters in small sample size. 156

Supplementary Material B

Supplementary Table B - 1: Genera Involved in SCFA Synthesis 162
Supplementary Table B - 2: Samples Used in SCFA Analysis 163
Supplementary Table B - 3: Proportional Contribution to Total SCFA Gene Abundance. 185
Supplementary Table B - 4: Median richness of selected SCFA-producing genera. 186

Supplementary Material C

Supplementary Table C - 1: Full lifestyle metadata for patients involved in this study 193
Supplementary Table C - 2: Raw Escherichia reads from extraction negatives and PCR blanks 194

List of Figures

Chapter 1

No Figures

Chapter 2

Figure 2 - 1: Rio Zape coprolite network. 28
Figure 2 - 2: Functional diversity of the Rio Zape coprolite for SCFA synthesis. 29
Figure 2 - 3: Networks for the ancient dental calculus datasets 31
Figure 2 - 4: Functional diversity in ancient calculus datasets. 34

Chapter 3

Figure 3 - 1: Map of Human Gut Microbiome Metagenomes Analyzed 50
Figure 3 - 2: Relative Abundance of SCFA Genes Compared Between Lifestyle Categories 57
Figure 3 - 3A-C: Taxonomic Diversity of SCFA-encoding Taxa 59
Figure 3 - 4A-F: Hill Numbers for SCFA-encoding Taxa 60
Figure 3 - 5A-C: Proportion of Genes Classified to a Taxon for Each SCFA 62

Chapter 4

Figure 4 - 1: Heatmap of the 25 most abundant genera in the vaginal microbiome 79
Figure 4 - 2A-D: Log-transformed odds ratio vaginal microbiome dominance. 79
Figure 4 - 3: Lactobacillus iners dominates in small gross residual disease. 81
Figure 4 - 4: Unweighted UniFrac distances (PC1 and PC2) of fecal microbiomes. 83
Figure 4 - 5A-B: Phylogenetic diversity in fecal microbiomes. 83

Chapter 5

No Figures

Supplementary Material A

Supplementary Figure A - 1A-D: Rio Zape coprolite DNA damage patterns for keystone taxa 122
Supplementary Figure A - 2: SourceTracker results for novel ancient dental calculus samples 123

Supplementary Figure A - 3A-B: Nuragic dental calculus DNA damage plots for keystone taxa. 124
Supplementary Figure A - 4A-C: Maya dental calculus DNA damage plots for keystone taxa. 125
Supplementary Figure A - 5: Co-occurrence heatmap of oral taxa of interest 126
Supplementary Figure A - 6A-F: Functional diversity in ancient and modern microbiome samples 127
Supplementary Figure A - 7A-B: Small sample size effect network properties 128
Supplementary Figure A - 8: Small sample size hinders identification of keystone taxa 129
Supplementary Figure A - 9A-C: Visual representation of network properties 130

Supplementary Material B

Supplementary Figure B - 1: Gini-Simpson Index Values for Taxa Encoding SCFAs 159
Supplementary Figure B - 2: Proportion of All Genes Classified to A Taxon at Different Phylogenetic Levels 160
Supplementary Figure B - 3: Genus:Species Relative Mapping Index. 161

Supplementary Material C

Supplementary Figure C - 1A-C: Weighted UniFrac beta diversity of all vaginal microbiome samples 189
Supplementary Figure C - 2A-B: Proportional contribution of most abundant phyla and genera in the vaginal microbiome in this study. 189
Supplementary Figure C - 3: Stacked bar chart of the proportion of samples within each study group that were dominated by different taxa 190
Supplementary Figure C - 4: Lactobacillus abundance has a positive association with log-transformed cell density in each sample. 190
Supplementary Figure C - 5A-B: Proportional contribution of most abundant phyla and genera in the gut microbiome in this study. 191
Supplementary Figure C - 6A-C: Weighted UniFrac beta diversity for gut microbiome samples in this study. 191
Supplementary Figure C - 7: Genera at high abundance in fecal outlier group 192
Supplementary Figure C - 8: Prevotella abundance in gut microbiome of ovarian cancer patients. 192
Supplementary Figure C - 9: Relationship between abundance of Escherichia in the vaginal and gut microbiomes. 193

Abstract

Human microbiomes are increasingly seen as a key to understanding human biology, whether it be in studying health/disease or in documenting human diversity. Anthropological interest in human microbiomes has primarily focused on inventorying shifts microbial abundance between lifestyles and over time. While undoubtedly valuable information, these taxonomic inventories lack application of theoretical frameworks that can provide a deeper understanding of human-microbiome interactions. In my work, I integrate microbiome data with ideas typically used in the study of macroecological systems. I present a background on ecological frameworks and their general applicability to human microbiome studies in Chapter 1. In Chapters 2-4, I present unique research on human microbiome ecology. Chapter 2 is a first-of-its-kind study on ecological dynamics in ancient human microbiomes, in which I demonstrate how network analysis can inform on general microbiome community structure in ancient coprolites and dental calculus, as well as demonstrate overlap in key ecological signatures between ancient and contemporary non-industrial populations. Chapter 3 is a demonstration of the value of interrogating potential microbiome stability and resilience in a specific niche, namely short-chain fatty acid (SCFA) production. Our research indicates that non-industrial gut microbiomes are more resilient for SCFA production and industrial SCFAs are encoded by a few, closely related species. These results come in the face of substantial database bias that inhibits study of non-industrial gut microbiomes. Finally, Chapter 4 describes vaginal and gut microbiomes in women with Ovarian Cancer (OC). We determine that women with OC have lower than expected vaginal *Lactobacillii* dominance, which creates ecological space for an opportunistic bacteria such as *Escherichia*, to thrive in women who suffer quick recurrence of cancerous growth after platinum chemotherapy.

Dissertation Keywords: Human microbiome, ecology, resilience, biological anthropology

Chapter 1 – Introduction

1.1 MICROBIOME BACKGROUND

Human health and disease are largely influenced by complex interactions among underlying human genetic architecture, personal habits/exposures/socioeconomics, and the microbes living in and on the human body. The resident microbes, referred to as the human microbiome, are increasingly seen as integral to human biology, where their influence ranges from impacting efficiency of energy harvest from diet, maintenance of vitamin homeostasis, to fending off pathogens (1, 2). Human gut microbiomes consist of hundreds to thousands of microbial species from a diverse range of bacterial phyla and many studies have focused on inventorying the presence/abundance of bacteria in different health/disease states, as well as between lifestyles (3-7). While various studies have identified specific microbes that are strongly influenced by lifestyle, such as diet, or found bacteria that are associated with disease, there is an increasing awareness that the metabolic capabilities of the human microbiome paint a more accurate picture of this ecosystem (8, 9). In other words, understanding the microbial aspect of human biology needs more information than just an inventory of what microbes are present, it requires knowledge of these microbes' metabolic capabilities (functional analysis).

Advances in next-generation DNA sequencing have facilitated great insights into the functional capabilities of the human microbiome via metagenomics and transcriptomics (2) by providing reams of data, resulting in ever-growing databases of annotations linking DNA sequences to proteins and metabolic processes. Yet the technological advances have not been matched by advances in theory. Similar to taxonomic inventories, functional inventories provide a useful overview of what is present in a microbiome, but it masks lower-level dynamics. There is

extensive room to address this theoretical gap through incorporation of frameworks from macroecology; namely, ecosystem resilience, stability in functional potential, and identification of keystone taxa that perform vital roles in the microbial system.

1.2 MICROBIOMES IN ANTHROPOLOGY

Ecological microbiome analysis is assuredly making inroads in the field (10-13) but there has been little incorporation of ecological frameworks into anthropological microbiome research. Anthropology has an important seat at the table in microbiome research because it is anthropologists' focus on human evolution and lifestyle diversity that have yielded a variety of biologically important findings. There is evidence that microbiome communities, and specific lineages of bacteria in particular, have co-speciated with their animal hosts (14, 15). This deep evolutionary relationship demonstrates the extent to which human biology and human microbiomes are intertwined because the depletion of bacteria that co-speciated with humans in evolutionary history is linked to gastrointestinal inflammation (16). Anthropology also plays a key role in the field by describing how lsubsistance practices have changed the human microbiome through the study of both archaeological specimens and contemporary non-industrial populations (17, 18). The non-industrial populations studied in microbiome research are quite diverse – including South American and African Hunter-Gatherers, as well as rural agriculturalists and pastoralists from various parts of the world – but they share a common characteristic of little to no processed food consumption (17, 19-21). Consistently, gut microbiome samples from industrialized populations (think Europe, North America, East Asia) demonstrate lower richness, than non-industrial populations and the absence, or great reduction in abundance, of extirpated bacteria, namely: *Treponema, Catenibacterium, Prevotella,* and *Succinovibrio* (17, 19-21). While the exact role of the extirpated bacteria in the gut microbiome

remains to be elucidated, they are hypothesized to aid digestion in fiber-rich diets (22). The loss in richness observed in industrial populations, which has been tied to various autoinflammatory and metabolic diseases (23), is commonly linked to differences in diet between industrial and non-industrial groups (17, 19, 20, 24); however, living environment (25, 26) and consumption of pharmaceutical drugs (27-30) also contribute to differences in microbiomes between industrial and non-industrial populations.

Anthropological human microbiome research's focus on documenting extirpated bacteria, tracking community richness, and characterizing changes in taxonomic abundance between populations leading different lifestyles is undoubtedly important and built the foundations for future microbiome research. Listing the bacterial taxa and the functions that they genetically encode is useful because it provides a baseline to compare samples; however, there is a growing recognition that these descriptions do not accurately reflect how the community is behaving. Ultimately, the lack of studies on how ecological dynamics intrinsic to the microbiome are impacted by human lifestyle has led to an ignorance of how the microbial community is actually behaving and interacting in industrial and non-industrial populations. The absence of this more ecological approach is unfortunate because these data are important for understanding diversity in human health and biology. For example, bacteria have different growth rates, transcriptional activity, predator/prey interactions, and interphylum signaling that all influence how they interact with each other, as well as their influence on human biology (8, 31). Moreover, redundancy in the diversity and number of bacteria that encode the same gene may impact functional stability in a microbial community over time (8). Yet the scientific community does not adequately know how ecological dynamics within non-industrial gut microbiomes compare to those observed

within industrial gut microbiomes. Clearly, the next step for anthropologists interested in human microbiomes is incorporation macroecology frameworks to better understand the extent of human microbiome diversity.

1.3 MICROBIOME ECOLOGY

Incorporating macroecological theory into microbiome research provides the opportunity to advance our understanding of human microbiome variation and behavior in health and disease. Although microbiomes do not always function in the same manner of macroecosystems (32), there are sufficient similarities in community attributes that macroecological theory can, and should, be applied to microbiomes (33, 34). For example, researchers use the same 'currency' in their analysis: identifying and counting individuals, such as species or genes (not without challenges in either field of study). Similarly, as a main component of their research, ecologists describe how species are distributed throughout a defined area and what factors (density, resource, human-perturbation) influence these species-area distributions, which is fundamentally the same as how microbiome researchers study how microbiomes vary within and between populations (34).

Applying ecological theory to the microbiome has only recently gained traction. While there is no unified or singular approach to study human microbial ecology, the complex ecological processes and interactions that shape community structure include concepts such as functional redundancy, community assembly, succession, response to disturbance, restoration, response diversity, keystone taxa and genes, and co-occurring networks of bacteria (11, 31, 35-39). Each of these concepts leads to estimating ecological variables, providing a more nuanced view of human microbiomes, including how they are formed, what factors drive change, and which

bacteria play important roles in maintaining homeostasis or stable conditions. It is important to reiterate that the field is currently far from a unified theory to study microbial ecology.

Much ecological work on the human microbiomes has focused on resilience. As the field of microbiome ecology is still developing, the term resilience is used with a surprising amount of variation in meaning (40) but I take a broader view of resilience as the ability for a community to experience an outside disruption but still maintain its functional capabilities (9, 13). Ecosystem resilience is seen as vital for microbiome health because it provides a buffer against the regular insults that microbiomes face, such as the consumption of pharmaceutical drugs and exposure to pathogens (13, 31). In the face of a disruption, resilient communities may initially suffer a shift in the structure and/or function of the community but then revert back to the original state, while non-resilient communities undergo an ecosystem state change and the functional capabilities of the microbiome change. As a common example, antibiotic consumption triggers microbiome ecosystem state changes through the overall loss of taxonomic diversity and functional profile in oral and gut microbiome communities (41-43). In oral microbiomes, the loss of diversity may allow orange or red complex bacteria (which are bacteria associated with the progression of periodontitis) to thrive, while in gut microbiomes, a loss of diversity can allow for pro-inflammatory bacteria to proliferate and negatively impact host physiology (44, 45).

An extreme example of an ecosystem state change in the gut microbiome is when a non-resilient microbiome shifts to a *Clostridium difficile* dominated community after the consumption of antibiotics in a hospital/clinical setting, which can cause life-threatening inflammation in the colon (46-48). *C. difficile* is found in 5-15% of healthy adults (49, 50) and evidence suggests that

there are microbial subnetworks of commensal bacteria (*Lachnospiraceae, Barnesiella, Akkermansia*) that influence *C. difficile* colonization and infection progression (47). Specific conditions are necessary for *C. difficile* to proliferate and antibiotics are strongly implicated in facilitating these conditions by altering the metabolic profile of the gut microbiome, resulting in a profile that favors germination of *C. difficile* spores via changes in bile acid production and altered competition for nutrients like sialic acid and glucose (51). A resilient microbiome may suffer an initial shift but then recover to prevent *C. difficile* expansion by maintaining homeostatic production of bile acids and resource competition. (52). Similarly, oral microbiome research on the red complex bacteria, which had been strongly associated with periodontitis, indicates that these bacteria are found in non-diseased individuals (53, 54) and it is only certain ecological conditions that these bacteria shift the oral microbiome into a diseased state. Importantly, resilience should not be only thought about in the context of maintaining a healthy state, as microbiomes may also be resilient to change when they are in a diseased state. Understanding the range within which a microbiome can vary while maintaining its functional capabilities is an important when characterizing a microbiome.

One approach for assessing microbiome resilience is through functional redundancy. Functional redundancy derives from the fact that for any given genetic pathway encoded by one bacterium, there are likely numerous bacterial taxa encoding that same function (8). This concept indicates that a simple description of taxa abundance is too reductionist of an approach (8). Indeed, changes in taxonomic abundance may not have any meaningful impact on the community's functional profile because the functions performed by one taxon may be performed by a suite of others. This produces a few dilemmas for researchers: if multiple species encode the same

function, and that holds true across the suite of genes encoded by the entire microbiome, does the taxonomic make-up of the microbiome matter at all? It is easy to envision a scenario where a group of bacteria are eliminated from a microbial community but the functional profile of the system does not change because other bacteria, that are still present in the community, are able to perform the same functions as the bacteria that were lost. This is of upmost interest to anthropologists due to the numerous bacterial taxa absent from industrialized populations compared to those living a more traditional lifestyle (17, 24). It has been assumed that taxonomic variation between industrial and non-industrial microbiomes results in highly different microbiome communities, yet the truth of how decreased diversity and the loss of extirpated observed in industrial gut microbiomes likely lies in assessing functional redundancy in microbiomes from diverse populations.

Response diversity is another useful tool to measure resilience that can be adapted from macroecology (55, 56) and it is closely related to functional redundancy. In functional redundancy, researchers may consider only the number of bacteria that encode a function as a measure of resilience, whereas response diversity describes the phylogenetic diversity of taxa performing the same function (55, 56). Response diversity captures the idea that a metabolic function that is encoded by a diverse range of bacteria is more likely to be resilient than a function encoded only by a few, closely related bacteria. For example, antibiotics often kill off related groups of bacteria. In a microbiome where these bacteria are the only taxa that perform a certain function, then that function is eliminated from the ecosystem. However, if a diverse group of bacteria encode that function, the loss of one clade of bacteria does not ultimately change the functional capabilities of the microbial community. Therefore, high response

diversity provides resilience (55, 56). Additionally, this example highlights the nuance between functional redundancy and response diversity, as functional redundancy does not describe the phylogenetic and genomic diversity of bacteria encoding a function, which may ultimately be meaningful for the community's resilience.

A further area of interest in microbiome ecology is through community assembly and succession in the microbial community. Bacteria are consistently being introduced to human niches through a variety of sources, such as air/environment, food stuffs, or other microbiomes near-by (metacommunities) (57-59). Many of these bacteria are transient, and there are a variety of factors that drive which bacteria ultimately become permanent residents in a microbiome. Initial assembly of a community requires early colonizing bacteria that often have physiological or metabolic traits that permit colonization by other bacteria, forming a system of community assembly and eventually succession of bacteria in the microbiome. Microbiome assembly and succession partially depends on abiotic elements of the environment that promote colonization of specific groups of bacteria. To provide an obvious example, assembly and succession of the fecal and oral microbiome clearly are shaped by very different events in their respective habitats, which include the microclimate of the respective bioreactor (gut and oral), such as acidity, the available substrates for metabolism, and interaction with other human biological systems. The early make-up of the human GI tract microbiome is influenced by the host, like mode of birth and nutrients in breast-milk/infant formula, to shape the taxonomic and functional make-up of the infant GI tract microbiome (11, 60, 61). The initial assembly of the GI microbiome then influences how microbial succession occurs in early childhood (11). In dental biofilms, community assembly is understood through the lens of early and late colonizing bacteria (62,

63). Early colonizers, such as some *Streptococcus* and *Actinomyces* species, have cell surface proteins that permit binding to the acquired enamel pellicle that forms on the tooth surface (64). The make-up of secondary and late colonizers to the plaque biofilms depends on binding compatibility between microbes, functional competition, and oxygen exposure, amongst other biotic and abiotic variables (62). The colonization of the 'red' complex bacteria, once thought to be associated with periodontal disease, depends on the presence of 'orange' complex bacteria, which may be thought of as bridging microbes that provide the ecological links between early colonizers and the more 'pathogenic' bacteria (37, 54, 65). Taken together, a fuller understanding of microbiome assembly and succession can inform about how the microbiome community was formed as well as identify key interactions that may drive the microbiome structure towards a certain community structure.

A final example of where ecology can be applied to microbiomes is through the identification of keystone taxa. In microbiomes, keystone taxa are bacteria that play important roles in ecosystem function because they are at the center of microbial interactions throughout the microbiome, or at least within different functional groups (38). Keystone taxa may be high or low abundance, but their signature feature is that they are at the heart of microbial interactions, which is often related to their ability to perform a narrow, specific function that multiple bacteria either feed into or rely on. Additionally, keystone taxa need not be keystones for health. For example, *Porphyromonas gingivalis* (a red complex bacterium) has been suggested as a keystone taxon for periodontitis because it alters immune system defense mechanisms and thus promotes a resilient periodontitis-inducing biofilm (54, 66, 67). Identification of keystone bacterium can provide

useful insights about microbial community structure and may cause major shifts in ecosystem function and interactions if they were eliminated from the community.

1.3 DISSERTATION STRUCTURE

This dissertation encompasses three studies (chapters 2 - 4) that use varying approaches to broaden the scientific community's understanding of human microbiome diversity and Chapter 5 is a conclusion/discussion section. Chapters 2 and 3 take an ecological approach that demonstrate that value and benefit of adapting macroecological theories to anthropological microbiome research, while Chapter 4 adopts a disease-marker approach of the human microbiome but still retains ecological interpretations.

Chapter 2 looks at metagenomic data from ancient dental calculus (Maya, 170 BCE-885 CE, Belize; Radcliffe Museum Dataset, 1770-1855 CE, UK; Nuragic, 1400-850 BCE, Sardinia, Italy) and coprolites (Rio Zape, 700 CE, Mexico). Ecological approaches have never been applied to ancient human microbiomes and in a first of its kind study, I applied network analysis to identify keystone taxa and evaluate community structure; in addition, I used functional analysis to assess functional redundancy and response diversity for specific genes of interest in the oral and gut microbiome. Chapter 2 also includes comparison of the ancient microbiome datasets to modern oral and gut microbiomes to provide context for our ecological analysis and draw comparisons between ancient and contemporary non-industrial gut microbiomes.

Chapter 3 incorporates human gut microbiome data from samples collected as part of a partnership with Centre Muraz Research Institute in Burkina Faso, where I worked with Dr. Thérèse Kagoné and her team in the summer of 2017. The Burkina Faso samples were compared

to a global gut metagenome panel of industrial and non-industrial populations to determine resilience in Short-Chain Fatty Acid (SCFA) production. SCFAs are a key molecule involved in human-microbiome interaction and are often positively associated with health. In Chapter 3, I demonstrate that non-industrial gut microbiomes have high resilience, as measured through response diversity, for SCFA production, which is to be expected; however, there is substantial database bias that hinders analysis of gut microbiomes from non-industrial contexts.

Chapter 4 is the result of a collaboration with Dr. Kathleen Moore and her team at the Stephenson Cancer Center at OU Health Sciences Center. In this study, vaginal and fecal samples were collected from women with ovarian cancer with differential responses to platinum-based chemotherapies. There are extensive links between colorectal cancer and the human microbiome but little is known about the relationship between ovarian cancer, response to chemotherapy, and the microbiome. I identified significantly lower than expected dominance of *Lactobacillus* in women with ovarian cancer, compared to similarly aged healthy women. Additionally, women who have a quick recurrence of cancerous growth after chemotherapy were more likely to have a vaginal community dominated by *Escherichia*. These two results indicate that ovarian cancer may cause an ecological disruption in the vaginal microbiome that decreases *Lactobacillus* abundance and permits proliferation of *Escherichia* in women with quick recurrence of cancerous growth.

Together, this dissertation shows the value of incorporating ecological theory into anthropologically focused human microbiome research and contributes to the growing body of literature about ecological dynamics of the human microbiome.

Chapter 2 – Functional Diversity of Microbial Ecologies Estimated from Ancient Human Coprolites and Dental Calculus[1,2]

2.1 ABSTRACT

Human microbiome studies are increasingly incorporating macroecological approaches, such as community assembly, network analysis, and functional redundancy to more fully characterize the microbiome. Such analyses have not been applied to ancient human microbiomes, preventing insights into human microbiome evolution. We address this issue by analyzing published ancient microbiome datasets: coprolites from Rio Zape (n = 7; 700 CE Mexico) and historic dental calculus (n = 44; 1770-1855 CE, UK), as well as two novel dental calculus datasets: Maya (n = 7; 170 BCE-885 CE, Belize) and Nuragic Sardinians (n = 11; 1400-850 BCE, Italy). Periodontitis-associated bacteria (*Treponema denticola*, *Fusobacterium nucleatum,* and *Eubacterium saphenum*) were identified as keystone taxa in the dental calculus datasets. Coprolite keystone taxa included known short-chain fatty acid producers (*Eubacterium biforme, Phascolarctobacterium succinatutens*) and potentially disease-associated bacteria (*Escherichia*, *Brachyspira)*. Overlap in ecological profiles between ancient and modern microbiomes was indicated by similarity in functional response diversity profiles between contemporary hunter-gatherers and ancient coprolites, as well as parallels between ancient Maya, historic UK, and modern Spanish dental calculus; however, the ancient Nuragic dental calculus shows a distinct ecological structure. We detected key ecological signatures from ancient microbiome data, paving the way to expand understanding of human microbiome evolution.

[1] Adapted from Jacobson et al. in press. Functional Diversity of Microbial Ecologies Estimated from Ancient Human Coprolites and Dental Calculus. *Philosophical Transactions of the Royal Society B.*
[2] See Supplementary Material A for full list of co-authors

2.2 INTRODUCTION

Host-associated microbiomes are complex ecosystems with diverse sets of interactions between microbes, the host, and abiotic features. Human microbiome research has primarily focused on documenting the genes/organisms present in a sample and differentiating microbiome communities using presence/absence and relative abundance data (4, 6, 7, 60, 68). Such contributions have undoubtedly advanced the understanding of human biology; however, a stronger focus on taxonomic co-occurrence, identification of taxa with disproportionate influence on community function, as well as overall resilience of metabolic pathways will provide a more nuanced view of the microbiome. An analogy can be drawn from mammalian ecology in the United States' Yellowstone National Park, where a focus on the role that wolves play as a keystone species yields greater clarification on cross-species interactions and dependencies. Wolves are not an abundant species in the ecosystem, yet their predator-prey relationships have tremendous downstream impacts on ecosystem production and stability (69-71). A simple taxonomic inventory does not present the full picture of the wolves' impact on the ecosystem, but an approach focused on their network of interactions demonstrates how they function as a keystone species that reshapes resource allocation and alters interspecies relationships throughout the ecosystem (69-71). In the absence of deeper modeling, wolves would remain a rare biome variant, without a sophisticated understanding of their role as a keystone species.

Human microbiome research can clearly benefit from a similar approach but the momentum in applying ecological theory to the microbiome has only recently gained traction. In this vein, microbiome focus is slowly shifting from describing what is present in a microbiome to understanding the factors that drive community membership, polyspecies interactions, functional

variation, and ecosystem stability. While there is no unified or singular approach, ecological concepts such as community assembly, succession, response to disturbance, restoration, functional redundancy, response diversity, keystone taxa and genes, and co-occurring networks have made inroads in human microbiome research (8, 11, 31, 36-39, 57-59, 72). Each of these concepts lead to estimating ecological variables, providing a more nuanced view of human microbiomes, including how they are formed, what factors drive change, and which bacteria play important roles in maintaining homeostasis or stable conditions. It may appear that such a heavy focus on microbial ecology takes the focus away from human biology; however, there is growing support for the holobiont paradigm. Holobiontism posits that macro-organismal development, health, and general function relies on microorganisms, and therefore, microbes play a role in macro-organism ecology and evolution (73-76); in other words, human-associated microbial ecology and human biology are inextricably intertwined.

The approaches for estimating and characterizing human microbiomes through ecological concepts are in an early stage of research but there have been numerous valuable insights. Community assembly- and succession-focused research has demonstrated that early life human gut microbiome composition is dynamic and strongly influenced by a variety of factors, including birth mode, nutrition, and exposure to antibiotics (10, 11, 60, 77), which can lead to downstream health effects (78). Network analysis has been used to identify potential therapeutic avenues for *Clostridium difficile* infections in the gut (47) and evaluate gut microbiome structure and stability (72). Taxonomic composition fluctuates over time in gut microbiomes (79, 80) but evidence suggests there is stability in metabolic activity (81) that may be driven by functional redundancy (8, 11, 77). Importantly, one should not consider resilience (and stability) as

resilience to remain in a healthy state; ecosystems, and thus microbiomes, can be resilient to change when they are in an alternative state (82-85). Similarly, keystone taxa need not be important for promoting a healthy state. For example, *Porphyromonas gingivalis* has been suggested as a keystone taxon for periodontitis because it alters immune system defense mechanisms and thus promotes a resilient periodontitis-inducing biofilm (54, 66, 67).

Stronger focus on ecological functions and interactions paints a more detailed picture of the role that taxa and functions play in different microbiome states. A logical next step is to apply these approaches to archaeological and paleogenomic microbiome data. These data are in a most unique position to impact the ecological understanding of the human microbiome as they permit exploration of how human microbiomes have responded to major changes in the human condition, such as epidemiological transitions, colonialism, biogeographic range expansions, and industrialization (86-94). In fact, the popularized roles the microbiome plays in human biology are deeply connected to "diseases of civilization", such as allergies, obesity, chronic inflammation, emerging infectious diseases, and the evolution of antibiotic resistance (3, 5, 6, 95-97). To understand the mechanisms behind these changes, we must know exactly what has changed in functional redundancy, keystone taxa, resilience, and assembly of human microbiomes. While datasets from non-human primates and extant non-industrialized people provide some progress towards that goal, there is simply no line more intuitive to understanding ancestral microbiomes than to study ancient populations.

Ecologically focused microbiome research with ancient biomaterials (primarily coprolites and dental calculus) will present unique challenges, such as DNA degradation, small sample size,

contamination, and lack of time-series data. However, coprolites (*i.e.* desiccated feces) and dental calculus (*i.e.* calcified dental plaque) have a long history of providing important information on human health and practices of the past and, in ideal conditions, preserve a record of the human microbiome (86, 88, 98-101). The first ancient microbiome study to apply the next-generation DNA sequencing technology was largely centered around the premise of whether detailed taxonomic information from ancient human gut microbiome was retrievable, and if so, whether these resembled the contemporary human gut (102). From an assemblage of pre-colonial coprolites from Mexico (Rio Zape), they observed a similar taxonomic profile to contemporary gastrointestinal (GI) tract microbiomes at the phylum level, as well concordance with contemporary non-industrialized populations at the genus level due to presence of *Treponema* and *Prevotella*, both of which are nearly absent from gut microbiomes of industrialized populations (102). A follow-up study (89) noted that the Rio Zape assemblage may be a rare find because coprolites from other archaeological sites, including coprolites directly extracted from well-preserved mummies, had very poor gut microbiome preservation, including a taxonomic profile that is not expected from any mammalian gut, let alone a human gut. Additionally, the Rio Zape coprolites are unique because of the relatively high number (n = 8) of samples with human GI microbiome signatures as compared to those from other archaeological sites (103).

Dental calculus has proved to be more reliable in reconstructing an accurate microbiome signature compared to coprolites (90, 91, 104) primarily because mineralization during life makes calculus a sturdy and rigid material lacking in organic nutrients (62, 63). Thus, dental calculus is more resistant to environmental contamination (62, 63). Often, more than 90% of the bacterial DNA found in dental calculus originates from known oral bacteria, whereas less than

half of coprolite bacterial DNA originates from known gut microbes (89, 104), and many of the challenges associated with studying coprolite microbiomes (including low DNA yields, soil contamination, and lack of a true human microbiome community) are less severe in ancient dental calculus. The first next-generation sequencing study of ancient dental calculus demonstrated that the oral microbiome could be reconstructed by amplifying the 16S rRNA gene from samples ranging from 5,500 BCE – 1600 CE (105); however, the use of 16S rRNA variable regions has been shown to be problematic for ancient microbiome datasets due to primer bias (106). Shotgun metagenomic approaches face fewer biases for taxonomic identification and additionally allow for the reconstruction of genomes and functional characterization (93). Along this line, metagenomics has been used to reconstruct genomes from *Tannerella forsythia* (93) and *Methanobrevibacter oralis* (94) as well as track diversity in functional and taxonomic profiles in the mammalian oral microbiome over time (91, 105, 107).

Here, we present an ecologically focused analysis on previously published Rio Zape human coprolites (n = 8, 700 CE) (89, 103) and historical dental calculus samples from the Radcliffe Infirmary Burial Ground, United Kingdom (n = 44, 1770-1855 CE) (107), as well as novel metagenomic dental calculus data generated as part of this study (see Methods) from Maya individuals from Belize (n = 7, 170 BCE-885 CE) and Nuragic individuals from Sardinia, Italy (n = 11, 1400-850 BCE). To best adapt ecological approaches to ancient coprolites and dental calculus, we focused on analyzing the structure and properties of microbiome networks, identification of keystone taxa, and functional diversity of specific functions of interest. Each of these can be evaluated without time-series data. Compositionally corrected networks using SparCC (108) were generated following the protocol suggested by Layeghifard et al. (109) using

species-level bacterial taxonomic inventories from MetaPhlAn2 (110). Each network was generated 100 times to estimate network properties (number of clusters, modularity, transitivity, and articulation points). Modularity and transitivity values were categorized as very low, low, medium, high, and very high based on the distribution modularity and transitivity values across the different networks we generated. Keystone taxa were identified using three techniques common to network evaluation: page rank (111, 112), hubs (113, 114), and closeness centrality (115, 116). Functional redundancy was evaluated with gene-level inventories generated by HUMAnN2 (117) using the UniRef50 (118) database. Finally, we compared the results from the ancient datasets to modern human microbiome datasets to evaluate our ability to take a deeper ecological approach with the former as well as to identify possible changes in microbiome structure and resilience between ancient and modern microbiomes.

2.3 METHODS

Archaeological Context of Novel Data

Maya samples: The ancient Maya occupied northern Central America and parts of eastern Mexico from around 1000 BCE up contact with the Spanish in the 1500s CE, and their descendants still occupy the region today. The earliest Maya lived in small, widely scattered farming villages during the Preclassic period (1000 BCE-250 CE), and by the Classic period (250-830 CE) Maya villages, towns, and cities covered the region. During the Terminal Classic period (830-1000 CE), heartland of the Maya area experienced a significant disruption as the Maya political system collapsed and populations declined precipitously (119). The seven Maya samples used in this study originate from burials at two sites in western Belize: Chan Chich, a moderately sized civic-ceremonial center, and Chan, a small farming community 50 km to the south. The Chan Chich samples derive from two burials (Burials CC-B12 and CC-B14) in the

same building, Structure D-1, one of three structures in a small special-purpose courtyard near the site's main plaza. Radiocarbon dating estimates Burial CC-B12 to cal. 713-885 CE, and Burial CC-B14 was interred slightly earlier based on burial context (120). The Chan samples derive from Burials CH6 and CH19. Burial CH6 was recovered the principal building of the site ceremonial center (cal. 170 BCE-50 CE). The individual in Burial CH19 was interred in an L-shaped structure in the West Plaza of the site (cal. 570-660 CE) (121-123).

Nuragic Sardinia samples: The Nuragic period lasted from Middle Bronze Age to early Iron Age in Sardinia, Italy (~1600–800 BCE) and the Sardinian Nuragic population shows a typical early European farmer ancestry profile, although with a contribution from groups of the eastern Mediterranean and North Africa, related to the commercial trade networks existing with these populations (124). The Nuragic society was substantially based on agriculture (cereals, legumes, grapes and figs) (125) and animal husbandry (sheep, goats, cattle, pigs) (126), whereas evidence of aquatic foods is limited (127). Collective burials were common as demonstrated by the minimum number of individuals found (MNI) at each of the archaeological sites from where the samples analyzed in this paper originated: Lu Maccioni (MNI = 40), Capo Pecora (MNI = 20), and Perdalba (MNI = 30) (128). Lu Maccioni (Alghero) is a natural cave located at sea level in Northern Sardinia (cal. 1126-825 BCE). Capo Pecora (Arbus) is a natural cave located in Southern Sardinia, at 63 m above sea level (cal.1384-936 BCE), and Perdalba (Sardara) is located in Central Sardinia at 163 m above sea level. It is a collective burial structure of the so-called "domus de janas" (*home of the witches*), prehistoric artificial hypogea characteristic to Sardinia, and can be ascribed archaeologically to Nuragic times. All samples analyzed in this paper belong to the Nuragic osteological collection housed in the Sardinian Museum of

Anthropology and Ethnography of the University of Cagliari (128).

Shotgun-sequencing of ancient dental calculus samples

Details of the dental calculus samples from Maya (n=7) and Nuragic individuals (n = 11) used in this study are provided in Supplementary Table A-1. All samples were processed at the Laboratories of Molecular Anthropology and Microbiome Research (LMAMR) at the University of Oklahoma following established protocols for ancient DNA (129). Up to 10 mg of dental calculus was used for DNA extraction, following an ancient DNA extraction protocol customized for dental calculus (130). DNA libraries were built using established protocols (130) with a modification: DNA extracts were partially treated with uracil DNA-glycosylase (UDG) as given in (131). Libraries were dual-indexed using the Kapa HiFi Uracil+ enzyme (Kapa Biosystems), quantified using the Fragment Analyzer (Agilent), and pooled in equimolar ratios. Size-selection was performed for a target range of 150-1000 bp using the PippinPrep (Sage Systems). Libraries were quantified using the Kapa Library Quantification kit (Kapa Biosystems) and sequenced on multiple runs (2 x 150 bp) of the Illumina HiSeq 3000 at the Oklahoma Medical Research Foundation, Oklahoma City, to an average of 16 million reads per sample (Supplementary Table A-1).

Data processing

Previously published shotgun-sequencing data for ancient dental calculus samples from individuals from the historical Radcliffe Infirmary Burial Ground collection at Oxford, UK (n = 44) and ancient human coprolite samples from Rio Zape, Mexico (n=8) were downloaded from the European Nucleotide Archive (Supplementary Table A-2). Previously published modern

human microbiome datasets (fecal and dental calculus) were downloaded from NCBI. Participants with diabetes and/or inflammatory bowel disease were excluded from the MetaHIT-China dataset (n = 38) (Supplementary Table A-2). Newly-generated as well as previously published data were processed using the same customized bioinformatics pipeline. Sequence reads were processed and merged using AdapterRemoval v2 (132), using a minimum overlap of 10. Reads were trimmed to remove Ns and low-quality bases and reads with a Phred score less than 30 were discarded.

Assessing preservation of ancient microbiome signatures

Post-processed reads were mapped to the GreenGenes (133) database of bacterial and archaeal 16S rRNA gene sequences using bowtie2 (134) with default parameters and the --no-unal option. Resulting SAM files were converted into BAM files, sorted, and duplicate reads were removed using SAMTools (135). Custom scripts were used to generate a FASTA file comprising all the unique reads across all samples; this FASTA file was used as input for closed-reference OTU picking at a 97% identity threshold, implemented in QIIMEv1.9 (136) using uclust. Taxonomic inventories at the genus level were generated using QIIME scripts and were used as input for SourceTracker2 (137) to determine the proportion of reads attributed to oral, gut, and other sources.

Authenticating ancient DNA

Reads from keystone taxa identified for each population were authenticated as ancient using the program MapDamage 2.0 (138). Analysis ready reads for all samples in the population were separately mapped to the reference genomes of the keystone taxa identified for that population

using bwa aln (139) with the following parameters: -n 0.03, -q 37, -l 1024, as suggested for ancient DNA (140). Duplicate reads were removed using DeDup (141) and the resulting BAM files were used as input for MapDamage. When the keystone taxa was identified to the species level, that bacterial species was used as the reference genome for MapDamage. For the keystone taxa only identified to the genus level, we identified species belonging to the respective genus using MetaPhlAn and used this species as a reference for MapDamage: in the Rio Zape coprolites, we used *E. coli* for *Escherichia* and *B. pilosicoli* for *Brachyspira.* For the Nuragic dataset, we used *O. uli* for *Olsenella*

Generating taxonomic and functional profiles

Metagenomic taxonomic inventories were generated from the post-processed reads using default parameters in MetaPhlAn v2.0 (110). Downstream analyses (networks, diversity) were conducted with the species-level data, excluding species with a mean abundance < 0.05% in each dataset. Functional profiles were generated from the post-processed reads using default parameters in HUMAnN2 (117) and the UniRef50 database (118). The gene family output tables were used for downstream analysis after normalizing each sample's gene abundance to copies per 1 million gene copies. These tables report abundance of each gene in every sample, as well as a stratified breakdown of how much each taxon contributes to the respective gene abundance.

Network Analysis

Filtered taxonomic tables for each sample in a dataset were combined into a single species-level taxonomic table for each dataset and were used as input for network generation in R (142). SparCC networks were generated for each dataset following the protocol outlined by Laygerfield

et al. (109) which accounts for compositionality in microbiome data by using sparse correlation coefficients. In brief, the input taxonomic table was used to create a sparse correlation coefficient matrix (using the SpiecEasi library) (143), which was in turn used as input to generate an undirected network implemented with the iGraph library (144). Edges connecting nodes in our networks represent positive Pearson correlations >0.3. Networks were generated 100 times for each dataset in order to provide estimates of keystone taxa, modularity, transitivity, and number of clusters. Keystone species scores for the different approaches (HubScore, PageRank, and Closeness) were generated for each taxon in the resulting network with default settings from the iGraph library (144). Keystone taxa with scores in the top 5 of all taxa were saved in each iteration of network generation, and the taxa that appeared in the top 5 in more than 80 of the 100 networks were determined to be potential keystone taxa. Cluster membership was determined using a walktrap algorithm, which performs random walks between nodes (144). Modularity and transitivity for each network were determined with default parameters from iGraph (144) and range in value between 0 and 1. Categorical modularity groups were defined as: very low (< 0.1), low (0.1 – 0.15), medium (0.15 – 0.2), high (0.2 – 0.3) and very high (> 0.3). Categorical transitivity values were defined as: very low (< 0.4), low (0.4 – 0.5), medium (0.5 – 0.6), high (0.6 – 0.7) and very high (> 0.7). These values were determined by the distribution of modularity and transitivity values in each of the networks we generated and are meant to provide relative categories for these network attributes across the datasets we analyzed. We defined the network distinctness ratio as modularity divided by transitivity, as a way to measure how these variables change between datasets and sample size. A more in-depth discussion discussing the techniques and theory used in microbiome network analysis is available in Supplementary Material A – Microbiome Network Analysis

The number of clusters, modularity, and transitivity values reported in Tables 1 and 3 are mean averages from the 100 network iterations. The co-occurrence cluster (Supplementary Figure A - 5) represents the number of times (out of 100 network iterations) that common oral taxon along the y-axis is found in the same cluster as the taxon of interest across the x-axis based on cluster membership determined with the walktrap algorithm.

HUMAnN2 gene family tables were used to source gene abundance data from each keystone taxon. We calculated the average gene abundance in each sample for every UniRef50 annotation from each dataset in R. Within each dataset, the top 50 most abundant genes from the keystone taxa were used to evaluate potential functional importance of each keystone (Supplementary Tables A – 3-6).

Functional Redundancy and Response Diversity Analysis

Gene abundance for each gene, or gene group, of interest (acetate kinase, butyrate kinase, methylmalonyl-coa decarboxylase, fimbrial proteins, flagellar proteins, and adhesin proteins) were acquired from the HUMANn2 gene family tables for each dataset. Gene-abundance not attributed to any taxa (i.e. unclassified) was removed from downstream analysis because we were focused on the diversity of taxa encoding each gene. Gene-abundance per taxon tables were used to determine species richness, phylogenetic diversity, and Gini-Simpson. Species richness was calculated as the number of taxa encoding each gene. For phylogenetic diversity, we created a FASTA file comprising complete 16S rRNA gene sequences from the EzBioCloud database (145) for all the taxa identified in dental calculus and feces. These sequences were aligned using

MAFFT with default parameters (146) and FastTree2 (147) was used within QIIMEv1.9 to create a phylogenetic tree, which was loaded into R. The vegan (148) and picante (149) libraries in R were used to calculate Gini-Simpson and phylogenetic diversity, respectively. The same process for calculating richness, phylogenetic diversity, and Gini-Simpson was used on the MetaPhlAn2 filtered output table to evaluate these metrics for the full community. Plots were generated in R using ggplot2 (150).

Sample size analysis

5, 10, and 20 samples were randomly subsampled in R from each of the modern fecal microbiome datasets. The MetaPhlAn2 taxonomic tables from each subsampling were filtered to remove taxa with < 0.05% mean abundance. Network generation and downstream analysis were performed in the same way as for the full dataset. The keystone taxa, number of clusters, modularity, and transitivity for each small sample size dataset was compared to the full datasets to evaluate the effect of small sample size on network properties.

Statistical Tests

All tests for statistical significance were carried out in R. Where reported, p-values were determined with Kruskal-Wallis tests and false discovery rate correction (151).

2.4 RESULTS

Coprolites

Network Analysis

We find a mean of 2.09 clusters (Table 2-1) across the network for the Rio Zape coprolites (Figure 1). Taking a broader view of network properties, the low modularity (mean = 0.11) in the

Rio Zape network, indicates that the two clusters are highly interconnected. Similarly, the high transitivity (mean = 0.67) demonstrates that nodes are highly connected to each other outside of central nodes. *Eubacterium biforme* and *Phascolarctobacterium succinatutens* were identified as potential keystone species in each of the approaches used to discover keystones (Table 2-2). Reads mapping to each keystone taxon were authenticated as ancient using MapDamage 2.0 (138, 152) (Supplementary Figure A – 1A-D). *Escherichia* and *Brachyspira* were also identified as keystone taxa; however, species-level resolution could not be obtained. Using MetaPhlAn2, we determined the presence of gut-associated members of these genera, such as *E. coli* and *B. pilosicoli*, respectively, in addition to other unclassified species.

Population	Sample Type	Number of Clusters	Modularity	Transitivity
Rio Zape (n = 8)	Coprolites	2.09 (sd = 0.43)	0.111 (sd = 0.010)	0.667 (sd = 0.003)
Maya (n = 7)	Dental Calculus	2.64 (sd = 0.67)	0.052 (sd = 0.008)	0.822 (sd = 0.004)
Nuragic (n = 11)	Dental Calculus	2.71 (sd = 0.87)	0.102 (sd = 0.013)	0.704 (sd = 0.003)
Radcliffe (n = 44)	Dental Calculus	14.14 (3.3)	0.063 (sd = 0.006)	0.738 (sd = 0.002)

Table 2-1: Network Properties of ancient microbiome ecology datasets.

Modularity was defined as: very low (< 0.1), low (0.1 – 0.15), medium (0.15 – 0.2), high (0.2 – 0.3) and very high (> 0.3). Similarly, transitivity was defined as: very low (< 0.4), low (0.4 — 0.5), medium (0.5 – 0.6), high (0.6 – 0.7) and very high (> 0.7). All ancient datasets have low or very low modularity and high or very high transitivity.

Population		Likely Keystone Taxa			
Rio Zape (n = 8)	Page Rank	Eubacterium biforme	Phascolarctobacterium succinatutens	Escherichia unclassified	Brachyspira unclassified
	Closeness Centrality	Eubacterium biforme	Phascolarctobacterium succinatutens	Escherichia unclassified	Brachyspira unclassified
	HubScore	Eubacterium biforme	Phascolarctobacterium succinatutens	Escherichia unclassified	Brachyspira unclassified
Maya (n = 7)	Page Rank	Fusobacterium nucleatum	Treponema denticola		
	Closeness Centrality	Fusobacterium nucleatum	Treponema denticola	Cardiobacterium_v alvarum	
	HubScore	Fusobacterium nucleatum	Treponema denticola	Cardiobacterium_v alvarum	

Nuragic (n = 11)	Page Rank	Eubacterium saphenum	Olsenella unclassified	Streptococcus gordonii	
	Closeness Centrality	Eubacterium saphenum	Olsenella unclassified		
	HubScore	Eubacterium saphenum	Olsenella unclassified		
Radcliffe (n = 44)	Page Rank	Treponema socranskii			
	Closeness Centrality	Treponema socranskii	Tannerella forsythia		
	HubScore	Treponema socranskii	Tannerella forsythia	Neisseria elongata	

Table 2-2: Keystone taxa identified from ancient microbiome datasets.

Likely keystone taxa were identified using three approaches (Page Rank, Closeness Centrality and Hubscore). Each network was generated 100 times and in each iteration the five most likely keystones from each approach were saved. The table above represents taxa that appear in at least 80 of the 100 iterations. There is strong agreement for each dataset's keystones, regardless of approach used.

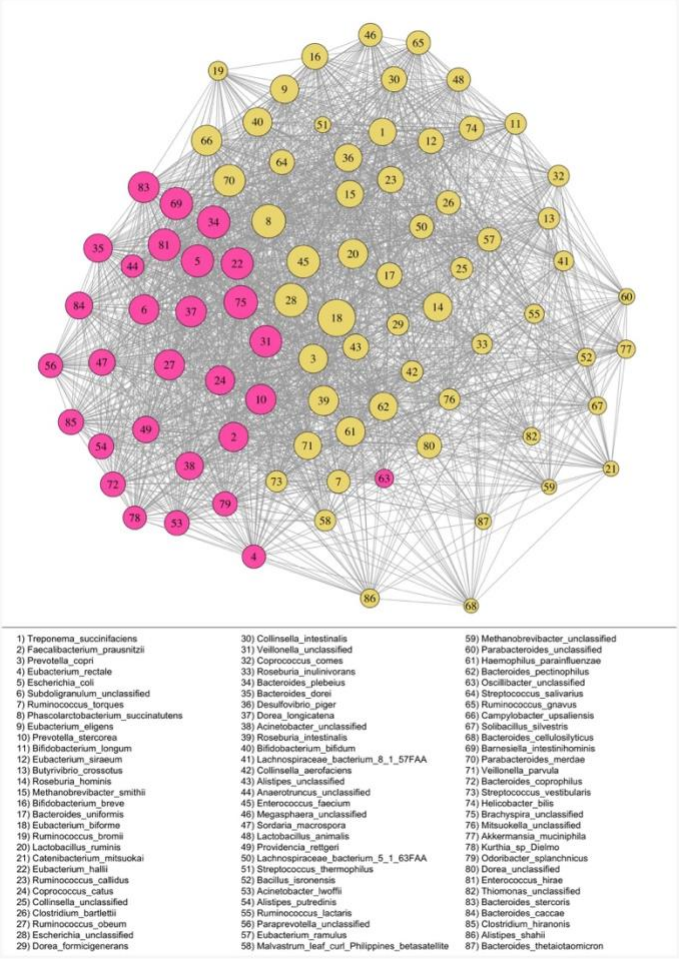


Figure 2 - 1: Rio Zape coprolite network.

Rio Zape coprolite network (n = 8) generated with SparCC. Clusters are differentially colored, keystones are outlined in black, and edges between nodes represent Pearson correlations >0.3. Refer to legend for taxa corresponding to each numbered node. Clusters and nodes are highly interconnected, which is consistent with the low modularity and transitivity values observed.

Keystone functions

The functional roles of these keystones were interpreted by identifying the top 50 most abundant genes found in each taxon. Each keystone taxon has a high abundance of typical housekeeping genes, such as genes involved in synthesis of ribosomal RNA, transferases, and transcriptional regulators (Supplementary Table A-3). We also identified genes involved in antibiotic-resistance mechanisms (MATE efflux proteins in *Brachyspira, E. biforme,* and *P. succinatutens*, and acriflavin-resistance proteins in *Brachyspira* and *P. succinatutens*). Toxin-antitoxin proteins were abundant in *Escherichia*, and transposases were abundant in all the keystone taxa (Supplementary Table A-3).

Functional redundancy

We used a gene-centric approach to evaluate functional redundancy and response diversity as estimators of resilience in the coprolite microbiome. Short-chain fatty acids (SCFAs) such as acetate, butyrate, and propionate, are critical for maintaining a properly functioning human gut microbiome (153-155) and therefore are an intuitive starting point for investigating gene-level functional diversity. We focused our analysis on three SCFA synthesis genes: acetate kinase (acetate), butyrate kinase (butyrate), and methylmalonyl-CoA decarboxylase (propionate). We observe higher diversity for acetate kinase in species richness (p-value $< 4 \times 10^{-7}$), phylogenetic diversity (p-value $< 2 \times 10^{-9}$), and Gini-Simpson (p-value < 0.003) (Figure 2-2 A-C). These

results point towards high response diversity (high number of phylogenetically diverse species) and more evenly distributed production for the taxa encoding acetate kinase, resulting in functionally redundant production of acetate kinase in the Rio Zape coprolites. Butyrate kinase and methylmalonyl-CoA decarboxylase are similar to each other in species richness and phylogenetic diversity (p-value > 0.05), while Gini-Simpson is higher for taxa encoding butyrate kinase (p-value < 0.03). Higher Gini-Simpson index values for butyrate production, compared to propionate, suggests a more even distribution of taxa encoding butyrate kinase and therefore greater protection against shifts in taxonomic abundance that may ultimately cause a decrease in propionate production.

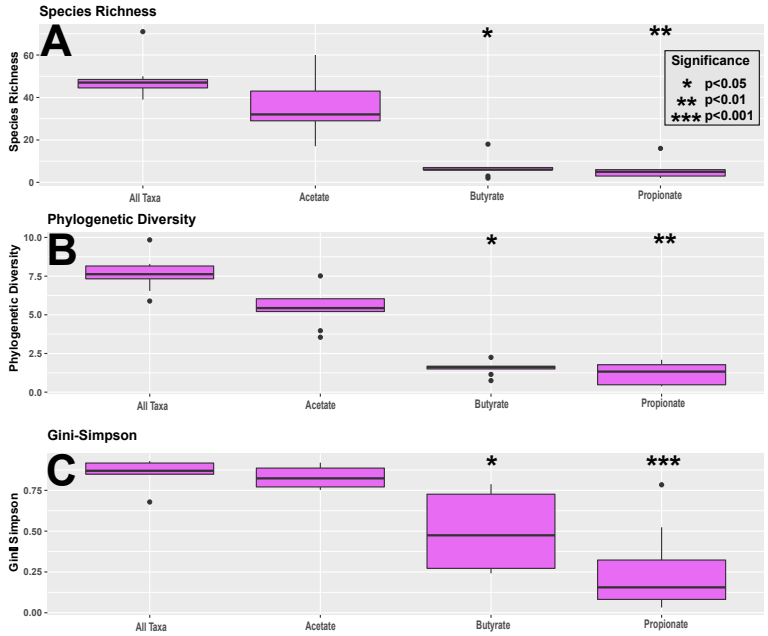


Figure 2 - 2: Functional diversity of the Rio Zape coprolite for SCFA synthesis.

Functional diversity in the Rio Zape coprolites for short chain fatty acid synthesis. A) High functional redundancy (richness), B) response diversity (phylogenetic diversity), and C) evenness (Gini-Simpson) are observed for acetate, indicating production of acetate was more resilient than butyrate and propionate in the Rio Zape population. Taxa encoding butyrate are more evenly distributed than those encoding propionate.

Dental Calculus

Shotgun metagenomic data were generated for Maya individuals (n = 7) and Nuragic individuals (n = 11). SourceTracker2 (137) analysis showed preservation of the oral microbiome signature in all samples, as evidenced by the proportion of reads attributed to taxa commonly found in subgingival or supragingival plaque (Supplementary Figure A - 2).

Network analysis

The two archaeological populations show similar network properties: the Maya population (Figure 2-3A) shows an average of 2.64 clusters, modularity of 0.052, and transitivity of 0.822. The Nuragic population (Figure 2-3B) shows 2.71 clusters, modularity of 0.102, and transitivity of 0.704 (Table 2-1). Modularity is significantly higher in the Nuragic population (p-value $< 2 \times 10^{-16}$) and transitivity is higher in the Maya population (p-value $< 2 \times 10^{-16}$), yet overall, both populations show very low or low modularity and very high transitivity compared to other networks generated in our analysis (Table 2-3). Low modularity values are consistent with a network that has highly interconnected clusters; the clusters lack independence. Similarly, the high transitivity values reflect the diverse paths to connect the bacterial species in each network, providing further evidence of high interconnectivity in the network. The historical Radcliffe population (Figure 2-3C) has significantly more clusters (14.1) compared to the Maya (2.64) and Nuragic (2.71) populations (p-value $< 2 \times 10^{-16}$), which is likely driven by higher sample size in the Radcliffe dataset (see *Sample Size Simulation* section in Results). Despite this, the Radcliffe dataset is similar to the archaeological dental calculus in having very low modularity and very high transitivity (Table 2-1).

Population	Biological Source	Sample Type	Number of Clusters	Modularity	Transitivity
Rio Zape (n = 8)	Feces	Ancient Coprolites	2.09 (sd = 0.43)	0.111 (sd = 0.010)	0.667 (sd = 0.003)
Matses (n = 26)		Modern Feces	6.46 (sd = 1.57)	0.178 (sd = 0.017)	0.465 (sd = 0.004)
HMP, USA (n = 50)			17.03 (sd = 3.23)	0.379 (sd = 0.021)	0.268 (sd = 0.009)
Hadza (n = 25)			7.79 (sd = 1.65)	0.199 (sd = 0.014)	0.402 (sd = 0.009)
China (n = 38)			9.77 (sd = 3.29)	0.256 (sd = 0.015)	0.377 (sd = 0.005)
Maya (n = 7)	Dental Calculus	Ancient Dental Calculus	2.64 (sd = 0.67)	0.052 (sd = 0.008)	0.822 (sd = 0.004)
Nuragic (n = 11)			2.71 (sd = 0.87)	0.102 (sd = 0.013)	0.704 (sd = 0.003)
Radcliffe (n = 44)			14.14 (sd = 3.3)	0.063 (sd = 0.006)	0.738 (sd = 0.002)
Spanish (n = 10)		Modern Dental Calculus	2.67 (sd = 0.84)	0.101 (sd = 0.008)	0.632 (sd = 0.002)

Table 2-3: Network properties of ancient and modern microbiome networks.

Basic network properties of the ancient and modern microbiome ecology datasets. Modularity was defined as: very low (< 0.1), low (0.1 – 0.15), medium (0.15 – 0.2), high (0.2 – 0.3) and very high (> 0.3). Similarly, transitivity was defined as: very low (< 0.4), low (0.4 – 0.5), medium (0.5 – 0.6), high (0.6 – 0.7) and very high (> 0.7). Modern gut microbiomes datasets have higher modularity and low transitivity than the Rio Zape coprolites. Modern dental calculus is similar to ancient dental calculus.

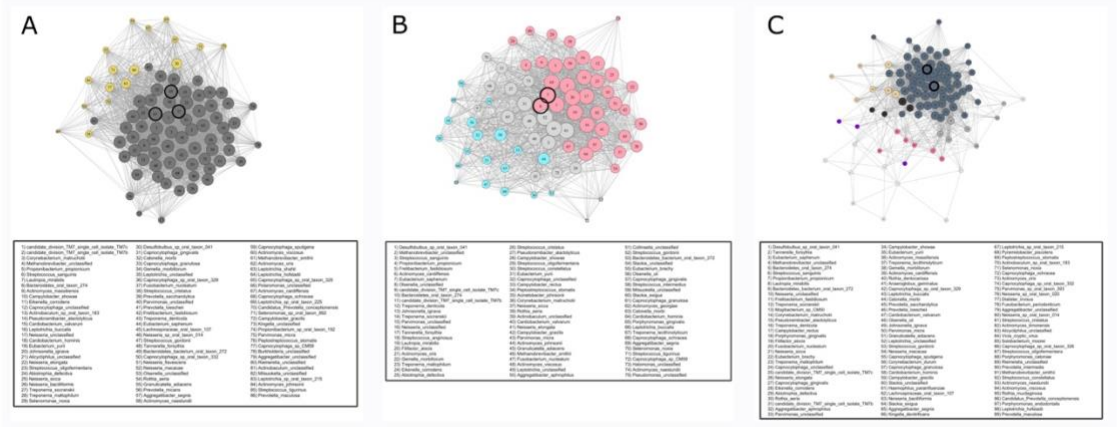


Figure 2 - 3: Networks for the ancient dental calculus datasets

Networks for the three dental calculus datasets, A) Maya, B) Nuragic, and C) Radcliffe. Clusters are differentially colored, keystones are outlined in black, and edges between nodes represent Pearson correlations >0.3. Refer to legend for taxa corresponding to each numbered node. The high number of clusters in the Radcliffe network is likely related to increased sample size in this

dataset. Highly interconnected clusters and nodes in each network is consistent with the low modularity and transitivity values observed.

The keystone species identified in each population are known oral taxa (Table 2-2). Reads mapping to these taxa were authenticated as ancient for the Maya and Nuragic datasets on the basis of DNA damage patterns, generated using MapDamage 2.0 (Supplementary Figures A - 3,4), suggesting that the networks represent an accurate ancient oral ecology. Additionally, keystone species were consistent regardless of analytical approach used (Table 2-2). Taxa associated with periodontitis progression were identified as keystone in each of the populations: *Treponema socranskii* and *T. forsythia* in Radcliffe, *Eubacterium saphenum* and *Olsenella* sp. in Nuragic Sardinians, and *Fusobacterium nucleatum* and *Treponema denticola* in Maya. *Cardiobacterium valvarum* was also identified as a keystone in the Maya population.

Co-occurring taxa

We next evaluated the co-occurrence patterns of selected taxa of interest: early colonizing bacteria *Streptococcus gordonii, Streptococcus sanguinis,* and *Actinomyces naeslundii* (37, 65, 156), as well as periodontitis-associated bacteria *T. forsythia*, *T. denticola,* and *P. gingivalis* (37, 53, 157-159). These bacteria were chosen to study cluster co-occurrence because they play an important role in ecological interactions and functions; early colonizers are among the first bacteria to colonize the dental surface and periodontitis-associated bacteria can shift the community to a disease state. We documented how often common members of the oral microbiome are found in the same cluster as each taxon of interest (Supplementary Figure A - 5). In the Maya and Radcliffe populations, early colonizers like *S. gordonii* and *S. sanguinis* co-occur in the same cluster as the periodontitis-associated bacteria *T. forsythia* and *T. denticola*. In

the Nuragic population, we observed a similar trend but *S. gordonii* does not co-occur with the other oral taxa and another early colonizer, *A. naeslundii,* appears to take *S. gordonii's* place. *A. naeslundii* does not co-occur with the above-mentioned oral taxa in the Maya and Radcliffe populations.

Keystone functions

Similar to the coprolites, the most abundant genes encoded by the keystones identified in each of the dental calculus samples include transporters, transferases, and ribosomal proteins. Efflux-related proteins linked to antibiotic-resistance (MATE efflux and RND efflux) were identified as highly abundant genes in each of the Maya and Radcliffe keystones, but in neither of the Nuragic keystone taxa (Supplementary Tables A - 4-6). The Nuragic keystones both encode putative pathogenic genes: bacteriocin in *E. saphenum* and virulence activator in *Olsenella*. *F. nucleatum* in the Maya samples was the only other keystone found to encode similar genes (hemolysin and ethanolamine utilization). Finally, the keystones *C. valvarum* (Maya), *T. denticola* (Maya), and *Olsenella* (Nuragic Sardinians) were found to encode toxin-antitoxin genes and other stress-response genes in high abundance.

Functional redundancy

For gene-centric analyses, we focused on proteins involved in dental calculus formation via cell-cell binding (adhesins, flagellar, and fimbrial proteins) (62) to give a better understanding of the functional redundancy of proteins involved in dental calculus formation. The Radcliffe and Maya populations have alpha diversity similar profiles for each binding protein and metric, while the Nuragic population has significantly lower richness than both Radcliffe and Maya populations

for each gene (p-value < 0.04; Figure 2-4A). The Nuragic population has significantly lower phylogenetic diversity for every gene (p-value < 0.0005) and Gini-Simpson for fimbrial and flagellum genes (p-value < 0.0008) compared to Radcliffe (Figure 2-4 B-C). The Maya population has significantly greater phylogenetic diversity for fimbrial genes (p-value < 0.003) and greater Gini-Simpson for adhesin and flagellum genes (p-value < 0.03) as compared to the Nuragic Sardinians (Figure 2-4 B-C).

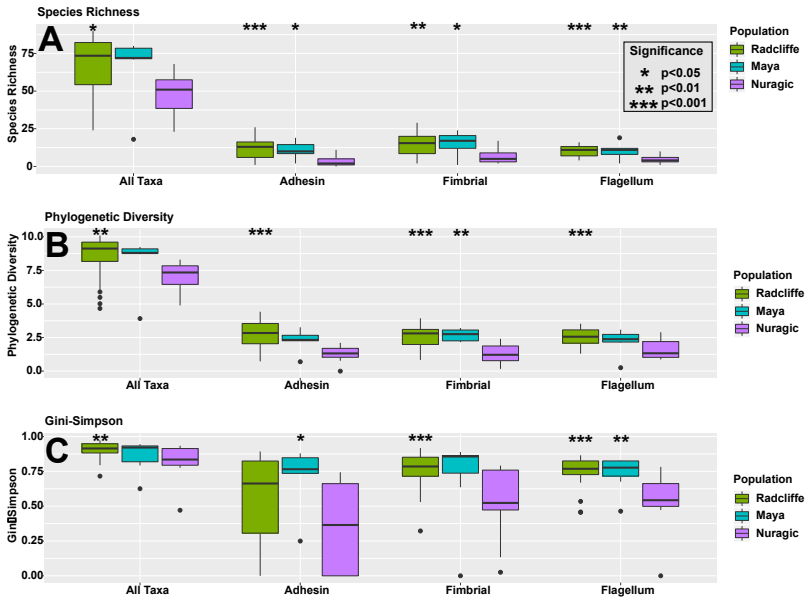


Figure 2 - 4: Functional diversity in ancient calculus datasets.

Functional diversity in the ancient calculus datasets for genes involved in bacterial cell adhesion and cell-cell binding. In general, the Maya and Radcliffe datasets have greater A) functional redundancy, B) response diversity, and C) evenness compared to the Nuragic samples for each gene of interest. These oral ecosystems may have been more robust in terms of dental calculus deposition and growth. Significant p-values are given in reference to the Nuragic dataset.

Articulation points

Neither the coprolites nor dental calculus networks have articulation points. This is likely related to the low modularity in the networks: the interconnectivity of clusters within each network makes it likely for clusters to be connected to each other through multiple nodes.

Comparison to Modern Microbiomes

The coprolite and dental calculus networks inform about general properties (numbers of clusters, connectedness, and articulation points) and identify keystone species. We compared these data to modern microbiome datasets to assess the viability of ancient networks. The Rio Zape coprolites were compared to modern fecal microbiome datasets that represent hunter gatherers (Hadza (24) and Matses (17)) and industrialized populations (MetaHIT-China (160) and Human Microbiome Project (161)). The ancient dental calculus was compared to modern Spanish dental calculus (107). It is important to compare ancient dental calculus to modern dental calculus and not modern dental plaque, as dental calculus is distinct from dental plaque in maturation stage and ecology (107). The small number of datasets mean that broad interpretations may be limited but it is a useful practice, nonetheless. The coprolite dataset showed fewer clusters (p-value $< 2 \times 10^{-16}$), lower modularity (p-value $< 2 \times 10^{-16}$), and higher transitivity (p-value $< 2 \times 10^{-16}$) than modern fecal datasets (Table 2-3). Unlike the low modularity and high transitivity found in the Rio Zape coprolites, the modern fecal microbiome networks had medium to very high modularity and low to very low transitivity (Table 2-3). Both modern and ancient dental calculus datasets each have low to very low modularity and high to very high transitivity, while a high number of clusters is only found in the Radcliffe ancient dental calculus dataset (Table 2-3). The Human Microbiome Project fecal microbiome network was the only network that had articulation points. Keystone taxa were not shared between ancient and modern datasets, except for *P. succinatutens* serving as a keystone in the Rio Zape coprolites and modern Hadza hunter gatherers.

The Rio Zape coprolites had similar response diversity and redundancy profiles when compared to the modern Matses and Hadza hunter gatherers for taxa encoding SCFA synthesis genes. Overall, acetate kinase had the highest alpha diversity in each dataset, regardless of metric used (p-value < 1×10^{-6}) (Supplementary Figure A - 6A-C). No significant differences were observed between the coprolites and Hadza (p-value > 0.05), but the Matses hunter gatherers had significantly greater phylogenetic diversity for propionate synthesis (p-value < 4×10^{-4}) and Gini-Simpson for butyrate synthesis (p-value < 7×10^{-10}) compared to the ancient coprolites. Industrialized populations were significantly more diverse than the ancient coprolites for all metrics in butyrate kinase and methmalonyl-CoA decarboxylase, as well as for species richness in acetate kinase (p-value < 0.03). This observation is likely related to ascertainment bias that hinders annotation and taxonomic identification in non-industrial gut metagenomes (162), but this area bears further study.

Modern Spanish dental calculus had higher alpha diversity for all genes of interest in each metric when compared to the prehistoric Nuragic dental calculus (p-value < 0.05) (Supplementary Figure A - 6D-F). Likewise, modern dental calculus had greater richness than Radcliffe and Maya populations (p-value < 0.012), with the exception of taxa encoding flagella in the Maya population. Modern dental calculus had significantly higher phylogenetic diversity than Maya and Radcliffe datasets for fimbrial production (p-value < 0.006) (Supplementary Figure A - 6E) but there were no significant differences between Radcliffe, Maya, and modern dental calculus in Gini-Simpson (Supplementary Figure A - 6F).

Sample Size Simulation

Most archaeological sites will provide small sample sizes. In our study, to address the effect of sample size on uncovering ecological interactions from human microbiomes, we simulated the effect of small sample size using modern GI tract microbiome data (See Methods). In brief, we randomly subsampled five, ten, and twenty samples from each dataset, then filtered taxa and generated the networks in the same way as we did for the full datasets. We found that both the number of clusters ($r^2 = 0.91$) and network distinctness ratio ($r^2 = 0.88$) increase with sample size (Supplementary Table A - 7, Supplementary Figure A - 7A-B). A high network distinctness ratio means high modularity and low transitivity, and therefore increased sample size leads to more clusters that are highly distinct from each other. We performed the same small sample size simulation with the Radcliffe ancient dental calculus dataset, which was our only dental calculus dataset with more than 20 samples. Similar to the fecal microbiome datasets, the number of clusters increased with sample size ($r^2 = 0.94$); however, there was no increase in the network distinctness ratio ($r^2 = 0.52$) (Supplementary Figure A - 7A-B), indicating there are more clusters but those clusters are still highly interconnected. For both the fecal microbiomes and dental calculus, the keystones found in the full sample dataset were not found in any of the five-sample datasets and only rarely found in the 10-sample datasets (Supplementary Table A - 8, Supplementary Figure A - 8). The keystones identified in the 20-sample dataset were similar to the keystones found in the full datasets (Supplementary Table A - 8, Supplementary Figure A - 8).

The gene-specific approaches do not appear to be hindered by small sample size. As discussed above, we observed similar profiles between modern fecal datasets (Hadza, n = 26 and Matses, n

= 25) and small coprolite datasets (Rio Zape = 8), as well as similar profiles between large ancient dental calculus datasets (Radcliffe, n =44) and small ancient and modern dental calculus datasets (Maya, n = 7 and Spain, n =10). Therefore, analysis of functional diversity in ancient human microbiome datasets remains robust even when few samples are recovered archaeologically.

2.5 DISCUSSION

Archaeologists have made use of coprolites and dental calculus to study human biology, nutrition, and cultural behavior (86-88, 90, 91, 104, 163). Applying ecological approaches to ancient human microbiomes from these materials is a clear next step to provide a deeper understanding of biology in the past. As research on modern human microbiome ecology is still in its infancy, it is expected that ancient microbiome ecology research will lag behind, but it should not be ignored. We have found that by focusing on ecological elements that can be interpreted from single time-point samples, such as ecological network properties, clusters of bacteria, keystone species, functional redundancy, and response diversity, we can gain a glimpse of ecological interactions and functional diversity in ancient human microbiomes.

The four keystone taxa identified in the Rio Zape coprolites are known members of the contemporary human gut microbiome, which provides a validation for prehistoric keystone taxa. *E. biforme* and *P. succinatutens* are commensals that can produce the SCFAs butyrate (155, 164) and propionate (155, 165), respectively, in addition to performing other functions. The role of *Escherichia* in the gut is variable and has been identified in both disease and health-associated states (68, 166). Lastly, *Brachyspira* is primarily found in the GI-tract of pigs (167), chickens

(168), and humans (169) and is associated with diarrhea and other GI-tract maladies (170, 171); however, members of this genus can survive in soil for up to four months after fecal shedding (172). While our analysis of keystone taxa was unable to provide species-level resolution for *Brachyspira,* a MetaPhlAn2 analysis showed that one of the species identified in the coprolites was *B. pilosicoli*. *B. pilosicoli* causes intestinal spirochaetosis in humans (169) and reads mapping to *B. pilosicoli* were authenticated as ancient using MapDamage 2.0, suggesting that *B. pilosicoli* could be a keystone species in this population. The diverse roles of the coprolite keystone taxa suggest that they may dominate distinct niches that lead to different impacts on human biology.

High response diversity and redundancy for acetate kinase is expected as acetate is the most abundant SCFA found in the human gut microbiome and is known to be encoded by diverse groups of bacteria (153-155, 173) . Nevertheless, it is encouraging that we observed this trend in coprolites as further support that we picked up a gut microbiome profile. The lower Gini-Simpson values for methylmalonyl-coa decarboxylase indicates that a few species dominate production of propionate, while production of butyrate is more evenly distributed between taxa. From an ecological perspective, the Rio Zape ancient microbiomes were likely more prone to loss of propionate production than acetate and butyrate because only a few, non-phylogenetically diverse bacteria dominated propionate production.

The keystone taxa identified in the Radcliffe and Maya dental calculus datasets (*T. forsythia* and *T. socranskii* and in Radcliffe and *T. denticola* and *F. nucleatum* in Maya) are members of the red and orange complex group of bacteria associated with periodontitis. Red complex bacteria

are associated with driving periodontitis (159, 174), and orange complex bacteria can function as bridging microbes that facilitate proliferation of red complex bacteria (175). However, the simple presence of 'orange' and 'red' complex bacteria does not guarantee periodontitis progression, as the disease is complex (37, 53, 157, 159). Nevertheless, the presence of these bacteria as keystones, as well as other disease-associated keystone taxa in the Nuragic population (*E. saphenum* and *Olsenella* sp.) (157), indicates that such ancient oral microbiomes are prone to periodontitis. *C. valvarum,* found to be a keystone species in the Maya population, has been associated with endocarditis (176, 177) and also has been isolated from the oral cavity of patients with periodontitis (178). While we do not have any information on the cardiovascular health of the Maya individuals included in this study, the presence of *C. valvarum* in ancient dental calculus further supports the idea that the oral cavity has long hosted bacteria known to be involved in cardiovascular disease (93).

The Nuragic dental calculus is generally similar to the other two ancient dental calculus datasets for network properties; yet it is distinct in functional diversity and patterns of co-occurrence, highlighting the benefit of using multiple approaches to study ecological variation in the microbiome. There is significantly lower response diversity and redundancy in the Nuragic population for each gene of interest. These genes are involved in bacterial cell-cell binding and development of biofilms, which suggests that this population had unique ecological interactions during dental calculus deposition and growth. Along those lines, *S. gordonii* does not cluster with other oral bacteria and is replaced with *A. naeslundii* in the Nuragic Sardinian dental calculus. In the other datasets, *A. naeslundii* does not cluster with the other oral bacteria, while *S. gordonii* does. These two bacteria are early colonizers of the dental surface and therefore may

represent alternative paths to early ecological interactions involved in ancient dental calculus formation.

The keystone taxa found in the coprolites and dental calculus are enriched for antibiotic resistant genes. The presence of antibiotic-resistance proteins in coprolites and dental calculus is anticipated; antibiotic-resistance is a natural result of millions of years of microbial evolution. However, it is noteworthy that three of the four keystone taxa in the Rio Zape coprolites are enriched for antibiotic-resistance proteins, suggesting how this mechanism may be important in a gut microbiome ecology. The Nuragic dental calculus was once again distinct due to the lack of antibiotic resistant genes found in its keystone taxa, yet the Nuragic keystones were enriched for pathogenic genes. The explanation for why the Nuragic population exhibits a seemingly distinct oral ecological community remains elusive. Host genetics may play a role, as Nuragic Sardinians had very low genetic diversity (179) and host genetics does have an impact on the make-up of the human oral microbiome (180); however, we did not analyze human genetics in our study and therefore we cannot provide further resolution for this idea. The unique oral microbiome in Nuragic Sardinians could also result from extensive use of copper mined from the island during the Bronze Age (181, 182). Copper has antimicrobial properties (183) and copper oxide, which is a product of heating copper (184) and has been found in Sardinian Bronze Age artifacts (185), is antimicrobial and has been shown to inhibit oral biofilm formation (186). It is possible that copper affected Nuragic Sardinian oral microbiomes, such as through direct, accidental inhalation while working with the material or through copper leaching into water/food; however, we did not examine copper content of the dental calculus. These hypotheses may be of interest to future anthropological research.

The lack of articulation points in the ancient microbiome datasets means that there are no specific weak-link taxa that would result in a disconnected network if they were removed. Such flexibility in ecological structure can be beneficial but may also mean less stability in taxonomic and functional interactions. However, not much is understood about articulation points in microbiome networks, let alone ancient microbiome networks, and more work needs to be done to develop the theory in this area. While this may point to greater ecological stability, it is more likely a result of flexibility in the network structure (meaning low modularity), which is directly tied to sample size.

Contemporary hunter gatherers shared a keystone species with the coprolites (*P. succinatutens*). Additionally, the contemporary and ancient hunter gatherer fecal microbiomes had similar response diversity and redundancy profiles for SCFA production. Both observations indicate a potential overlap in ecological community structure and function in contemporary and ancient hunter gatherers. The similarity in ecological profiles is exciting because it also demonstrates that ecological interpretation is feasible with ancient microbiome datasets. A similar conclusion can be drawn from comparing the modern Spanish dental calculus to the Maya and Radcliffe ancient dental calculus. Each of these dental calculus datasets have low modularity, high transitivity, and similar phylogenetic diversity and Gini-Simpson values for each gene of interest. There are likely different factors driving similarity in ancient and modern gut microbiomes than the factors driving similarity in ancient and modern dental calculus, but these observations present opportunities for deeper investigations into how lifestyle changes over time influence variation in ecological interactions and functional redundancy within microbiomes.

While we were able to demonstrate key ecological signatures of ancient human microbiomes, we remain cautious about interpretation and application of ecological approaches to studying these biomaterials. A primary concern is sample size. In fecal datasets with small sample sizes ($n < 10$), we observed fewer clusters, lower modularity, and high transitivity. This pattern means that clusters will consist of many taxa and the clusters will be highly interconnected, which may obfuscate more nuanced ecological interactions. However, in the dental calculus dataset, we observed fewer clusters at small sample size, but no change in the network distinctness ratio with sample size, meaning increased sample size does not result in more separation between clusters and nodes. Nevertheless, both the fecal and dental calculus small-sample datasets report different keystone taxa than the respective full datasets. Given the current data available, archaeological studies with representative microbiome samples greater than 20 is strongly suggested for such analyses.

Unfortunately, excavating more than 20 coprolites with sufficient microbiome data to perform ecological analysis from a single site is unlikely. A major challenge is the presence of soil and non-GI tract bacteria in coprolites. Even in the best cases, human GI tract microbiome bacteria make up less than 75% of microbial DNA in coprolites (89); thus, improved methods to isolate gut-derived molecules are required. Furthermore, among the coprolites that are consistent with the gut microbiome, the gut may not be solely human; for instance, dogs are coprophagic and suspected human coprolites may, in fact, be from dogs. Additionally, host DNA content may leach between coprolites, as well as other sources (187). Fortunately, recently developed bioinformatic approaches are improving our ability to distinguish human from non-human

coprolites (103, 188). Finally, coprolites are relatively delicate and often expose the sample to processes that alter DNA sequences and fragment nucleic acids. While microbiome data has been successfully recovered from coprolites in diverse sets of environments (89, 189-191), we would expect such success to be an exception rather than the rule. Even when a set of coprolites do prove to retain a GI microbiome community, small sample size may hinder ecological interpretation. Microbiomes from mummies initially provided an intuitive avenue to study ecology of ancient human gut microbiomes but was ultimately discovered to be misleading as the human gut, upon death, continues to be a moist, warm, enclosed bioreactor shaping the ecology to resemble that expected of compost (89). Because ecology focuses so closely on taxa-taxa abundances and taxa-gene interactions, the preservation issues of a coprolites presents a major challenge, but Rio Zape proved to be an exception, as our results show that we can still study resilience and redundancy with small samples sizes.

A further challenge is authenticating that communities are in fact ancient human microbiomes. Importantly, previously published datasets used in this analysis validated their sequencing reads and we did the same for our newly generated data with SourceTracker2 (137), where the majority of our reads come from expected oral microbes. As expected for ancient DNA, all samples had reads which could not be assigned to known taxa in the database (categorized as 'unknown' in SourceTracker2). We included these reads in our analyses, since a majority of them likely originate from ancient oral microbes but cannot be confidently assigned due to existing databases being biased towards reference strains from modern, industrialized populations. These ‘unknown’ reads may belong to taxa performing important functions and therefore removing them may bias results. An additional validation for the recovery of an ancient microbiome is to

analyze post-mortem DNA damage patterns for reads mapping to the keystone taxa, by using programs such as MapDamage 2.0 (138). Damage patterns consistent with ancient DNA lend strong credibility that the taxa at the center of ecological interactions (i.e. the keystone taxa) are truly ancient, and not arising from recent contamination. Contamination from modern sources, either environmental or from human microbiomes during lab work, would be evident in both the types of microbes identified as keystone taxa, as well as a lack of the prototypical ancient DNA damage (192) in these keystones. Keystone taxa indicative of recent contamination would be bacteria found at high abundance in soil and/or human skin microbiomes. However, our results indicate that we are profiling an ancient microbial ecosystem because our keystone taxa are gut/oral microbes and have prototypical ancient DNA damage.

A greater interest in the maturing of ecological theory for microbiomes is needed, but applying such theory effectively requires a serious investment in mitigating ascertainment biases that burden current reference databases. Publicly available reference databases are skewed towards microbiomes from modern, industrial settings, of often health-associated microbiomes, which bias functional annotation of ancient and non-industrial studies. This ascertainment bias explains why we observe high taxonomic diversity for the industrialized gut microbiome datasets and provide at least a partial explanation of 'unknown' reads in ancient dental calculus results. Our functional diversity approach relies on mapping to marker genes identified from reference taxa. Poor reference representation from non-industrialized populations will lead to bacterial genes and taxa being missed and categorized as 'unknown'. Future microbiome initiatives must avoid exacerbating these biases, with an attention to data that informs, and contextualizes, the microbial ecology.

Chapter 3 – Non-Industrial Gut Microbiomes Provide a More Resilient Ecology for Short-Chain Fatty Acid Production[3,4]

3.1 ABSTRACT

High taxonomic diversity in non-industrial human gut microbiomes is often interpreted as beneficial; however, it is unclear if taxonomic diversity engenders ecological resilience (i.e. community stability, metabolic continuity). We estimate resilience through taxonomic richness, phylogenetic diversity, and evenness in short-chain fatty acid (SCFA) production among a global gut metagenome panel of 11 populations (n = 451) representing industrial and non-industrial lifestyles, including novel metagenomic data from Burkina Faso (n = 90). We observe significantly higher genus-level resilience in non-industrial populations, while SCFA production in industrial populations is driven by a few phylogenetically closely related species (belonging to *Bacteroides* and *Clostridium*), meaning industrial microbiomes have low resilience. Additionally, database bias obfuscates resilience estimates, as we were 2-5x more likely to identify SCFA-encoding species in industrial microbiomes compared to non-industrial. We observe high ecological diversity in non-industrial gut microbiomes, and thus high SCFA resilience, despite database biases that favor industrial populations, while SCFA production in industrial populations is driven by a few phylogenetically closely related species (belonging to *Bacteroides* and *Clostridium*), meaning industrial microbiomes have low resilience. Additionally, database bias obfuscates resilience estimates, as we were 2-5x more likely to identify SCFA-encoding species in industrial microbiomes compared to non-industrial. We

[3] Adapted from Jacobson et al. in review. Non-Industrial Gut Microbiomes Provide a More Resilient Ecology for Short-Chain Fatty Acid Production. *Scientific Reports*

[4] See Supplementary Material B for full list of authors and affiliations

observe high ecological diversity in non-industrial gut microbiomes, and thus high SCFA resilience, despite database biases that favor industrial populations

3.2 INTRODUCTION

Lifestyle alterations have repeatedly coincided with biological changes throughout the human past (193) and this is particularly true for how industrialization changed the relationship between humans and our resident microbes (23). Compared to industrial human gut microbiomes, non-industrial gut microbiomes have higher taxonomic richness, functional enrichment of amino acid metabolism, greater diversity of genes involved in complex carbohydrate metabolism, and higher amounts of short chain fatty acids in stool (17, 194). These trends have been linked to diets rich in plants and fibers, infrequent consumption of highly processed foods, and low exposure to pharmaceutical drugs, such as antibiotics, in non-industrial populations (195).

Higher diversity in the gut microbiome is typically considered healthy, all other factors being equal (23), which would imply that a non-industrial gut is healthier than the industrial gut, in the absence of pathogens and other confounding variables. Yet, commonly used diversity statistics oversimplify more complex microbial associations. Ecological approaches that provide context for microbe-microbe interactions, and present insights into how taxonomic shifts influence microbial and host metabolic processes, are making progress towards mitigating this issue. Taxa-gene relationships are at the heart of deeper ecological understandings of human microbiomes and can be assessed through functional diversity and redundancy (35, 77). Functional diversity, which is similar to the macroecological concept of response diversity (56), refers to the abundance and phylogenetic diversity (PD) of taxa that encode specific genes. It conceptualizes

structure-function relationships in the microbiome by tying together taxonomic and metagenomic gene abundance data. Similarly, redundancy can be thought of as the total number of taxa encoding a function, as well as how evenly the production of any given protein is spread amongst taxa.

Functional diversity can be multi-layered, ranging from a fine-tuned focus on individual genes to a broad genome-wide approach. Gene-centric approaches present the opportunity for niche-specific interpretations, while a broader approach allows for study of how entire microbiomes may shift in the face of outside perturbations. No matter the depth and focus of study, high functional diversity is found in microbiomes where phylogenetically diverse bacteria encode the same functions. Under an idealized model, phylogenetically diverse taxa will have an equal contribution to gene production, leading to high redundancy. Functional diversity and redundancy are intertwined and together estimate microbiome resilience. Shifts in taxonomic abundance are less likely to alter the functional potential of a resilient community because any given function is encoded by a wide range of bacteria and production is distributed between these diverse taxa. Consequently, the loss of one phylogenetic branch of bacteria within the ecosystem will not cause a loss of function that those bacteria encode; however, communities with low functional diversity and redundancy may suffer ecosystem-wide functional changes during minor taxonomic perturbations. Accurately quantifying functional diversity is therefore a necessary part of ecologically-minded microbiome research because it more deeply describes how structure-function relationships influence resilience in a microbiome.

Short-chain fatty acids (SCFAs) synthesis is the most intuitive area of study for understanding the ecological differences between industrial and non-industrial gut microbiomes, given the trends attributed to high-fiber diets among non-industrial populations. SCFAs are important byproducts of microbial metabolism and fermentation in the human gut. The three most prominent SCFAs in the human gut (acetate, butyrate, and propionate) are vital for maintaining tight junction integrity between epithelial cells in the gastrointestinal (GI) tract, serve as an energy source for colonocytes, and signal immune cells, amongst a number of other functions (153, 196). Unsurprisingly, variation in SCFA abundance is a classic link to human health. For example, high butyrate levels are found to decrease diastolic blood pressure via regulating inflammation, and acetate abundance is tied to appetite, thus impacting metabolic regulation (153, 196).

Studying SCFA functional diversity is particularly intriguing as it provides a line of evidence as to whether estimates of taxa/gene diversity and ecological resilience are concordant. Research suggests that non-industrial populations have high SCFA abundance, which is attributed to dietary composition (194, 195). It is assumed that non-industrial gut microbiomes bear an ecology that is resilient for SCFA production due to high overall taxonomic diversity and high SCFA levels in stool, but this has not been demonstrated. We address this gap by using metagenomic data from industrial (European/North American and Central/East Asian), pastoral, rural agricultural, and hunter-gatherer populations to compare functional diversity and redundancy of SCFA synthesis genes (Figure 3 - 1). We chose to evaluate SCFA genes that are involved in end-stage synthesis in different pathways for each SCFA: acetate kinase (*ackA*) for acetate (197), butyrate kinase (*buk*) and butyryl-CoA:acetate CoA transferase (*but*) for butyrate

(198, 199), and methylmalonyl-CoA decarboxylase *(mmdA)*, lactoyl-CoA dehydratase (*lcdA*), and CoA-dependent propionaldehyde dehydrogenase (*pduP*) for propionate (165, 198). These genes are known to be encoded by a diverse range of bacteria (Supplementary Table B - 1), which permits ecological investigation into the resilience of SCFA production. Our study includes previously published metagenomic datasets from industrial and non-industrial populations, as well as novel gut microbiome metagenomic data generated from fecal samples collected from rural agriculturalists living in central Burkina Faso (n = 90). We functionally profiled these metagenomes using HUMAnN2 (200).

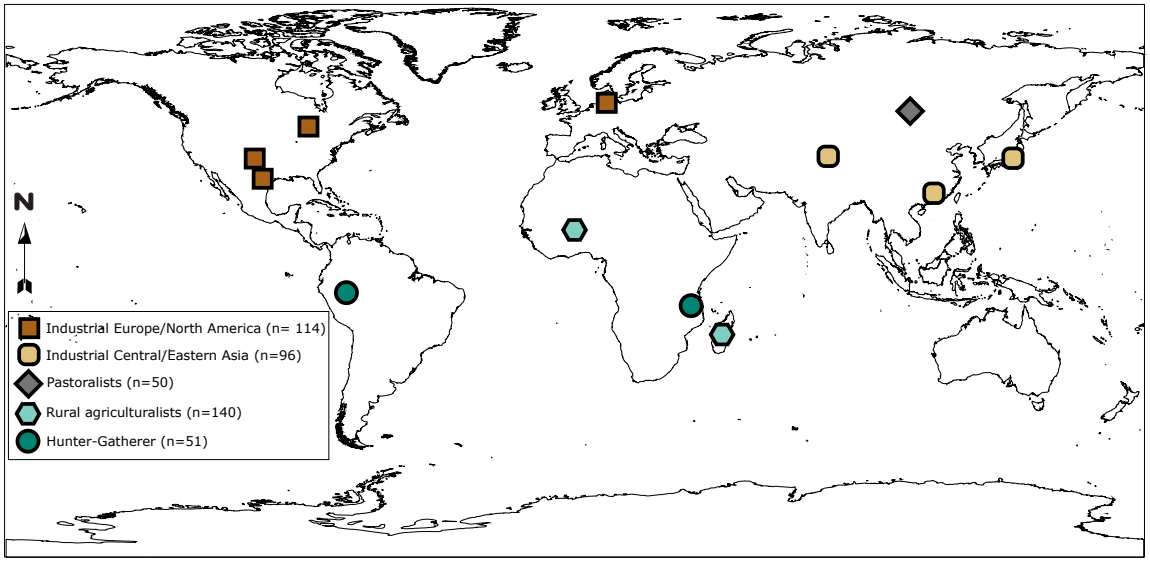


Figure 3 - 1: Map of Human Gut Microbiome Metagenomes Analyzed

Analyzed microbial metagenomes originated from the following populations/datasets: Industrial North American/European – Human Microbiome Project(201) (Missouri, Texas – USA, n = 50), Oklahoma(17) (USA, n = 21), Northern Europe(202)(n = 43) ; Industrial Central/East Asia – China(203) (Guandong Province, China, n = 38), Tokyo(204) (Japan, n = 32), Astana(205) (Kazakhstan, n = 26); Pastoral – Mongolia(206) (Khentii Province, n = 50); Rural Agriculturalist – Burkina Faso (n = 90), Madagascar (n = 50); Hunter Gatherer – Matses(17) (Peru, n = 25), Hadza(24) (Tanzania, n = 26).

Each gene we analyzed is involved in terminal or near-terminal steps of production of each respective SCFA (165, 197-199). For analytical purposes, taxonomic-gene abundance for *but* and *buk* were combined for butyrate and data for *mmdA, lcdA,* and *pduP* were aggregated for propionate. Response diversity was estimated through taxonomic richness and phylogenetic diversity, while the Gini-Simpson index and Hill Numbers were used to assess redundancy by documenting evenness in the community. For our purposes, the Gini-Simpson index represents the probability that two sequencing reads originate from different taxa, and therefore values close to 1 represent a community with a high diversity of taxa encoding the SCFA, while values close to 0 indicate that the SCFA is encoded almost entirely by one taxon. Hill numbers (207) are a diversity measure that allows for interpretation of effective taxonomic richness at different levels: at diversity order 0 the Hill number richness is equal to the total number of taxa, at diversity order 1 the Hill number richness is the effective number of commonly occurring taxa, and at diversity order 2 the Hill number richness is the effective number of dominant taxa. Therefore, Gini-Simpson and Hill number investigations permit analyzing how evenly SCFA production is distributed between taxa.

3.3 MATERIALS AND METHODS

Study Design

Sample Size*:* Datasets were chosen because they represent a wide diversity of lifestyles, have a minimum of 20 samples per population, and were sequenced to an average read depth of 10 million reads per sample. We used 20 samples as a threshold based on previous research (Jacobson et al. *in review*) showing that at least 20 samples per population are required for the types of ecological analyses pursued in this study. Similarly, 10 million reads was chosen as a

threshold to allow for sufficient read depth to attain coverage of as many genes as possible from each metagenome. The number of reads generated for the Burkina Faso samples is available in Supplementary Table B - 2 and the SRA accession number and number of reads analyzed for the comparative datasets are provided in Supplementary Table B – 2.

Data inclusion/exclusion criteria*:* For datasets with available metadata, we included only healthy adults (i.e. non-obese BMI, non-diabetic) in the analysis; however, children were included in the Matses and Hadza datasets due to limitation in sample size.

Outliers: Outliers were included in all analysis

Research objectives*:* Our research objective was to assess resilience in SCFA production across different lifestyles. SCFAs are a key component of human-microbiome interaction and taxonomic diversity is higher in non-industrial populations. Our pre-specified hypothesis was that resilience would be higher in non-industrial populations. After our first phase of analysis, we uncovered the contradictory results between genus and species level resilience and we hypothesized this was due to reference database bias.

Research subjects: All participants from Burkina Faso were healthy volunteers. For datasets with available metadata, we included only healthy adults (i.e. non-obese BMI, non-diabetic) in the analysis; however, children were included in the Matses and Hadza datasets due to limitation in sample size.

Experimental design*:* Human fecal microbiome samples were collected with informed consent from residents of a single village in central Burkina Faso under ethical approval granted by Centre MURAZ Research Institute in Burkina Faso (No. 31/2016/CE-CM). Gut metagenomic data were generated as given in Borry et al. (188). Sex and age was recorded for each individual and is reported in Supplementary Table B – 2.

Randomization*:* We randomly subsampled 50 individuals, in R, from each of the Madagascar, Human Microbiome Project, and Mongolian datasets, due to the much higher numbers of individuals in these studies as compared to the remaining datasets. We did not want skew the different lifestyle groups with overrepresentation from a single dataset.

Statistical Analysis

Bioinformatic Processing

Metagenomic reads for the following datasets were downloaded from either the NCBI Sequence Read Archive or European Nucleotide Archive: hunter-gatherers (Matses from Peru (17) and Hadza from Tanzania (24)), pastoralists (residents of Khentii region, Mongolia (206)), rural-agriculturalists (Madagascar (208)), industrial European/North American populations (Human Microbiome Project (161), Europe (202, 203), and Oklahoma, USA (17)), and industrial Central/East Asian populations (Japan(204), China(203), and Kazakhstan (205)). Accession numbers can be found in Supplementary Table B – 2.

All metagenomic data (newly generated from Burkina Faso and downloaded) was processed as follows: AdapterRemoval v2 (132) was used to quality filter and merge reads (quality score >30, maxns = 0, minlength > 50, minalignmentlength = 10). The resulting FASTQ files (forward,

reverse, and merged) were used as input for HUMAnN2 (200) with default parameters and using the UniRef50 database (118). Briefly, pangenomes are generated for each taxa identified from metagenomic reads using MetaPhlAn2 (110). Metagenomic reads are mapped against those pangenomes to identify genes; reads not mapping to any pangenome are then mapped against the UniRef50 database (118) to identify 'unclassified' genes. Reads not mapping to neither the pangenomes nor UniRef50 database are termed 'UNAMAPPED'. Gene abundance is normalized to reads per kilobase (RPKs) to account for differences in reference database size. The gene family output at the species level from HUMAnN2 was used to perform downstream analysis. Each sample's output was normalized to gene abundance RPKs per 1 million DNA kilobases and then merged into a single file with all samples. The RPK gene abundance is further stratified by the abundance of the gene that is mapped to a species. Phylum, Family, and Genus level tables were created using the humann2_infer_taxonomy script from HUMAnN2. Acetate, Butyrate, and Propionate gene family tables were generated by pulling out all lines that were annotated with the gene names listed above from the respective normalized phylum, family, genus, species normalized gene abundance tables.

The following gene names were used to identify SCFAs: acetate kinase, butyrate kinase, butyryl-CoA:acetate CoA transferase, methylmalonyl-CoA decarboxylase, lactoyl-CoA dehydratase, and CoA-dependent propionaldehyde dehydrogenase. Acetate kinase (*ackA*)is the primary end stage enzyme for acetate synthesis (197). Butyrate kinase (*buk*) and butyryl-CoA:acetate CoA transferase (*but*) can both catalyze butyrate production from butyryl-CoA (199). Propionate synthesis can follow three different biochemical pathways: succinate, acrylate, and propanediol depending on the initial substrate (165). Methylmalonyl-CoA decarboxylase *(mmdA)* is a

biomarker for the succinate pathway, lactoyl-CoA dehydratase (*lcdA*) for the acrylate, and CoA-dependent propionaldehyde dehydrogenase (*pduP*) for the propanediol pathway (165). Taxonomic-gene abundance for *but* and *buk* were combined for butyrate and likewise data for *mmdA, lcdA,* and *pduP* combined for propionate to facilitate SCFA comparisons.

Total gene copies per 1 million DNA fragments was calculated in R (142) using the normalized 'UNMAPPED' gene abundance generated from HUMAnN2. The stratified gene family tables, after removal of 'UNMAPPED' abundance, were used for the remainder of analysis. The proportion of total gene abundance classified to a taxon was determined by summing the abundance of each gene that mapped to taxon and dividing that value by each sample's total gene abundance. This was repeated at each taxonomic level. The same procedures were applied to the Acetate, Butyrate, and Propionate gene family tables

Ecological Metrics

Richness, phylogenetic diversity (PD), Gini-Simpson, and Hill Numbers (207) values were generated using the vegan package (148) in R. Richness was determined as the number of taxa that have a gene abundance value > 0 for each SCFA. PD was calculated using the 16S rRNA gene as a proxy. The 16S rRNA gene for each taxon found across the full dataset was extracted from the EzTaxon (145) reference database and concatenated into a single 16S rRNA gene FASTA file. These sequences were aligned using mafft (146) with default parameters and a phylogenetic tree was built using FastTree (147) in QIIMEv1.9 (136). PD was calculated with resulting tree and gene family tables using vegan. Gini-Simpson and Hill Number values were determined using the gene family tables with vegan. P-values were determined using the

Kruskal-Wallis H test and the post-hoc Dunn Test. False discovery rate (FDR) was used to account for multiple testing. Plots were generated using ggplot2 (150).

3.4 RESULTS

Independent of lifestyle, acetate synthesis was significantly more abundant than the other two SCFAs (p-value < 8×10^{-84}) and butyrate was more abundant than propionate (p-value < 6×10^{-8}, Supplementary Table B - 3). The overall higher abundance of butyrate compared to propionate across the full dataset is driven by the non-industrial populations, as propionate and butyrate are at similar abundance in industrial populations (Supplementary Table B - 3). The relative abundance ratio of SCFA synthesis genes of acetate:butyrate:propionate (mean = 0.600 [standard error = 0.001] : 0.215 [SE = 0.001] : 0.184 [SE = 0.001]), supports the previous finding of a 60:20:20 ratio of SCFA molarity in stool (153) (Supplementary Table B - 3). Comparing between lifestyles, acetate (FDR-adjusted p-value < 3×10^{-19}, n = 451) and butyrate (FDR-adjusted p-value < 0.002, n = 451) synthesis genes were more abundant in each of the non-industrial populations (Figure 3 - 2). Propionate synthesis genes were similar between lifestyle groups, with the exception of being at significantly lower abundance in rural agriculturalists compared to all lifestyles (FDR-adjusted p-value < 0.006, n = 451, Figure 3 - 2). Within the general lifestyle categories, the rural agriculturalists had significantly lower abundance of butyrate and propionate compared to hunter-gatherers and pastoralists (FDR-adjusted p-value < 0.006, n = 241) while there was no significant difference between European/North American industrial and Central/East Asian industrial populations for any of the SCFA gene groups (FDR-adjusted p-value > 0.05, n = 210) (Figure 3 – 2).

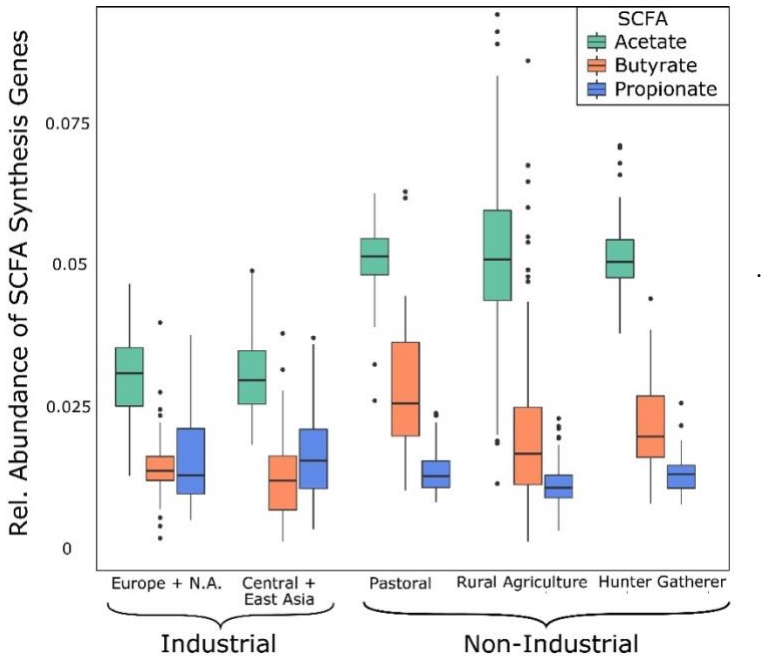


Figure 3 - 2: Relative Abundance of SCFA Genes Compared Between Lifestyle Categories

Acetate (FDR-adjusted p-value < 3 x 10^{-19}, n = 451) and butyrate (FDR-adjusted p-value < 0.002, n = 451) are significantly higher in each of the non-industrial populations compared to the industrial populations. Propionate is significantly lower in the rural agriculturalists compared to all other lifestyles (FDR-adjusted p-value < 0.006, n = 451) but there are no significant differences between the industrial and other non-industrial populations for propionate synthesis (p-value > 0.05, n = 311). Butyrate was significantly lower abundance in rural agriculturalists compared to hunter gatherers and pastoralists (FDR-adjusted p-value < 0.006, n = 241). There were no significant differences between the European/North American (Europe+N.A.) and Central/East Asian industrial population, and likewise, no significant differences between pastoralists and hunter-gatherers for any of the SCFA genes. Statistical comparisons were generated using the Kruskal-Wallis H test and the post-hoc Dunn Test. False discovery rate (FDR) was used to account for multiple testing.

The pastoralist and rural agricultural populations have significantly higher taxonomic richness at the genus level for acetate and butyrate synthesis compared to the industrial populations (FDR-adjusted p-value < 0.0009, Figure 3 - 3A, n = 401); however, hunter-gathers only have significantly greater abundance than the Central/Eastern Asia population for taxa involved in acetate synthesis (FDR-adjusted p-value = 0.013, Figure 3 - 3A, n = 261). Hunter-gatherers have significantly lower genus richness for propionate synthesis compared to both the industrial and

non-industrial populations (FDR-adjusted p-value < $3x10^{-5}$, Figure 3 - 3A, n = 261). The high taxonomic diversity observed at the genus-level in the non-industrial populations for acetate and butyrate is not observed at the species level, as every non-industrial population has significantly lower species richness for each SCFA gene (FDR-adjusted p-value < 0.05, Figure 3 – 3B, n = 451). Additionally, the rural agricultural and hunter-gatherer populations have significantly lower species richness than the pastoralists for acetate, butyrate, and propionate (FDR-adjusted p-value < $8x10^{-5}$, Figure 3 – 3B, n = 241). Similar to species richness, we observed lower species phylogenetic diversity (PD) in the non-industrial populations, but the effect sizes were not as large as in species richness (Figure 3 – 3C). The rural agriculturalists and hunter-gatherers had significantly lower PD for each SCFA, compared to the industrial populations (FDR-adjusted p-value < 0.05, Figure 3 – 3C, n = 401); however, the pastoralists only had significantly lower PD for acetate when compared to the European/North American industrial populations (FDR-adjusted p-value = 0.025, Figure 3 – 3C, n = 164). The small drop-off in species PD, compared to species richness, suggests there are a high number of closely phylogenetically related species in the industrial gut microbiome. We found *Bacteroides* and *Clostridium,* which are found at high abundance in industrial gut microbiomes, to have up to nine species encoding SCFAs, while known SCFA producers at high abundance in non-industrial gut microbiomes (*Prevotella, Faecalibacterium,* and *Phascolarctobacterium)* only had one or two species per each genus (Supplementary Table B - 4).

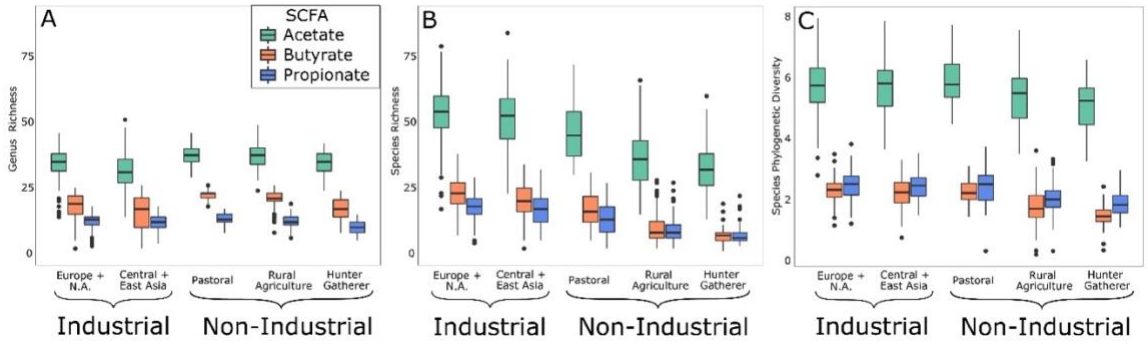


Figure 3 - 3A-C: Taxonomic Diversity of SCFA-encoding Taxa

*Genus richness (**A**), species richness (**B**), and species phylogenetic diversity (**C**) for taxa encoding the different SCFAs. **A**) Pastoralists and rural agriculturalists have higher genus richness for acetate and butyrate, hunter-gatherers have significantly lower genus richness for propionate (FDR-adjusted p-value < 0.05, n = 451). **B**) Species richness is significantly lower in non-industrial populations for each SCFA (FDR-adjusted p-value < 0.05, n = 451) and there is a steep drop-off in non-industrial populations. C) PD is lower in most non-industrial populations but the drop-off from industrial to non-industrial is not as steep as seen in species richness. Statistical comparisons were generated using the Kruskal-Wallis H test and the post-hoc Dunn Test. False discovery rate (FDR) was used to account for multiple testing.*

Genus evenness, as gauged through effective number of taxa at Hill numbers 1 (number of common genera) and 2 (number of dominant genera), tell a unique story for each SCFA. Hunter-gatherers and rural agriculturalists have higher effective number of common and dominant species compared to the industrial populations for butyrate and propionate, but diversity is only greater in non-industrial populations at Hill number 1 for acetate (FDR-adjusted p-value < 0.02, Figure 3 - 4A-C, n = 451). Additionally, the pastoralists have significantly higher diversity than the industrial populations for butyrate (FDR-adjusted p-value < 3×10^{-10}, n = 260) but lower diversity than the European/North American industrial population for acetate (FDR-adjusted p-value < 0.01, n = 260). These results demonstrate the industrial populations have only a few common and dominant genera encoding SCFAs, indicating that they are prone to loss in SCFA production if those common/dominant genera are lost. At the species level, evenness is

significantly lower for each SCFA in every non-industrial population compared to the industrial groups (FDR-adjusted p-value < 4 x 10^{-4}, Figure 3 – 4D-F, n = 451). Similar to the richness and Hill numbers, the Gini-Simpson index is higher in non-industrial populations at the genus-level but lower at the species level (Supplementary Figure B - 1).

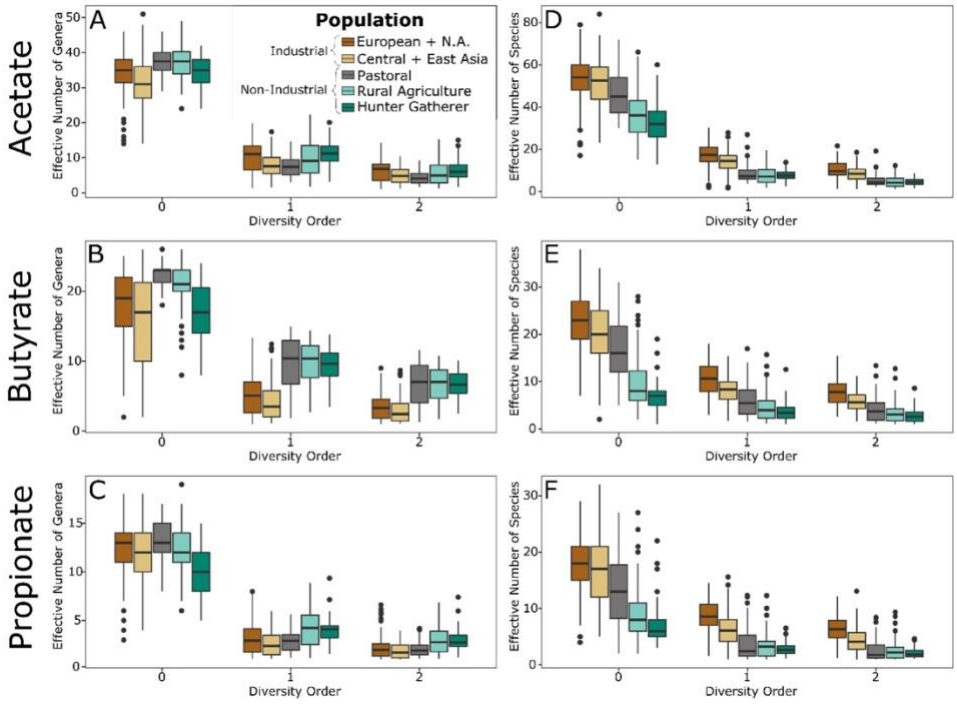


Figure 3 - 4A-F: Hill Numbers for SCFA-encoding Taxa

*Effective number of genera (**A-C**) and species (**D-F**) for each SCFA as determined through Hill numbers at diversity order 0 (total number of taxa), 1 (number of common taxa), 2 (number of dominant taxa). Rural agriculturalists and hunter-gatherers have significantly higher number of common and dominant genera than industrial populations for butyrate and propionate, as well as for acetate at diversity order 1. This means the distribution of SCFA production in non-industrial populations is more even. The number of effective number of species is significantly lower in the non-industrial populations for each SCFA (FDR-adjusted p-value < 4 x 10^{-4}, n = 451). Statistical comparisons were generated using the Kruskal-Wallis H test and the post-hoc Dunn Test. False discovery rate (FDR) was used to account for multiple testing.*

The discrepancy between genus and species results, particularly the drastic drop-off in diversity in non-industrial populations at the species level suggests a loss of information during annotation of non-industrial gut metagenomes. To probe this further, we assessed the proportion of total

DNA fragments that mapped to a gene between lifestyles, as well as the proportion of those gene-mapped fragments that were classified to a taxon. After normalizing gene abundance to genes per 1 million DNA molecules, genes are positively identified from approximately 75% of DNA fragments in the industrial populations, but gene identification drops to about 65% of DNA fragments in non-industrial populations (p-value < 5.41×10^{-15}, n = 451). The stratified HUMAnN2 output provides abundance of genes matched to a taxon ('classified'), as well as gene abundance not accounted for by any taxon ('unclassified'). Across all genes identified in each metagenome, there is a significant decrease in the proportion of gene abundance that is classified to a taxon from industrial to non-industrial populations (FDR-adjusted p-values: phylum < 5.13×10^{-9}, family < 6.27×10^{-9}, genus < 4.24×10^{-8} , species < 8.25×10^{-8}; Supplementary Figure B - 2, n = 451). Only 25-30% of gene abundance is classified to a species in hunter-gatherers and rural agriculturalists but upwards of 65% to 75% of genes are classified to species in industrial populations. Therefore, there are significantly fewer genes identified in non-industrial metagenomics and this loss of information is compounded by substantially worse identification of the taxa that encode those genes in non-industrial gut metagenomes.

The afore-mentioned pattern is replicated for each of the SCFA synthesis gene groups, as there is significantly lower classification percentage at every taxonomic level in the non-industrial populations (FDR-adjusted p-value < 0.05, Figure 3 - 5A-C, n = 451), with the exception of the pastoral populations for all taxonomic levels for acetate and at the phylum level for butyrate. Nevertheless, there are interesting trends for each of the SCFAs. Even though acetate is the most abundant SCFA, the classification percentage is essentially the same for the other two SCFAs (Figure 3 - 5A). For butyrate, the species-level information for non-industrial populations is by

far the lowest of any of the SCFAs (Figure 3 - 5B). All taxonomic levels have poor classification in rural agriculturalists and hunter-gatherers for propionate; phylum-level classification in hunter-gatherers is nearly half of species-level classification in industrial populations (Figure 3 - 5C). For every SCFA, there is a steeper drop-off from genus-level classification to species-level classification in the non-industrial populations compared to industrial populations (FDR-adjusted p-value < 0.001; Supplementary Figure B – 3, n = 451).

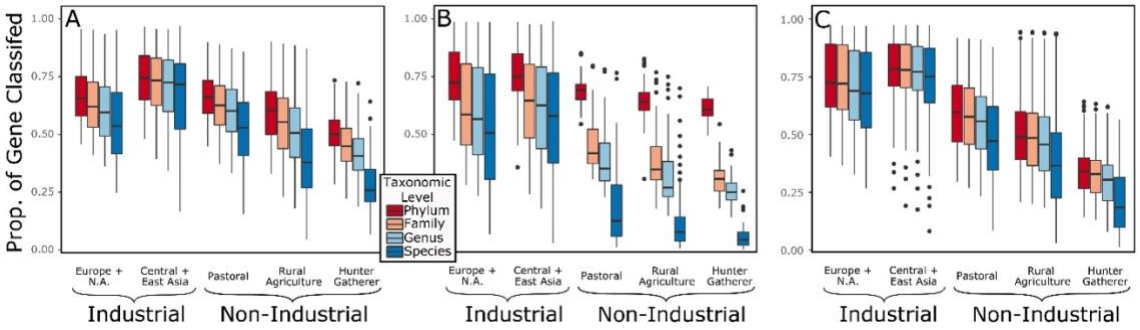


Figure 3 - 5A-C: Proportion of Genes Classified to a Taxon for Each SCFA

*Acetate (**A**), butyrate (**B**), and propionate (**C**). Rural agriculturalists and hunter-gatherers have significantly lower proportion of genes mapping to a taxon at each taxonomic level for each SCFA (FDR-adjusted p-value < 0.05, n = 401). Statistical comparisons were generated using the Kruskal-Wallis H test and the post-hoc Dunn Test. False discovery rate (FDR) was used to account for multiple testing.*

3.5 DISCUSSION

Our results are consistent with a non-industrial gut harboring a more resilient ecology with respect to SCFA production, while the industrial gut ecology would be vulnerable to disruption of such pathways, yet the pattern is complex and nuanced. The increased gene abundance in non-industrial populations and overall ratio of acetate:butyrate:propionate generally agrees with previous studies of SCFAs (153, 195). Similarly, the higher taxonomic diversity of taxa encoding acetate, compared to the other SCFAs, is expected and matches studies that have documented the

taxa that encode different SCFAs (165, 196, 198). The overall high taxonomic diversity, high diversity at Hill numbers 1 and 2, and high Gini-Simpson indices found in non-industrial populations at the genus level indicates a highly diverse and evenly distributed production of SCFAs. From an ecological perspective, uneven production of SCFA dominated by a few bacteria in industrial gut microbiomes means lower functional diversity and less redundancy, which ultimately leads to an expectation of decreased resilience. In other words, this study finds that industrial gut microbiomes are at a higher risk of reduced SCFA production because SCFA synthesis is dominated by only a few taxa. Given the lower resilience, factors that disrupt the gut ecology are expected to have a more extreme consequence to those living an industrial, relative to non-industrial, lifestyle.

While there is an overall trend of increased genus-level functional diversity and redundancy for SCFA production in non-industrial populations, variation exists when examining the SCFAs and populations individually. At the genus-level, the pastoral and rural agricultural populations have increased richness of taxa encoding genes involved in acetate and butyrate synthesis, while there is similarity between the different lifestyles for genus richness for propionate encoding taxa. Although hunter-gatherers have similar, or lower, taxonomic richness as industrial populations, they have significantly higher diversity at Hill number orders 1 and 2 and Gini-Simpson indices for butyrate and propionate. Additionally, the pastoralists have a generally similar profile to the industrial populations for Hill number diversity. This paints a complex picture. Non-industrial populations have a high diversity of taxa encoding butyrate synthesis, and butyrate production is spread more evenly across taxa in non-industrial populations than in industrial populations. Hunter-gatherers and rural agriculturalists have significantly greater evenness of propionate

production, even though they have fewer number of total genera encoding this SCFA. Finally, the richness and evenness of taxa encoding acetate is similar between industrial and non-industrial populations. Ecologically, we would expect the industrial populations to be more likely to suffer decreased production of butyrate and propionate when faced with a shift in taxonomic composition. Non-industrial populations may be only marginally more resilient for acetate production compared to industrial populations. The propionate and acetate results are intriguing because they are found at higher levels in industrial and non-industrial populations, respectively, highlighting the benefit of taking an ecological approach to understanding diversity of metabolic processes in the human microbiome.

The increased species-level alpha diversity in industrial populations runs counter to the genus-level results but genus and species level results are ultimately concordant after accounting for ecology and ascertainment bias. The substantially higher species richness in industrial populations is striking; however, differences in PD between industrial and non-industrial populations are not nearly as extreme. This difference means that high species richness in the industrial populations is driven by taxa that are closely related to each other phylogenetically. Indeed, we observed SCFA producing genera found at high abundance in industrial populations (*Bacteroides* and *Clostridium*) to have up to nine species encoding SCFAs, while highly abundant non-industrial genera only have one or two species. Therefore, what first appears to indicate high ecological resilience in SCFA production in industrial populations is actually the result of closely related species performing the same function. It follows that closely related species may be prone to the same types of ecological removal events, such as antibiotics or xenobiotics that target specific bacterial groups. While this result has ecological implications, it

is also likely the result of historical trends of microbiology research. Taxa at high abundance in non-industrial gut microbiomes have not been a focus of microbiological isolation and species identification until recently, therefore, we expect more species to be identified from non-industrial gut microbiomes in the future. Additionally, classification of bacteria into distinct genera and species is undergoing a revolution in the genomic era (209) meaning that the high number of species classified to *Bacteroides* and *Clostridium* may ultimately be reclassified to different genera. Nevertheless, the fact that we observe a large jump in species richness, but only a minor increase in species PD, in the industrial gut microbiomes suggests that the high industrial species richness is driven by closely related species.

Ascertainment bias extends to the databases used to identify taxa and genes: fewer genes were identified in non-industrial populations and a smaller proportion of these genes can be linked back to taxa at every taxonomic level, in non-industrial gut microbiomes. In some cases, such as butyrate synthesis genes, less than 10% of genes are identified to species for non-industrial populations, while over 50% of such identifications were possible for industrial populations. A decreased ability to identify the genus and species encoding SCFA synthesis genes in non-industrial populations means that the ecological metrics underestimate the true ecological diversity of these genes. The statistically significant differences observed at the genus-level send a strong signal of the high functional diversity, and potential resilience, of SCFA synthesis genes in non-industrial gut microbiomes.

The metagenome-wide poor performance in terms of gene identification and classifying SCFA genes to taxa indicates a bias in reference databases that underrepresents diversity in non-

industrial gut microbiomes, which is unsurprising. Bias is expected because the vast majority of human gut microbiome studies have used samples from industrial populations. There is an immense challenge in including non-industrial communities in biomedical research, including recruiting research participants, sustaining longitudinal sampling, building culturally appropriate community relationships, and even securing transport of samples (162). This has resulted in comparatively few metagenomic studies of human gut microbiomes from non-industrial settings (162). Nevertheless, our data demonstrate the extent of this bias and how it can hinder more in-depth study of human gut microbiome health. Given this sizable ascertainment bias favored industrial populations, the non-industrial populations are likely even more diverse, more resilient, than our databases can sufficiently characterize, making our genus-level results even stronger. Without a serious investment to include such populations, characterization of microbiomes will remain naive to the ecological breadth of the core, healthy, human gut. Imagine studying forest ecology, with only city parks at your disposal. This has been, overwhelmingly, the analogous practice of human microbiome research.

Relative lack of microbiome studies with non-industrial populations means an underrepresentation of not only metagenomic data and genome annotation but also fewer opportunities for cultivation and validation of novel species of bacteria. This underrepresentation ultimately leads to an inequality in the depth to which researchers can describe microbiome samples from non-industrial communities, compared to industrial microbiomes, as diverse groups of novel taxa may be grouped into a single group of “unknown” or “unclassified” bacteria. Similarly, an incomplete picture of microbial functional potential means that genes may be misidentified or even unannotated completely. Unknown taxa and misidentified genes may be

playing key roles in ecological and metabolic processes but researchers are unable to confidently identify them, let alone make statements about their importance in a microbial ecology (162). Recent human gut microbiome metagenome studies from diverse populations will undoubtedly improve database representation but the number of studies and metagenomic samples from non-industrial populations still pales in comparison to industrial gut microbiomes (162, 188, 208, 210).

Limitations in annotating the full extent of microbial diversity does not only affect researchers interested in microbiome ecology, there are real-life consequences. Microbial changes associated with industrialization, recently termed 'Microbiota Insufficiency Syndrome (MIS)' (23), have resulted in a mismatch between microbial communities and human biology. The decreased phylogenetic diversity and loss of specific taxa (e.g. Prevotellaceae, Succinivibrionaceae, and Spirochaetaceae) observed in industrial gut microbiomes may contribute to the increase in non-communicable chronic diseases found at higher prevalence in industrial populations. Our findings of decreased resilience in industrial populations, as well as species-level diversity driven by a few closely related species, fits in well with MIS. However, if we are unable to fully characterize non-industrial gut microbiomes then we will be unable to paint a complete picture of MIS. Currently, we have confidence that there is a wealth of undiscovered resilience in non-industrial gut microbiomes. Once we describe the extent of this diversity/resilience, through increased sampling and focus on partnerships with research institutes in industrializing countries, we will have a more complete picture of MIS and possibly develop therapeutic approaches to combat non-communicable chronic diseases related to the human gut microbiome.

Lack of sample diversity is not unique to human microbiome research, as human genetics research has been grappling with this very issue for decades. In 2009, 96% of individuals included in human genome-wide association studies (GWAS) claimed European ancestry, as compared to 78% in 2019 (211). Thus, while there have been improvements, GWAS clearly fail to reflect the breadth of human diversity. Incorporating diverse populations in human genome and microbiome research has the potential to greatly benefit the scientific community's understanding of human biology and develop treatments that are based on human diversity rather than European-ancestry genetics and microbiomes. A key component of increasing representation in genetics and microbiome studies is that these studies are designed as partnerships with minority and/or indigenous communities in a manner that builds both trust between the community and researchers, as well as facilitates the ability for sample donors to exercise their rights on how data are treated and shared (212).

Chapter 4 – Shifts in Gut and Vaginal Microbiomes Associated with Platinum-Free Interval Length in Women with Ovarian Cancer[5,6]

4.1 ABSTRACT

Many studies investigating the human microbiome-cancer interface have focused on the gut microbiome and gastrointestinal cancers. Outside of human papillomavirus driving cervical cancer, little is known about the relationship between the vaginal microbiome and other gynecological cancers, such as ovarian cancer. In this retrospective study, we investigated the relationship between ovarian cancer, platinum-free interval (PFI) length, and vaginal and gut microbiomes. We observed that *Lactobacillus*-dominated vaginal communities were less common in women with ovarian cancer, as compared to existing datasets of similarly aged women without cancer. Primary platinum-resistance (PPR) disease is strongly associated with survivability under one year and we found over one-third of patients with PPR (PFI<6 months, n = 17) to have a vaginal microbiome dominated by *Escherichia* (> 20% relative abundance), while only one platinum super sensitive (PFI>24 months, n = 23) patient had an *Escherichia*-dominated microbiome. Additionally, *L. iners* was associated with little, or no, gross residual disease, while other *Lactobacillus* species were dominant in women with > 1 cm gross residual disease. In the gut microbiome, we found patients with PPR disease to have lower phylogenetic diversity than platinum-sensitive patients. The trends we observe in women with ovarian cancer and PPR disease, such as the absence of *Lactobacillus* and presence of *Escherichia* in the vaginal microbiome as well as low gut microbiome phylogenetic diversity have all been linked to other diseases and/or pro-inflammatory states, including bacterial vaginosis and autoimmune

[5] Adapted from Jacobson et al. (Submitted). Shifts in Gut and Vaginal Microbiomes Associated with Platinum-Free Interval Length in Women with Ovarian Cancer. *PeerJ.*
[6] See Supplementary Material C for full list of authors and affiliations

disorders. Future prospective studies are necessary to explore the translational potential and underlying mechanisms driving these associations.

4.2 INTRODUCTION

Ovarian cancer is the most deadly gynecological cancer (213); it kills approximately 14,000 women in the United States annually, accounting for 4.9% of all cancer-related deaths in females in the United States. In the majority of cases (> 80%), ovarian cancer is not detected until stage III or later, primarily due to the nonspecific nature of ovarian cancer symptoms and lack of informative biomarkers (214, 215). Early-stage (I or II) detection results in substantially greater five-year survivability compared to late-stage diagnosis (III or IV): 70% versus 36% survival rate, respectively (216), highlighting the importance of discovering early-disease biomarkers.

The standard course of primary treatment in ovarian cancer is cytoreductive surgery (CRS) in combination with platinum-based chemotherapy (217), which causes cytotoxicity through formation of intra- and inter- strand adducts on DNA in cancer cells (218). The diameter of the remaining tumor after CRS, referred to as gross residual disease, is an important predictor of patient outcome, as individuals with no residual disease or residual disease < 1 cm have improved survivability compared to those with tumors > 1 cm after CRS (219). The combination of CRS and platinum-based chemotherapy is highly effective with approximately 80% of patients showing no evidence of disease at the conclusion of therapy; however, recurrences occur in 70-80% of advanced stage patients and 20-25% of early-stage patients (220). Primary platinum resistance (recurrence of cancerous growth within six months of primary treatment cessation) develops in about 20% of patients, and is highly problematic because it is associated with a survivability of under one year and fewer effective treatment options (221). Other patients

may remain free of cancerous growth for more than two years but the risk of recurrence and eventual development of treatment-resistant cancer is still unacceptably high (222-225). This problem merits a focus on discovering biomarkers of ovarian cancer and drivers of platinum-resistance to facilitate early cancer detection as well as better understand variation in treatment outcomes.

Recent evidence suggests that the human microbiome is an important factor in tumorigenesis, carcinogenesis, and effectiveness of chemotherapy (5, 226-228). Gut microbiome dysbiosis can influence colorectal carcinogenesis via production of genotoxic metabolites, such as colibactin, and through promotion of a pro-inflammatory state, which contributes to cancer cell proliferation, angiogenesis, and metastasis (5, 226, 229-231). While most studies on the microbiome-cancer relationship have focused on the gut microbiome, there is growing evidence in support of the relationship between the vaginal microbiome and gynecological cancers. For example, human papillomavirus (HPV) is a known causative agent of cervical cancer (232-234), and pelvic inflammatory disease, which is associated with shifts in vaginal microbiome composition (235), has been linked to ovarian cancer development (236). Yet, there are still many unknowns about the relationship between the vaginal microbiome and gynecological cancers (232-234, 237). Likewise, links between the vaginal microbiome and platinum-sensitivity remain elusive; however, a previous study demonstrated that cancerous growths in mice with antibiotic-depleted gut microbiomes were less susceptible to platinum-chemotherapies compared to those with high gut microbiome diversity (227). The microbiome has a strong, bi-directional relationship with the host immune system (238, 239) and immune cells in the mice with a depleted gut microbiome produced fewer reactive oxygen species (ROS) than control

mice (227). Decreased ROS production and platinum-resistance in mice with low gut microbiome diversity suggests a role for the microbiome in response to platinum-chemotherapy because ROS play a part in cell apoptosis after exposure to platinum chemotherapeutic agents (240).

In this study, we assessed how the vaginal and gut microbiomes vary in ovarian cancer patients with different platinum-sensitivities, with the aim of determining whether the human microbiome can be used as a biomarker of platinum-sensitivity.

4.3 MATERIALS & METHODS

Study Population

Patients who carried a diagnosis of advanced (Stage III/IV) epithelial ovarian cancer and who were classified as primary platinum resistant (platinum-free interval [PFI] from completion of primary platinum based chemotherapy < 6 months) or platinum super sensitive (PFI > 24 months) and were being treated at the Stephenson Cancer Center at the University of Oklahoma Health Sciences Center were approached to participate in this study. We approached patients when they either developed primary platinum resistant disease or when they were identified as platinum super sensitive. We also included patients who had already been diagnosed with primary platinum resistant disease or as platinum super responders but had moved on to additional therapy, as well as patients on active anti-cancer therapy or in surveillance. We excluded patients if: 1) they were taking antibiotics at the time of sample collection or within 14 days prior to sample collection, or 2) they had active vaginal bleeding or known entero-vaginal fistulae. We also collected samples from five individuals who were referred for ovarian cancer

treatment at the Stephenson Cancer Center but ultimately had benign tumors; these served as a non-chemotherapy exposed control group. A brief demographic and medical treatment history for participants in this study (n = 45, median age 62.2, age range 33-83) is provided in Table 4 - 1. This study was approved by the University of Oklahoma Health Sciences Center Institutional Review Board (February 22nd, 2016, reference #6458).

Table 4 - 1: Demographic and clinical data for individuals in this study

ID	Age (years)	Ethnicity	PFI (Months	Cancer Stage	Histology	Cytoreductive Surgery	Residual Disease	Months Last Pla	Neuropathy
ocm002	54.9	White	>24	III	Serous	iterative CRS (iCRS)	No Gross Residual (NGR)	29	none
ocm003	69.5	Native American	<6	III	Serous	primary CRS (pCRS)	>1cm	25	none
ocm004	52.9	White	>24	III	Serous	primary CRS (pCRS)	>1cm	65	none
ocm005	68.9	White	>24	III	Serous	iterative CRS (iCRS)	<1cm	82	none
ocm007	68.5	Native American	>24	III	Serous	primary CRS (pCRS)	No Gross Residual (NGR)	53	>grade2
ocm009	52.8	White	>24	III	Serous	primary CRS (pCRS)	<1cm	98	none
ocm012	59.1	White	>24	III	Serous	primary CRS (pCRS)	No Gross Residual (NGR)	124	none
ocm014	52.8	White	Benign	N/A	N/A	iterative CRS (iCRS)	>1cm	N/A	N/A
ocm021	64.8	White	<6	III	Serous	iterative CRS (iCRS)	No Gross Residual (NGR)	7	UNK
ocm022	77.7	White	>24	III	Serous	primary CRS (pCRS)	No Gross Residual (NGR)	106	none
ocm024	76.5	White	>24	III	Serous	primary CRS (pCRS)	No Gross Residual (NGR)	40	none
ocm026	45.5	White	<6	III	Serous	iterative CRS (iCRS)	No Gross Residual (NGR)	10	UNK
ocm027	59.2	White	<6	III	Serous	iterative CRS (iCRS)	<1cm	7	UNK
ocm028	71.9	White	>24	III	Serous	primary CRS (pCRS)	No Gross Residual (NGR)	45	none
ocm029	68.3	White	<6	III	Serous	iterative CRS (iCRS)	<1cm	4	UNK
ocm031	47.2	White	<6	III	Serous	iterative CRS (iCRS)	<1cm	7	none
ocm032	56.9	White	<6	IV	Serous	iterative CRS (iCRS)	>1cm	6	>grade2
ocm033	73.66	African American	>24	IV	Serous	primary CRS (pCRS)	No Gross Residual (NGR)	68	>grade2
ocm034	77.9	White	>24	IV	Serous	primary CRS (pCRS)	No Gross Residual (NGR)	61	>grade2
ocm036	51	White	>24	IV	Serous	primary CRS (pCRS)	No Gross Residual (NGR)	39	>grade2
ocm043	70.2	White	>24	IV	Serous	iterative CRS (iCRS)	>1cm	67	none
ocm044	58.2	White	>24	III	Serous	primary CRS (pCRS)	No Gross Residual (NGR)	18	>grade2
ocm045	67.9	White	>24	III	Serous	primary CRS (pCRS)	No Gross Residual (NGR)	42	none
ocm046	64.4	White	>24	III	Serous	primary CRS (pCRS)	No Gross Residual (NGR)	54	none
ocm047	70.2	White	>24	III	Serous	primary CRS (pCRS)	<1cm	87	none
ocm049	47.5	White	>24	IV	Serous	iterative CRS (iCRS)	No Gross Residual (NGR)	8	>grade2
ocm051	45.5	White	>24	IV	Serous	primary CRS (pCRS)	No Gross Residual (NGR)	45	none
ocm057	74.9	White	<6	III	Serous	primary CRS (pCRS)	No Gross Residual (NGR)	18	UNK
ocm058	79	Native American	Benign	N/A	N/A	N/A	>1cm	N/A	N/A
ocm060	60.9	White	>24	III	Serous	primary CRS (pCRS)	>1cm	150	none
ocm061	64.6	White	<6	III	Serous	iterative CRS (iCRS)	>1cm	9	none
ocm063	58.6	White	>24	III	Serous	primary CRS (pCRS)	No Gross Residual (NGR)	N/A	UNK
ocm070	65.4	White	<6	III	Serous	primary CRS (pCRS)	No Gross Residual (NGR)	38	UNK
ocm071	57.8	White	Benign	N/A	N/A	N/A	>1cm	N/A	N/A
ocm075	62.5	White	<6	IV	Serous	No Surgery	>1cm	6	UNK
ocm084	33.7	Native American	Benign	N/A	N/A	N/A	>1cm	N/A	N/A
ocm088	53.6	African American	<6	IV	Serous	iterative CRS (iCRS)	>1cm	1	none
ocm089	82.7	White	<6	III	Serous	iterative CRS (iCRS)	<1cm	6	none
ocm091	38.5	White	<6	III	Serous	iterative CRS (iCRS)	No Gross Residual (NGR)	35	UNK
ocm093	59.7	White	>24	III	Serous	primary CRS (pCRS)	No Gross Residual (NGR)	72	>grade2
ocm094	73.6	White	<6	III	Serous	iterative CRS (iCRS)	>1cm	3	UNK
ocm096	63.6	White	>24	III	Serous	primary CRS (pCRS)	<1cm	41	>grade2
ocm098	75.1	White	<6	IV	Serous	iterative CRS (iCRS)	<1cm	15	>grade2
ocm099	61.9	White	Benign	N/A	N/A	N/A	>1cm	N/A	N/A
ocm106	58	White	<6	IV	Serous	iterative CRS (iCRS)	<1cm	3	UNK

All patients were treated initially with platinum and taxane chemotherapy for a planned six to eight cycles. These regimens included paclitaxel and carboplatin given every 21 days, paclitaxel

given weekly with every 21st day carboplatin, or intraperitoneal administration of either cisplatin or carboplatin with intravenous and intraperitoneal paclitaxel. In patients with platinum resistant disease, standard of care options after recurrence included pegylated liposomal doxorubicin (PLD), weekly paclitaxel, gemcitabine, topotecan or bevacizumab given as monotherapy or with chemotherapy. Patients were also screened for eligibility for clinical trials. Patients with a PFI > 24 months had not recurred at the time of study participation and were followed every 6 months with surveillance of Ca-125 values and exams. For those who had recurred beyond 24 months, treatment options included several platinum-based doublets including carboplatin and PLD given every 28 days, carboplatin and paclitaxel given every 21 days or carboplatin and gemcitabine given on a day one and day eight or day one and day fifteen schedule. Each patient completed a minimum of six cycles of treatment and could undergo more cycles as long as the patient was responding and tolerating therapy.

Sample Collection

Samples were collected during standard of care exams in the gynecologic oncology clinic at the Stephenson Cancer Center in Oklahoma City, OK. Catch-All Sample Collection Swabs (Epicentre) were used to collect vaginal and fecal samples. Vaginal swabs were collected from three sites per individual: vaginal introitus (VIT), mid-vagina (MDV), and posterior fornix (VPF), and then placed into a dry sample collection tube. Fecal samples were collected via a rectal digital exam, after which any stool collected was placed on a Catch-All swab and placed in a dry collection tube. Two swabs were collected from each site (bilaterally from the vaginal sites and sequentially for the rectal samples). Each participant completed a quality of life survey

regarding their medical treatment history, antibiotic use within the past year, vitamin consumption, socioeconomic status, and other lifestyle metadata (Supplementary Table C - 1).

Laboratory Methods

DNA was extracted from the left-side vaginal swab and first fecal swab from each patient, using the MoBio PowerSoil DNA Isolation Kit (now Qiagen DNeasy PowerSoil Kit), following manufacturer's protocols with the addition of a ten-minute incubation at 65°C prior to the initial bead-beating step, as recommended in the Manual of Protocols for the Human Microbiome Project (241). A quantitative PCR (qPCR), using the SYBR Green PCR Master Mix (Applied Biosystems) and primers targeting the V4 region of the bacterial 16S rRNA gene (242), was conducted; dilutions of *Escherichia coli* DNA corresponding to known 16S rRNA gene copy numbers were used as quantification standards for the DNA extracts. DNA extracts were amplified in triplicate, using Phusion High-Fidelity DNA polymerase (ThermoFisher Scientific) and Illumina-compatible primers 515F and 806R (V4 region of the16S rRNA gene) with error-correcting Golay barcodes incorporated into the 806R reverse primer (242). PCR products were pooled in equimolar concentrations, purified with the MinElute PCR purification kit (Qiagen), quantified using KAPA Biosystems Illumina library quantification kit, and sequenced across multiple runs of an Illumina MiSeq (500 cycles paired-end sequencing, v2 reagent kit).

Bioinformatic Methods

AdapterRemoval (v2) (132) was used to filter out reads with uncalled bases, reads with Phred quality threshold < 30, and reads less than 150 bp in length. Quality filtered paired-end reads were merged using AdapterRemoval (v2) (132) and then demultiplexed with QIIME (v1.9),

followed by removal of chimeric sequences and low-abundance (<5 total sequences) reads (136). The remaining sequences were used for *de novo* Operational Taxonomic Unit (OTU) clustering with USEARCH (v10) at 97% sequence similarity (243). Taxonomy was assigned to each OTU representative using the EzBioCloud 16S rRNA gene database (145). The resulting OTU table was rarefied to 9000 reads and downstream analysis was performed in QIIME (v1.9) (136). The post-rarefaction sample breakdown was: PFI>24 (n = 23), PFI<6 (n = 17), and benign (n = 5).

Statistical Methods

Phylogenetic diversity and weighted/unweighted UniFrac (244) metrics were generated in QIIME (v1.9) with FastTree2 (147). Tests for significance between study groups for alpha and beta diversity were performed using Kruskal-Wallis and PERMANOVA tests, respectively, in R (142). Vaginal samples were classified into clusters by the dominant bacterial taxon found in each sample, as determined by Ward hierarchical clustering (245), and visualized as a heatmap via the gplots package in R (246). If there were no dominant bacteria, the sample was classified as diverse. Median-unbiased estimated odds ratios were calculated to determine whether study groups had significantly different odds of dominant bacteria; reported odds ratios and 95% confidence intervals were log-transformed. Kruskal-Wallis tests with a Benjamini and Hochberg false-discovery rate adjustment were used to evaluate differential abundance of individual taxa between study groups. Odds ratios were calculated using epitools (247) in R. Plots were generated using the ggplot2 (150) and ColorBrewer (248) packages in R.

4.4 RESULTS

Vaginal microbiome

Samples from different vaginal sites (VIT, MDV, VIT) that originated from the same individual showed similar taxonomic beta-diversity profiles (Supplementary Figure C - 1A-C). Due to this similarity, we grouped sequencing reads from each individual's three vaginal samples into a single representative vaginal microbiome sample per individual for downstream analysis, unless otherwise noted. Firmicutes was the most dominant phylum in the vaginal microbiome and it was found at over 50% relative abundance in 40% of individuals (Supplementary Figure C - 2A), while Proteobacteria, Bacteroidetes, and Actinobacteria were the next most abundant phyla and found at >50% relative abundance in the vaginal microbiome of 13.3%, 13.3% and 6.7% of individuals, respectively. At the genus level, *Lactobacillus, Prevotella, Escherichia, Gardnerella,* and *Streptococcus* were the most dominant bacteria and accounted for 57.2% of all reads (Supplementary Figure C - 2B).

Vaginal microbiome communities were clustered into five community-dominance groups using Ward hierarchical clustering: *Lactobacillus* cluster, *Escherichia* cluster, *Gardnerella* cluster, *Prevotella* cluster, and a high diversity cluster (Figure 4 - 1). Dominance groups were evenly distributed between patients with PFI < 6 months, PFI > 24 months, and benign, with the exception of higher than expected dominance of *Escherichia* in patients with PFI < 6 months: five of the six Escherichia-dominated vaginal communities identified with hierarchical clustering belonged to patients with PFI < 6 months (Supplementary Figure C - 3). Vaginal microbiomes dominated by *Escherichia* had higher odds of occurring in PFI < 6 months compared to PFI > 24

months (log odds ratio [OR] = 2.812, 95% CI 1.027 – 15.059, p-value = 0.043, Figure 4 - 2A). Additionally, one of the patients with PFI < 6 months and one of the patients with benign pathology identified with a 'diverse' vaginal microbiome had *Escherichia* at greater than 20% relative abundance, while no other patients with PFI > 24 months had *Escherichia* relative abundance above 5% (Figure 4 - 1). In total, 35.3% of the patients with PFI < 6 months had *Escherichia* at greater than 20% relative abundance in the vaginal microbiome, compared to 4.34% of PFI > 24 months and only one of five benign individuals. Apart from platinum resistance, *Escherichia* showed no association with any of the other health or lifestyle factors (Figure 4 - 2A). Although *Escherichia* is common lab-grown bacterium and found in feces, our analysis demonstrates *Escherichia* abundance in the vaginal samples is biological and not a technical artifact (Supplementary Material C).

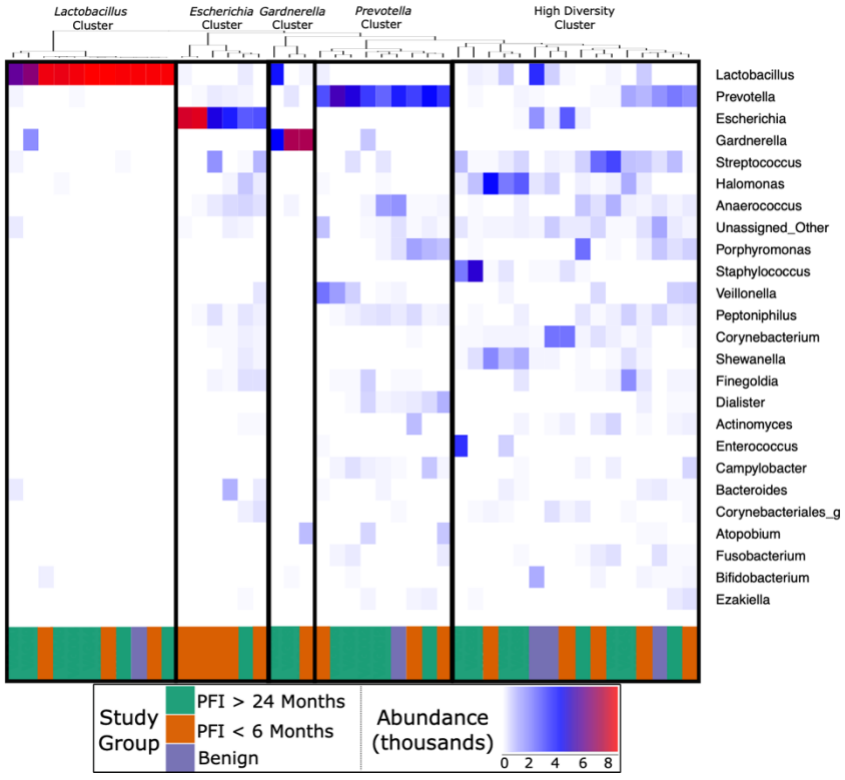


Figure 4 - 1: Heatmap of the 25 most abundant genera in the vaginal microbiome

Each column represents a single individual's vaginal microbiome, color coded by study group. Samples were clustered together based on similarity of vaginal microbiome using Ward hierarchical clustering. 11 microbiomes were Lactobacillus-dominated, six Escherichia-dominated, three Gardnerella-dominated, nine Prevotella, and 16 highly diverse.

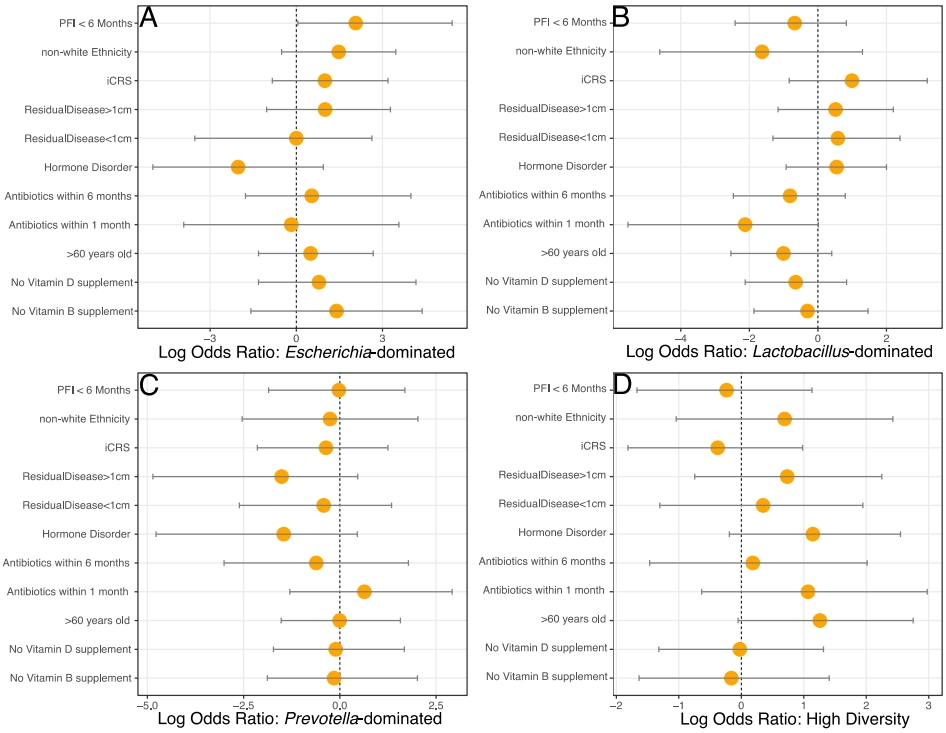


Figure 4 - 2A-D: Log-transformed odds ratio vaginal microbiome dominance.

A) Escherichia had significantly higher odds of occurring in PFI < 6 months individuals (p-value = 0.043). B) Antibiotics within one month was negatively associated with Lactobacillus-dominance, but not significantly (p-value > 0.05). There were no differences between medical/health/lifestyle variables in Prevotella (C) and highly diverse (D) vaginal microbiomes. Gardnerella-dominated communities were not included in this analysis due to small sample size (n = 3).

Approximately 24% (11 of 45) of patients in this study had *Lactobacillus*-dominated communities, which is significantly lower as compared to studies of similarly aged women without ovarian cancer (p-value = 0. 037) (249, 250). Other studies have found *Lactobacillus* to be less abundant in Black and Hispanic women (251) and our study consisted of 39 women who

self-reported ethnicity as white, two self-reported as Black, and four self-reported as Native American (Table 4 - 1). Each of the Black and Native American women had a non-*Lactobacillus* dominated vaginal microbiome (Supplementary Table C - 1) but ethnicity was not a statistically significant determinant of *Lactobacillus*-dominance (log OR = -1.63, 95% CI -4.62 – 1.30, p-value = 0.27). High microbial cell density, as gauged through qPCR with a standard curve generated from controls with known cell density, was positively correlated with vaginal *Lactobacillus*-dominance, although somewhat weakly (R^2 = 0.242, Supplementary Figure C - 4). Consumption of antibiotics within the past month was associated with a lack of *Lactobacillus*-dominance (Figure 4 - 2B); however, this relationship was not significant (p-value = 0.0515) and there were no other significant associations between *Lactobacillus*-dominance and any health or lifestyle factors, including PFI (Figure 4 - 2B). Only 20% (1 of 5) of patients with benign pathology had a *Lactobacillus*-dominated microbiome but small sample size prohibits statistical inference.

Previous studies have indicated that different *Lactobacillus* species in the vaginal microbiome may have different roles and differential influence on host biology (252-254). In our study, we identified *L. iners, L. brevis, L. mucosae, L. reuteri, L. zeae,* and *L. delbrueckii* in the vaginal microbiome; however, over 99% of the *Lactobacillus* reads in our study were either unclassified at the species level or mapped to *L. iners*. In individuals with high *Lactobacillus* abundance (n = 11), we found *L. iners* at significantly higher abundance in patients with either no gross residual disease or residual disease < 1cm (n = 7), compared to patients with residual disease > 1cm (n = 4, p-value = 0.0359, Figure 4 - 3).

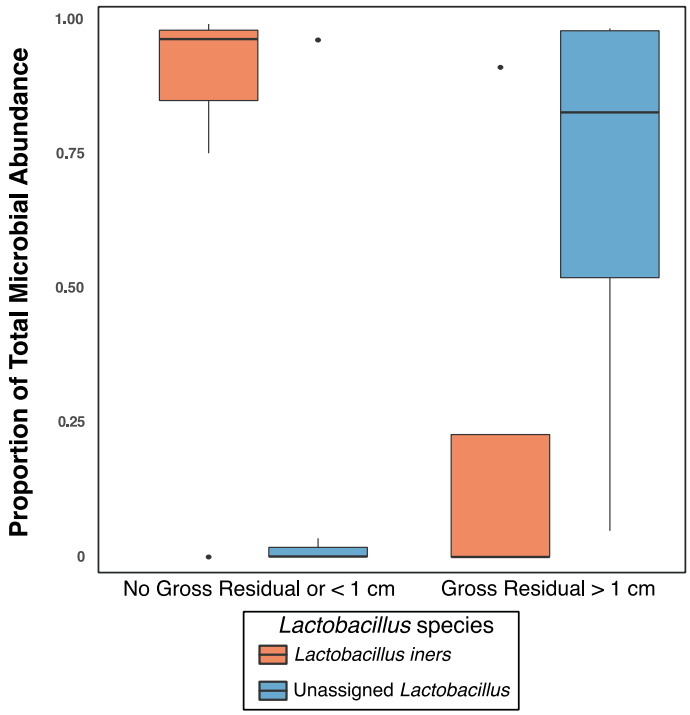


Figure 4 - 3: Lactobacillus iners dominates in small gross residual disease.

In individuals with Lactobacillus dominated vaginal microbiomes, L. iners was at significantly higher relative abundance (p-value = 0.0359) in patients with no gross residual disease or residual disease under 1 cm.

Prevotella was less common in those with residual disease > 1cm and those with a history of hormonal disease (Figure 4 – 2C), while a highly diverse vaginal microbiota was more common in women over 60 years old and in patients with a history of hormonal disorders, such as thyroid disease (Figure 4 – 2D); however, none of these associations were statistically significant (p-value > 0.05).

Gut microbiome

The gut microbiome was colonized by typical members of the gut microbiome at the phylum (Bacteroidetes, Firmicutes, Proteobacteria) and genus (*Bacteroides, Akkermansiac, Faecalibacterium, Ruminococcus,* and *Prevotella*) levels (Supplementary Figure C - 5A-B). The

PFI > 24 months and PFI < 6 months groups were not significantly different with respect to fecal unweighted and weighted UniFrac beta diversity distances as tested through PERMANOVA (Figure 4 - 4, Supplementary Figure C - 6A-C); however, similar to the vaginal microbiome samples, there was a small subset of individuals (n = 9) with a unique microbiome signature. Patients with PFI < 6 months individuals had higher odds (log OR = 1.85, 95% CI 0.85 4.497, p-value = 0.12) of being in this unique/outlier fecal microbiome group. These fecal outliers have significantly lower phylogenetic diversity compared to the other fecal samples (p-value = 0.001, Figure 4 - 5A) and have increased abundance of genera belonging to the order Clostridiales (*Lachnospira,* unidentified Ruminococceae genus, and *Subdoligranulum*) (Supplementary Figure C - 7). Of the six platinum-resistant patients in this outlier subset, two also had *Escherichia*-dominated vaginal communities. Eight of the nine patients in this subgroup reported consuming antibiotics within the past 6 months but this was not significantly different compared to the other individuals in this study (p-value = 0.49).

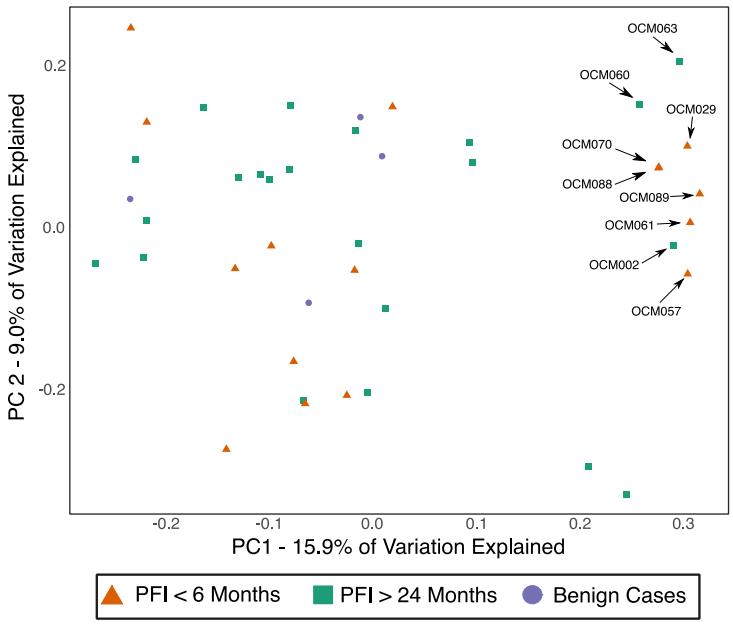


Figure 4 - 4: Unweighted UniFrac distances (PC1 and PC2) of fecal microbiomes.

Each shape represents a single sample and shapes clustering together have similar gut microbiome taxonomic composition. There was no significant difference in overall microbiome community structure between sample groups (PERMANOVA p-value > 0.05); however, there are nine samples (6 PFI < 6 months, 3 PFI > 24 months) that form an outlier group along PC1 (Figures A,B). These individuals are labelled.

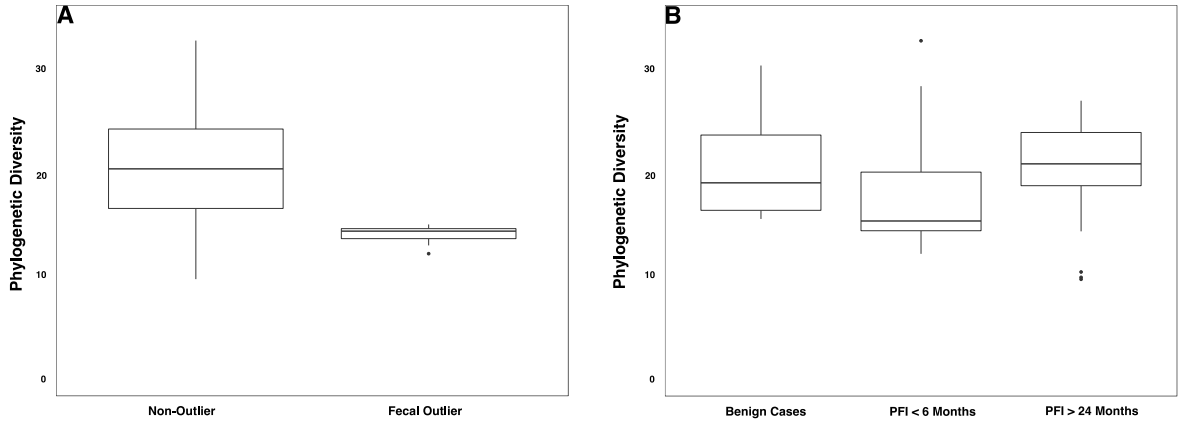


Figure 4 - 5A-B: Phylogenetic diversity in fecal microbiomes.

A) Samples that formed the outlier group in Figure 4A-B had lower phylogenetic diversity compared to the remainder of gut microbiome samples. B) Patients with PFI < 6 months had lower phylogenetic diversity than benign and platinum-sensitive patients.

Overall, patients with platinum-resistant disease had lower phylogenetic diversity compared to patients with platinum-sensitive disease but the difference was not significant (Figure 4 - 5B, p-value = 0.18). Regardless of PFI, patients with ovarian cancer had significantly higher relative abundance of *Prevotella* in the gut microbiome compared to benign individuals (p-value = 0.028, Supplementary Figure C - 8). Outside of the above-mentioned associations, there were no other significant associations between the gut microbiome and platinum sensitivity or other health/lifestyle variables.

4.5 DISCUSSION

One major findings of this study is the inverse relationship between a *Lactobacillus*-dominant vaginal microbiome and ovarian cancer. In our study, fewer women than expected (24.4%) have *Lactobacillus*-dominated vaginal communities compared to similarly aged, healthy women (47.2%) from other studies (p-value = 0.037) (249, 250). *Lactobacillus* dominance is not found in any of the Black (n = 2) nor Native American (n = 4) women in our study but only 11 of the 39 white women have *Lactobacillus*-dominated vaginal microbiomes; therefore, ethnicity was not a driving factor in *Lactobacillus* abundance in this study (p-value = 0.27). These results suggest the possibility that the low abundance of *Lactobacillus* may be indicative of a broad relationship between ovarian cancer and the vaginal microbiome. This finding corroborates a previous study that also observed a reduced frequency of the *Lactobacillus*-dominated vaginal microbiome in women with ovarian cancer, particularly in women under 50 (250). A partial explanation may be that *Lactobacillus*-dominance, while typically viewed as a healthy state in the female genital tract, is a non-resilient ecology prone to disruption by variable factors, including changes in glycogen availability (255, 256), antibiotic exposure (257), and shifts in hormone abundance induced during stress responses (258). While we observed no statistical relationships between *Lactobacillus* dominance and platinum-sensitivity or other lifestyle/medical variables, there was a nearly significant decrease in *Lactobacillus*-dominance associated with taking antibiotics within the past month, as well as a positive association between *Lactobacillus*-dominance and microbial cell density.

Lactobacillus maintains a low pH in the vaginal environment by producing lactic acid as a byproduct of glycogen metabolism and this low pH inhibits growth of pro-inflammatory bacteria

(232, 233, 259). Vaginal *Lactobacillus* may protect from gynecological cancers by inhibiting pro-inflammatory bacteria, such as those implicated in pelvic inflammatory disease, and by reducing inflammatory cytokines IL-1β and IL-6 (260). The low *Lactobacillus* levels we observed may be related to glycogen availability - all women, except one, in this study were postmenopausal, and vaginally produced glycogen is known to decrease after menopause; likewise chemotherapy can inhibit ovarian estrogen production and result in lower glycogen levels (249, 259, 261, 262). More importantly, many women in this study have had at least one ovary surgically removed during initial cancer treatment. Ovarian removal leads decreased estrogen production and thus a likely decrease in vaginal glycogen levels; however, more research is needed to fully explain the ovarian-estrogen-glycogen dynamic (255). Glycogen abundance may also help explain the relationship between vaginal microbiome cell density and *Lactobacillus,* as widely-available glycogen may encourage a densely-colonized *Lactobacillus* vaginal community due to high nutrient availability (262). However, we did not document glycogen levels in our study. The retrospective nature of our study means that we were unable to assess *Lactobacillus* levels in women with ovarian cancer before they progressed to stage III/IV, or prior to chemotherapy. Therefore, we could not investigate anti-gynecological cancer properties of vaginal *Lactobacillus*; nevertheless, by comparing patients in this study to similarly aged women without ovarian cancer, we present further evidence that low *Lactobacillus* levels are more common in women with ovarian cancer.

The presence, and size, of residual disease is strongly correlated with decreased survivability in ovarian cancer (219) and therefore our finding that *L. inters* was at significantly higher abundance in patients with either no gross residual disease or residual disease < 1cm after

treatment may point to *L. iners* as a potential path toward a biomarker. *L. iners* is a common vaginal bacterium (251, 263, 264) but its role in health and disease is sometimes contradictory (232); *L. iners* has been found at high relative abundance in low-grade squamous intraepithelial lesions in the cervix but at low abundance in high-grade squamous intraepithelial lesions (237). Yet, another study found *L. iners* at high abundance in women with normal cytology when compared to women with squamous intraepithelial lesions (265). Other studies have found *L. iners* to be positively associated with cervical cancer (266, 267) but *L. iners* is also linked to clearance of HPV (268), which is a causative agent of cervical cancer. Further research is necessary to better understand the role of *L. iners* in gynecological cancers in general, and in the potential inhibition of gross residual disease in ovarian cancer.

Lactobacillus is known to inhibit colonization and growth of *Escherichia* in the vaginal microbiome (269, 270). The low abundance of *Lactobacillus* found in our study may present ample opportunity for typically low abundance vaginal bacteria, such as *Escherichia*, to thrive and proliferate in the absence of competition. Overgrowth of *Escherichia* only occurred in 17.8% of patients in our study, yet 75% of those patients were in the PFI < 6 months group. To put it another way, 35.3% of patients with PFI < 6 months had *Escherichia* at greater than 20% relative abundance, compared to 4.34% of PFI > 24 months and only 1 of 5 benign cases. The explanation for why *Escherichia* was significantly more common in patients with platinum-resistant tumors is unclear, and because this study was retrospective, we were unable to track *Escherichia* abundance before we knew each patient's platinum-sensitivity. One possible pathway is via interactions between the microbiome, immune system, and how platinum-based chemotherapies induce cancer cell death. Platinum chemotherapies partially rely on ROS

produced by host myeloid cells (227). Microbes strongly influence immune system function, and hence, alteration in ROS production may be more common in *Escherichia*-dominant vaginal microbiomes, which may render platinum-based chemotherapies less effective, leading to platinum-resistance. Vaginal *Escherichia* may also cause an increased inflammatory response, such as during pelvic inflammatory disease (271), and promote cancerous growth, resulting in a shortened PFI. The effect of *Escherichia* on platinum-sensitivity warrants further investigation.

We also observed differences in the gut microbiome of women with platinum-resistant tumors, compared to benign and platinum-sensitive. Similar to the prevalence of *Escherichia*-dominant vaginal microbiomes, 35.3% of patients with PFI < 6 months were fecal beta-diversity outliers compared to the remainder of the fecal samples, while only 13.0% of PFI > 24 and none of the benign cases fell into this cluster. Genera belonging to the Clostridiales order (*Subdoligranulum* and *Lachnospira)* were at higher abundance in this group of outlier samples. *Subdoligranulum,* has been found at high abundance in the stool of individuals with gastrointestinal neoplasms (272) and has a positive association with blood-based markers of inflammation (273). *Lachnospira* has been found to be positively correlated with a plant-based diet (274) and at high abundance in healthy controls in a study investigating chronic kidney disease (275); yet *Lachnospira* is also found at high abundance in women with metabolic disorder and obesity (274). Additionally, we note lower phylogenetic diversity in the gut microbiome of platinum-resistant tumors compared to both platinum-sensitive and benign tumors. Chemotherapy is a well-documented driver of decreased gut microbiome alpha diversity (276-278); however, we did not observe a shift in alpha diversity with time since last cycle of chemotherapy. High gut microbiome alpha diversity is typically associated with improved human health (7, 279),

although recent studies have started to question this paradigm (280). Nevertheless, the unique gut microbiome beta diversity profile in a subset of PFI < 6 months, and the decreased alpha diversity in the full PFI < 6 months population indicates that there may be long term relationship between platinum-resistance and the gut microbiome.

Prevotella is enriched in both the platinum-resistant and platinum-sensitive study groups, compared to the benign group. *Prevotella* is typically only found at high abundance in the gut microbiomes of non-industrial, traditional populations (281) and it is associated with consumption of a plant-rich diet (281). Yet, some strains of *Prevotella* are found in the gut microbiomes of industrial populations and are linked to pro-inflammatory states in the gut microbiome (282, 283). Similar to decreased *Lactobacillus* in the vaginal microbiome of patients with ovarian cancer, the relatively high abundance of *Prevotella* in the gut microbiome of women with ovarian cancer indicates a notable shift in microbial composition. Studies with larger control groups are necessary to address this relationship.

4.6 CONCLUSIONS

Our results demonstrate an association between the vaginal and gut microbiomes and platinum-sensitivity in women with ovarian cancer. *Escherichia*-dominant vaginal communities are significantly more likely to be present in patients with platinum-resistant tumors but the explanatory mechanism for this relationship is currently unclear. Lab and/or collection contamination does not appear to play a role in vaginal *Escherichia* abundance, which indicates that finding *Escherichia* at high relative abundance in patients with PFI < 6 months is a biological trend.

We also observed shared vaginal and gut microbiome profiles in women with ovarian cancer, with decreased dominance of *Lactobacillus* and increased relative abundance of *Prevotella,* respectively, regardless of platinum-sensitivity. These results suggest shifts in microbiome composition that are related to the ovarian cancer disease state, which may possibly be related to chemotherapy, but the retrospective nature of our study does not allow us to distinguish the exact mechanism of action.

Our results call for deeper investigation into the relationship between the vaginal and gut microbiomes and ovarian cancer. A future avenue for research is a prospective, longitudinal study that tracks how the vaginal and gut microbiomes change throughout the course of ovarian cancer therapy, with an aim to disentangle how *Escherichia*-abundance impacts response to chemotherapy. Similarly, a study tracking *Lactobacillus* abundance in an aged-matched, lifestyle-matched cohort of women with and without ovarian cancer may provide insights into how microbial risk factors impact occurrence and outcomes of ovarian cancer. This work must also investigate why high *L. iners* abundance is found nearly exclusively in cases with < 1 cm or no gross residual disease, while other *Lactobacillus* species are found with > 1 cm residual disease. Finally, ovarian cancer microbiome research also presents an opportunity for microbial metagenomics and metabolomics to provide a fuller picture of the vaginal and gut microbiome ecosystems in health and disease.

Chapter 5 – Conclusions

The research presented in this dissertation is connected by a common theme of applying ecological perspectives to current impactful topics in microbiome research: diversity, function, and health association. Microbiomes are diverse ecosystems with dynamic interactions between bacteria and the three studies presented in this dissertation highlight how ecological study of human microbiomes provides a deeper understanding of microbiomes and human biology. Importantly, this dissertation indicates that numerous types of human microbiome samples, whether they be ancient (Chapter 2), modern metagenomes (Chapter 3), or amplicon (Chapter 4), can be used for ecological analysis and present unique and novel insights into human biological variation.

In Chapter 2 “Functional Diversity of Microbial Ecologies Estimated from Ancient Human Coprolites and Dental Calculus”, both dental calculus and coprolites were analyzed using network analysis and functional redundancy approaches. These techniques revealed similarites between ancient and modern dental calculus with respect to network properties and the types of keystone taxa identified. The Nuragic Sardinian ancient dental calculus was unique from the other dental calculus, which may be due to certain lifestyle variables, like exposure to copper. Furthermore, Chapter 2 demonstrated shared functional redundancy profiles and keystone taxa between ancient coprolites and modern hunter-gatherers. Taken together, Chapter 2 shows that ancient microbiome samples can be used for ecological analysis and these novel approaches yield valuable insights into microbiome change over time.

Chapter 3 – "Non-Industrial Gut Microbiomes Provide a More Resilient Ecology for Short-Chain Fatty Acid Production" compares newly generated fecal microbiome metagenome data from Burkina Faso to a panel of industrial and non-industrial gut micorbiomes in an attempt to outline how lifestyle influences resilience in short-chain fatty acid production. Short-chain fatty acids (SCFAs) are a key component of human-microbiome interactions in the GI tract and higher levels of SCFAs have been linked to improved health. In Chapter 3, the taxa encoding SCFAs were found to be more diverse and evenly distributed in non-industrial populations at the genus level; however, there is a steep drop-off in diversity at the species level in non-industrial populations. The relative inabilty to attribute SCFA synthesis genes to species in non-industrial populations (and to a lesser degree at the genus level as well) is due to a lack of microbiome metagenome datasets from non-industrial datasets in existing databases, which can be directly linked to non-industrial populations typically being a low research priority in the genomics research community. While Chapter 3 shows strong SCFA resilience in non-industrial populations, there is still a long way to go before we can have confidence that we are capturing the full range of human microbiome diversity.

In Chapter 4, "Shifts in Gut and Vaginal Microbiomes Associated with Platinum-Free Interval Length in Women with Ovarian Cance", *Lactobacillus* dominated vaginal communities is observed to be less common in women with ovarian cancer compared to similarly aged healthy women. A missing keystone species, like *Lactobacillus*, allows for opportunistic bacteria to flourish in the community and *Escherichia* was found to be more likely to dominate vaginal microbiomes of women with a Platinum-Free Interval (PFI) length of less than 6 months, compared to women with a PFI > 24 months in Chapter 4. A short PFI length often leads to

death more quickly than a long PFI length and hence finding ecological biomarkers of PFI length has importance in survivability. Vaginal *Escherichia* is often associated with infections and inflammation and while we could not pin-down the exact cause for the higher chance of *Escherichia* in PFI<6 months indviduals the absence of *Escherichia* from PFI>24 months individuals indicates some relationship between PFI length and vaginal microbiome community make-up.

This dissertation covers three distinct approaches/techniques to provide a better understanding of ecological interactions in the human microbiome. A clear future direction is in ecological analysis of ancient human microbiomes, which had never been done before our analysis in Chapter 2. Ancient DNA presents many challenges to researchers, but this should not prohibit theoretical advancements in the field. Ecological analysis of coprolites will be much more difficult due to the genuine lack of samples found at a single site and difficulty in recovering high quality DNA. Dental calculus, on the other hand, presents ample opportunity for ecological analysis because it can typically be linked back to an individual, is found in many individuals in museums/archaeological sites, and it yields high quality and high biomass DNA.

Studying resilience in contemporary non-industrial gut microbiomes is valuable because it allows for a more complete picture of human gut microbiome diversity. An obvious next step to ecological analysis of human gut metagenomes is to study resilience in antibiotic resistant genes in diverse populations. Populations with little or no antibiotic exposure have already been found to harbor antibiotic resistance, which should not be a surprise because antibiotic resistance is a natural phenomenon. However, it is vital to track ecological diversity in antibiotic resistance in non-industrial contexts to identify bacteria at high risk for developing multidrug resistance.

Finally, clinical studies can use ecological approaches, particularly in longitudinal studies, to evaluate how the microbiome responds in different contexts and identify potential keystone taxa that mark recovery and/or disease progression. Microbiomes are increasingly seen as important players in pharmaceutical efficacy and ecological analysis of microbiome shifts in a clinical setting has the potential to drastically improve health outcomes. Understanding the full picture of the human microbiome necessitates an ecological approach and this dissertation demonstrates the valuable insights gained from applying different ecological techniques to a wide variety of sample types.

Bibliography

1. Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. Current understanding of the human microbiome. Nature medicine. 2018;24(4):392.
2. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciolek T, McCall L-I, McDonald D. Best practices for analysing microbiomes. Nature Reviews Microbiology. 2018:1.

3. Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. Obesity alters gut microbial ecology. Proc Natl Acad Sci U S A. 2005;102(31):11070-5. Epub 2005/07/22. doi: 10.1073/pnas.0504978102. PubMed PMID: 16033867; PMCID: PMC1176910.
4. Mueller NT, Shin H, Pizoni A, Werlang IC, Matte U, Goldani MZ, Goldani HA, Dominguez-Bello MG. Birth mode-dependent association between pre-pregnancy maternal weight status and the neonatal intestinal microbiome. Scientific reports. 2016;6:23133.
5. Schwabe RF, Jobin C. The microbiome and cancer. Nat Rev Cancer. 2013;13(11):800-12. Epub 2013/10/18. doi: 10.1038/nrc3610. PubMed PMID: 24132111; PMCID: PMC3986062.
6. Turnbaugh PJ, Bäckhed F, Fulton L, Gordon JI. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell host & microbe. 2008;3(4):213-23.
7. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI. A core gut microbiome in obese and lean twins. Nature. 2009;457(7228):480-4. Epub 2008/12/02. doi: 10.1038/nature07540. PubMed PMID: 19043404; PMCID: PMC2677729.
8. Moya A, Ferrer M. Functional redundancy-induced stability of gut microbiota subjected to disturbance. Trends in microbiology. 2016;24(5):402-13.
9. Sommer F, Anderson JM, Bharti R, Raes J, Rosenstiel P. The resilience of the intestinal microbiota influences health and disease. Nature Reviews Microbiology. 2017;15(10):630.
10. Bäckhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, Li Y, Xia Y, Xie H, Zhong H. Dynamics and stabilization of the human gut microbiome during the first year of life. Cell host & microbe. 2015;17(5):690-703.
11. Guittar J, Shade A, Litchman E. Trait-based community assembly and succession of the infant gut microbiome. Nat Commun. 2019;10(1):512. Epub 2019/02/03. doi: 10.1038/s41467-019-08377-w. PubMed PMID: 30710083; PMCID: PMC6358638.
12. McNally L, Brown SP. Microbiome: Ecology of stable gut communities. Nat Microbiol. 2016;1(1):15016. Epub 2016/08/31. doi: 10.1038/nmicrobiol.2015.16. PubMed PMID: 27571760.
13. Shaw LP, Bassam H, Barnes CP, Walker AS, Klein N, Balloux F. Modelling microbiome recovery after antibiotics using a stability landscape framework. The ISME journal. 2019;13(7):1845-56.
14. Davenport ER, Sanders JG, Song SJ, Amato KR, Clark AG, Knight R. The human microbiome in evolution. BMC biology. 2017;15(1):127.
15. Moeller AH, Caro-Quintero A, Mjungu D, Georgiev AV, Lonsdorf EV, Muller MN, Pusey AE, Peeters M, Hahn BH, Ochman H. Cospeciation of gut microbiota with hominids. Science. 2016;353(6297):380-2.
16. Groussin M, Mazel F, Sanders JG, Smillie CS, Lavergne S, Thuiller W, Alm EJ. Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. Nature communications. 2017;8:14319.
17. Obregon-Tito AJ, Tito RY, Metcalf J, Sankaranarayanan K, Clemente JC, Ursell LK, Zech Xu Z, Van Treuren W, Knight R, Gaffney PM, Spicer P, Lawson P, Marin-Reyes L, Trujillo-Villarroel O, Foster M, Guija-Poma E, Troncoso-Corzo L, Warinner C, Ozga AT, Lewis CM. Subsistence strategies in traditional societies distinguish gut microbiomes. Nat Commun. 2015;6:6505. Epub 2015/03/26. doi: 10.1038/ncomms7505. PubMed PMID: 25807110; PMCID: PMC4386023.

18. Warinner C, Herbig A, Mann A, Fellows Yates JA, Weiß CL, Burbano HA, Orlando L, Krause J. A robust framework for microbial archaeology. Annual review of genomics and human genetics. 2017;18:321-56.
19. Clemente JC, Pehrsson EC, Blaser MJ, Sandhu K, Gao Z, Wang B, Magris M, Hidalgo G, Contreras M, Noya-Alarcon O, Lander O, McDonald J, Cox M, Walter J, Oh PL, Ruiz JF, Rodriguez S, Shen N, Song SJ, Metcalf J, Knight R, Dantas G, Dominguez-Bello MG. The microbiome of uncontacted Amerindians. Sci Adv. 2015;1(3):e1500183. Epub 2015/08/01. doi: 10.1126/sciadv.1500183. PubMed PMID: 26229982; PMCID: PMC4517851.
20. Gomez A, Petrzelkova KJ, Burns MB, Yeoman CJ, Amato KR, Vlckova K, Modry D, Todd A, Jost Robinson CA, Remis MJ, Torralba MG, Morton E, Umana JD, Carbonero F, Gaskins HR, Nelson KE, Wilson BA, Stumpf RM, White BA, Leigh SR, Blekhman R. Gut Microbiome of Coexisting BaAka Pygmies and Bantu Reflects Gradients of Traditional Subsistence Patterns. Cell Rep. 2016;14(9):2142-53. Epub 2016/03/01. doi: 10.1016/j.celrep.2016.02.013. PubMed PMID: 26923597.
21. Zhang J, Guo Z, Lim AAQ, Zheng Y, Koh EY, Ho D, Qiao J, Huo D, Hou Q, Huang W. Mongolians core gut microbiota and its correlation with seasonal dietary changes. Scientific reports. 2014;4:5001.
22. Moeller AH. The shrinking human gut microbiome. Current opinion in microbiology. 2017;38:30-5.
23. Sonnenburg ED, Sonnenburg JL. The ancestral and industrialized gut microbiota and implications for human health. Nat Rev Microbiol. 2019;17(6):383-90. Epub 2019/05/16. doi: 10.1038/s41579-019-0191-8. PubMed PMID: 31089293.
24. Rampelli S, Schnorr SL, Consolandi C, Turroni S, Severgnini M, Peano C, Brigidi P, Crittenden AN, Henry AG, Candela M. Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota. Curr Biol. 2015;25(13):1682-93. Epub 2015/05/20. doi: 10.1016/j.cub.2015.04.055. PubMed PMID: 25981789.
25. Ruiz-Calderon JF, Cavallin H, Song SJ, Novoselac A, Pericchi LR, Hernandez JN, Rios R, Branch OH, Pereira H, Paulino LC, Blaser MJ, Knight R, Dominguez-Bello MG. Walls talk: Microbial biogeography of homes spanning urbanization. Sci Adv. 2016;2(2):e1501061. Epub 2016/03/05. doi: 10.1126/sciadv.1501061. PubMed PMID: 26933683; PMCID: PMC4758746.
26. Fragiadakis GK, Smits SA, Sonnenburg ED, Van Treuren W, Reid G, Knight R, Manjurano A, Changalucha J, Dominguez-Bello MG, Leach J. Links between environment, diet, and the hunter-gatherer microbiome. Gut Microbes. 2019;10(2):216-27.
27. Bokulich NA, Chung J, Battaglia T, Henderson N, Jay M, Li H, Lieber AD, Wu F, Perez-Perez GI, Chen Y. Antibiotics, birth mode, and diet shape microbiome maturation during early life. Science translational medicine. 2016;8(343):343ra82-ra82.
28. Fujimura KE, Demoor T, Rauch M, Faruqi AA, Jang S, Johnson CC, Boushey HA, Zoratti E, Ownby D, Lukacs NW. House dust exposure mediates gut microbiome Lactobacillus enrichment and airway immune defense against allergens and virus infection. Proceedings of the National Academy of Sciences. 2014;111(2):805-10.
29. Song SJ, Lauber C, Costello EK, Lozupone CA, Humphrey G, Berg-Lyons D, Caporaso JG, Knights D, Clemente JC, Nakielny S. Cohabiting family members share microbiota with one another and with their dogs. elife. 2013;2:e00458.

30. Tun HM, Konya T, Takaro TK, Brook JR, Chari R, Field CJ, Guttman DS, Becker AB, Mandhane PJ, Turvey SE. Exposure to household furry pets influences the gut microbiota of infants at 3–4 months following various birth scenarios. Microbiome. 2017;5(1):40.
31. Rosier B, Marsh P, Mira A. Resilience of the oral microbiota in health: mechanisms that prevent dysbiosis. Journal of dental research. 2018;97(4):371-80.
32. Koskella B, Hall LJ, Metcalf CJE. The microbiome beyond the horizon of ecological and evolutionary theory. Nature ecology & evolution. 2017;1(11):1606.
33. Maier L, Pruteanu M, Kuhn M, Zeller G, Telzerow A, Anderson EE, Brochado AR, Fernandez KC, Dose H, Mori H. Extensive impact of non-antibiotic drugs on human gut bacteria. Nature. 2018;555(7698):623.
34. Shade A, Dunn RR, Blowes SA, Keil P, Bohannan BJ, Herrmann M, Küsel K, Lennon JT, Sanders NJ, Storch D. Macroecology to unite all life, large and small. Trends in ecology & evolution. 2018.
35. Moya A, Ferrer M. Functional Redundancy-Induced Stability of Gut Microbiota Subjected to Disturbance. Trends Microbiol. 2016;24(5):402-13. Epub 2016/03/22. doi: 10.1016/j.tim.2016.02.002. PubMed PMID: 26996765.
36. Stein RR, Bucci V, Toussaint NC, Buffie CG, Rätsch G, Pamer EG, Sander C, Xavier JB. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. PLoS computational biology. 2013;9(12):e1003388.
37. Teles FR, Teles RP, Uzel NG, Song XQ, Torresyap G, Socransky SS, Haffajee AD. Early microbial succession in redeveloping dental biofilms in periodontal health and disease. J Periodontal Res. 2012;47(1):95-104. Epub 2011/09/08. doi: 10.1111/j.1600-0765.2011.01409.x. PubMed PMID: 21895662; PMCID: PMC3253172.
38. Banerjee S, Schlaeppi K, van der Heijden MG. Keystone taxa as drivers of microbiome structure and functioning. Nature Reviews Microbiology. 2018;16(9):567-76.
39. Layeghifard M, Hwang DM, Guttman DS. Disentangling Interactions in the Microbiome: A Network Perspective. Trends Microbiol. 2017;25(3):217-28. Epub 2016/12/06. doi: 10.1016/j.tim.2016.11.008. PubMed PMID: 27916383; PMCID: PMC7172547.
40. Hodgson D, McDonald JL, Hosken DJ. What do you mean,'resilient'? Trends in ecology & evolution. 2015;30(9):503-6.
41. Francino MP. Antibiotics and the Human Gut Microbiome: Dysbioses and Accumulation of Resistances. Front Microbiol. 2015;6:1543. Epub 2016/01/23. doi: 10.3389/fmicb.2015.01543. PubMed PMID: 26793178; PMCID: PMC4709861.
42. Francino MP, Moya A. Effects of antibiotic use on the microbiota of the gut and associated alterations of immunity and metabolism. EMJ Gastroenterol. 2013;1:74-80.
43. Gomez-Arango LF, Barrett HL, McIntyre HD, Callaway LK, Morrison M, Nitert MD. Antibiotic treatment at delivery shapes the initial oral microbiome in neonates. Scientific reports. 2017;7:43481.
44. Lewis JD, Chen EZ, Baldassano RN, Otley AR, Griffiths AM, Lee D, Bittinger K, Bailey A, Friedman ES, Hoffmann C. Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric Crohn's disease. Cell host & microbe. 2015;18(4):489-500.
45. Knoop KA, McDonald KG, Kulkarni DH, Newberry RD. Antibiotics promote inflammation through the translocation of native commensal colonic bacteria. Gut. 2016;65(7):1100-9.

46. Chang JY, Antonopoulos DA, Kalra A, Tonelli A, Khalife WT, Schmidt TM, Young VB. Decreased diversity of the fecal Microbiome in recurrent Clostridium difficile-associated diarrhea. J Infect Dis. 2008;197(3):435-8. Epub 2008/01/18. doi: 10.1086/525047. PubMed PMID: 18199029.
47. Steinway SN, Biggs MB, Loughran TP, Jr., Papin JA, Albert R. Inference of Network Dynamics and Metabolic Interactions in the Gut Microbiome. PLoS Comput Biol. 2015;11(5):e1004338. Epub 2015/06/24. doi: 10.1371/journal.pcbi.1004338. PubMed PMID: 26102287; PMCID: PMC4478025.
48. Theriot CM, Koenigsknecht MJ, Carlson Jr PE, Hatton GE, Nelson AM, Li B, Huffnagle GB, Li JZ, Young VB. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection. Nature communications. 2014;5:3114.
49. Ozaki E, Kato H, Kita H, Karasawa T, Maegawa T, Koino Y, Matsumoto K, Takada T, Nomoto K, Tanaka R, Nakamura S. Clostridium difficile colonization in healthy adults: transient colonization and correlation with enterococcal colonization. J Med Microbiol. 2004;53(Pt 2):167-72. Epub 2004/01/20. doi: 10.1099/jmm.0.05376-0. PubMed PMID: 14729940.
50. Galdys AL, Nelson JS, Shutt KA, Schlackman JL, Pakstis DL, Pasculle AW, Marsh JW, Harrison LH, Curry SR. Prevalence and duration of asymptomatic Clostridium difficile carriage among healthy subjects in Pittsburgh, Pennsylvania. Journal of clinical microbiology. 2014;52(7):2406-9.
51. Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gobourne A, No D, Liu H, Kinnebrew M, Viale A, Littmann E, van den Brink MR, Jenq RR, Taur Y, Sander C, Cross JR, Toussaint NC, Xavier JB, Pamer EG. Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. Nature. 2015;517(7533):205-8. Epub 2014/10/23. doi: 10.1038/nature13828. PubMed PMID: 25337874; PMCID: PMC4354891.
52. Theriot CM, Young VB. Interactions between the gastrointestinal microbiome and Clostridium difficile. Annual review of microbiology. 2015;69:445-61.
53. Curtis MA, Zenobia C, Darveau RP. The relationship of the oral microbiotia to periodontal health and disease. Cell Host Microbe. 2011;10(4):302-6. Epub 2011/10/25. doi: 10.1016/j.chom.2011.09.008. PubMed PMID: 22018230; PMCID: PMC3216488.
54. Hajishengallis G, Lamont RJ. Beyond the red complex and into more complexity: the polymicrobial synergy and dysbiosis (PSD) model of periodontal disease etiology. Molecular oral microbiology. 2012;27(6):409-19.
55. Elmqvist T, Folke C, Nyström M, Peterson G, Bengtsson J, Walker B, Norberg J. Response diversity, ecosystem change, and resilience. Frontiers in Ecology and the Environment. 2003;1(9):488-94.
56. Mori AS, Furukawa T, Sasaki T. Response diversity determines the resilience of ecosystems to environmental change. Biol Rev Camb Philos Soc. 2013;88(2):349-64. Epub 2012/12/12. doi: 10.1111/brv.12004. PubMed PMID: 23217173.
57. Costello EK, Stagaman K, Dethlefsen L, Bohannan BJ, Relman DA. The application of ecological theory toward an understanding of the human microbiome. Science. 2012;336(6086):1255-62. Epub 2012/06/08. doi: 10.1126/science.1224203. PubMed PMID: 22674335; PMCID: PMC4208626.
58. Miller ET, Bohannan BJ. Life between patches: incorporating microbiome biology alters the predictions of metacommunity models. Frontiers in Ecology and Evolution. 2019;7:276.

59. Miller ET, Svanbäck R, Bohannan BJ. Microbiomes as metacommunities: understanding host-associated microbes through metacommunity ecology. Trends in ecology & evolution. 2018;33(12):926-35.
60. Bokulich NA, Chung J, Battaglia T, Henderson N, Jay M, Li H, A DL, Wu F, Perez-Perez GI, Chen Y, Schweizer W, Zheng X, Contreras M, Dominguez-Bello MG, Blaser MJ. Antibiotics, birth mode, and diet shape microbiome maturation during early life. Sci Transl Med. 2016;8(343):343ra82. Epub 2016/06/17. doi: 10.1126/scitranslmed.aad7121. PubMed PMID: 27306664; PMCID: PMC5308924.
61. Wampach L, Heintz-Buschart A, Fritz JV, Ramiro-Garcia J, Habier J, Herold M, Narayanasamy S, Kaysen A, Hogan AH, Bindl L. Birth mode is associated with earliest strain-conferred gut microbiome functions and immunostimulatory potential. Nature communications. 2018;9(1):1-14.
62. Jin Y, Yip HK. Supragingival calculus: formation and control. Crit Rev Oral Biol Med. 2002;13(5):426-41. Epub 2002/10/24. doi: 10.1177/154411130201300506. PubMed PMID: 12393761.
63. White DJ. Dental calculus: recent insights into occurrence, formation, prevention, removal and oral health effects of supragingival and subgingival deposits. European journal of oral sciences. 1997;105(5):508-22.
64. Kolenbrander PE, London J. Adhere today, here tomorrow: oral bacterial adherence. Journal of bacteriology. 1993;175(11):3247.
65. Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent RL, Jr. Microbial complexes in subgingival plaque. J Clin Periodontol. 1998;25(2):134-44. Epub 1998/03/12. doi: 10.1111/j.1600-051x.1998.tb02419.x. PubMed PMID: 9495612.
66. Hajishengallis G, Darveau RP, Curtis MA. The keystone-pathogen hypothesis. Nat Rev Microbiol. 2012;10(10):717-25. Epub 2012/09/04. doi: 10.1038/nrmicro2873. PubMed PMID: 22941505; PMCID: PMC3498498.
67. Olsen I, Lambris JD, Hajishengallis G. Porphyromonas gingivalis disturbs host-commensal homeostasis by changing complement function. J Oral Microbiol. 2017;9(1):1340085. Epub 2017/07/28. doi: 10.1080/20002297.2017.1340085. PubMed PMID: 28748042; PMCID: PMC5508361.
68. Salosensaari A, Laitinen V, Havulinna AS, Meric G, Cheng S, Perola M, Valsta L, Alfthan G, Inouye M, Watrous JD. Taxonomic Signatures of Long-Term Mortality Risk in Human Gut Microbiota. medRxiv. 2020:2019.12. 30.19015842.
69. Fortin D, Beyer HL, Boyce MS, Smith DW, Duchesne T, Mao JS. Wolves influence elk movements: behavior shapes a trophic cascade in Yellowstone National Park. Ecology. 2005;86(5):1320-30.
70. Ripple WJ, Beschta RL. Wolf reintroduction, predation risk, and cottonwood recovery in Yellowstone National Park. Forest Ecology and Management. 2003;184(1-3):299-313.
71. Ripple WJ, Beschta RL. Trophic cascades in Yellowstone: the first 15 years after wolf reintroduction. Biological Conservation. 2012;145(1):205-13.
72. Coyte KZ, Schluter J, Foster KR. The ecology of the microbiome: Networks, competition, and stability. Science. 2015;350(6261):663-6. Epub 2015/11/07. doi: 10.1126/science.aad2602. PubMed PMID: 26542567.

73. Bordenstein SR, Theis KR. Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. PLoS Biol. 2015;13(8):e1002226. Epub 2015/08/19. doi: 10.1371/journal.pbio.1002226. PubMed PMID: 26284777; PMCID: PMC4540581.
74. Rosenberg E, Zilber-Rosenberg I. Microbes Drive Evolution of Animals and Plants: the Hologenome Concept. mBio. 2016;7(2):e01395. Epub 2016/04/02. doi: 10.1128/mBio.01395-15. PubMed PMID: 27034283; PMCID: PMC4817260.
75. Roughgarden J, Gilbert SF, Rosenberg E, Zilber-Rosenberg I, Lloyd EA. Holobionts as units of selection and a model of their population dynamics and evolution. Biological Theory. 2018;13(1):44-65.
76. Simon JC, Marchesi JR, Mougel C, Selosse MA. Host-microbiota interactions: from holobiont theory to analysis. Microbiome. 2019;7(1):5. Epub 2019/01/13. doi: 10.1186/s40168-019-0619-4. PubMed PMID: 30635058; PMCID: PMC6330386.
77. Gilbert JA, Lynch SV. Community ecology as a framework for human microbiome research. Nature medicine. 2019:1.
78. Durack J, Kimes NE, Lin DL, Rauch M, McKean M, McCauley K, Panzer AR, Mar JS, Cabana MD, Lynch SV. Delayed gut microbiota development in high-risk for asthma infants is temporarily modifiable by Lactobacillus supplementation. Nature communications. 2018;9(1):1-9.
79. Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, Knights D, Gajer P, Ravel J, Fierer N, Gordon JI, Knight R. Moving pictures of the human microbiome. Genome Biol. 2011;12(5):R50. Epub 2011/06/01. doi: 10.1186/gb-2011-12-5-r50. PubMed PMID: 21624126; PMCID: PMC3271711.
80. Sinha R, Goedert JJ, Vogtmann E, Hua X, Porras C, Hayes R, Safaeian M, Yu G, Sampson J, Ahn J, Shi J. Quantification of Human Microbiome Stability Over 6 Months: Implications for Epidemiologic Studies. Am J Epidemiol. 2018;187(6):1282-90. Epub 2018/04/03. doi: 10.1093/aje/kwy064. PubMed PMID: 29608646; PMCID: PMC5982812.
81. Thaiss CA, Itav S, Rothschild D, Meijer MT, Levy M, Moresi C, Dohnalova L, Braverman S, Rozin S, Malitsky S, Dori-Bachash M, Kuperman Y, Biton I, Gertler A, Harmelin A, Shapiro H, Halpern Z, Aharoni A, Segal E, Elinav E. Persistent microbiome alterations modulate the rate of post-dieting weight regain. Nature. 2016;540(7634):544-51. Epub 2016/12/03. doi: 10.1038/nature20796. PubMed PMID: 27906159.
82. Fukami T, Nakajima M. Community assembly: alternative stable states or alternative transient states? Ecol Lett. 2011;14(10):973-84. Epub 2011/07/28. doi: 10.1111/j.1461-0248.2011.01663.x. PubMed PMID: 21790934; PMCID: PMC3187870.
83. Zaneveld JR, McMinds R, Vega Thurber R. Stress and stability: applying the Anna Karenina principle to animal microbiomes. Nat Microbiol. 2017;2(9):17121. Epub 2017/08/25. doi: 10.1038/nmicrobiol.2017.121. PubMed PMID: 28836573.
84. Zmora N, Zilberman-Schapira G, Suez J, Mor U, Dori-Bachash M, Bashiardes S, Kotler E, Zur M, Regev-Lehavi D, Brik RB, Federici S, Cohen Y, Linevsky R, Rothschild D, Moor AE, Ben-Moshe S, Harmelin A, Itzkovitz S, Maharshak N, Shibolet O, Shapiro H, Pevsner-Fischer M, Sharon I, Halpern Z, Segal E, Elinav E. Personalized Gut Mucosal Colonization Resistance to Empiric Probiotics Is Associated with Unique Host and Microbiome Features. Cell. 2018;174(6):1388-405 e21. Epub 2018/09/08. doi: 10.1016/j.cell.2018.08.041. PubMed PMID: 30193112.

85. Schröder A, Persson L, De Roos AM. Direct experimental evidence for alternative stable states: a review. Oikos. 2005;110(1):3-19.
86. Callen EO, Cameron T. A prehistoric diet revealed in coprolites. New Scientist. 1960;8(190):35-40.
87. Iniguez AM, Araujo A, Ferreira LF, Vicente AC. Analysis of ancient DNA from coprolites: a perspective with random amplified polymorphic DNA-polymerase chain reaction approach. Mem Inst Oswaldo Cruz. 2003;98 Suppl 1:63-5. Epub 2003/04/12. doi: 10.1590/s0074-02762003000900012. PubMed PMID: 12687765.
88. Reinhard KJ, Bryant VM. Coprolite analysis: a biological perspective on archaeology. Archaeological method and theory. 1992;4:245-88.
89. Tito RY, Knights D, Metcalf J, Obregon-Tito AJ, Cleeland L, Najar F, Roe B, Reinhard K, Sobolik K, Belknap S, Foster M, Spicer P, Knight R, Lewis CM, Jr. Insights from characterizing extinct human gut microbiomes. PLoS One. 2012;7(12):e51146. Epub 2012/12/20. doi: 10.1371/journal.pone.0051146. PubMed PMID: 23251439; PMCID: PMC3521025.
90. Warinner C, Speller C, Collins MJ, Lewis CM, Jr. Ancient human microbiomes. J Hum Evol. 2015;79:125-36. Epub 2015/01/07. doi: 10.1016/j.jhevol.2014.10.016. PubMed PMID: 25559298; PMCID: PMC4312737.
91. Brealey JC, Leitão HG, Xu W, Dalén L, Guschanski K. Dental calculus as a tool to study the evolution of the oral microbiome in mammals. bioRxiv. 2019:596791.
92. Warinner C, Hendy J, Speller C, Cappellini E, Fischer R, Trachsel C, Arneborg J, Lynnerup N, Craig OE, Swallow DM, Fotakis A, Christensen RJ, Olsen JV, Liebert A, Montalva N, Fiddyment S, Charlton S, Mackie M, Canci A, Bouwman A, Ruhli F, Gilbert MT, Collins MJ. Direct evidence of milk consumption from ancient human dental calculus. Sci Rep. 2014;4:7104. Epub 2014/11/28. doi: 10.1038/srep07104. PubMed PMID: 25429530; PMCID: PMC4245811.
93. Warinner C, Rodrigues JF, Vyas R, Trachsel C, Shved N, Grossmann J, Radini A, Hancock Y, Tito RY, Fiddyment S, Speller C, Hendy J, Charlton S, Luder HU, Salazar-Garcia DC, Eppler E, Seiler R, Hansen LH, Castruita JA, Barkow-Oesterreicher S, Teoh KY, Kelstrup CD, Olsen JV, Nanni P, Kawai T, Willerslev E, von Mering C, Lewis CM, Jr., Collins MJ, Gilbert MT, Ruhli F, Cappellini E. Pathogens and host immunity in the ancient human oral cavity. Nat Genet. 2014;46(4):336-44. Epub 2014/02/25. doi: 10.1038/ng.2906. PubMed PMID: 24562188; PMCID: PMC3969750.
94. Weyrich LS, Duchene S, Soubrier J, Arriola L, Llamas B, Breen J, Morris AG, Alt KW, Caramelli D, Dresely V, Farrell M, Farrer AG, Francken M, Gully N, Haak W, Hardy K, Harvati K, Held P, Holmes EC, Kaidonis J, Lalueza-Fox C, de la Rasilla M, Rosas A, Semal P, Soltysiak A, Townsend G, Usai D, Wahl J, Huson DH, Dobney K, Cooper A. Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. Nature. 2017;544(7650):357-61. Epub 2017/03/09. doi: 10.1038/nature21674. PubMed PMID: 28273061.
95. Stone VN, Xu P. Targeted antimicrobial therapy in the microbiome era. Mol Oral Microbiol. 2017;32(6):446-54. Epub 2017/06/14. doi: 10.1111/omi.12190. PubMed PMID: 28609586; PMCID: PMC5697594.
96. Fujimura KE, Lynch SV. Microbiota in allergy and asthma and the emerging relationship with the gut microbiome. Cell host & microbe. 2015;17(5):592-602.
97. Khosravi A, Mazmanian SK. Disruption of the gut microbiome as a risk factor for microbial infections. Current opinion in microbiology. 2013;16(2):221-7.

98. Dobney K, Brothwell D. Dental calculus: its relevance to ancient diet and oral ecology. Teeth and anthropology. 1986;291:55-81.
99. Linossier A, Gajardo M, Olavarria J. Paleomicrobiological study in dental calculus: Streptococcus mutans. Scanning Microsc. 1996;10(4):1005-13; discussion 14. Epub 1996/01/01. PubMed PMID: 9854852.
100. Poinar HN, Kuch M, Sobolik KD, Barnes I, Stankiewicz AB, Kuder T, Spaulding WG, Bryant VM, Cooper A, Paabo S. A molecular analysis of dietary diversity for three archaic Native Americans. Proc Natl Acad Sci U S A. 2001;98(8):4317-22. Epub 2001/04/11. doi: 10.1073/pnas.061014798. PubMed PMID: 11296282; PMCID: PMC31832.
101. Vandermeersch B. Middle Paleolithic dental bacteria from Kebara, Israel. CR Acad Sci Paris. 1994;319:727-31.
102. Tito RY, Macmil S, Wiley G, Najar F, Cleeland L, Qu C, Wang P, Romagne F, Leonard S, Ruiz AJ, Reinhard K, Roe BA, Lewis CM, Jr. Phylotyping and functional analysis of two ancient human microbiomes. PLoS One. 2008;3(11):e3703. Epub 2008/11/13. doi: 10.1371/journal.pone.0003703. PubMed PMID: 19002248; PMCID: PMC2577302.
103. Hagan RW, Hofman CA, Hubner A, Reinhard K, Schnorr S, Lewis CM, Jr., Sankaranarayanan K, Warinner CG. Comparison of extraction methods for recovering ancient microbial DNA from paleofeces. Am J Phys Anthropol. 2020;171(2):275-84. Epub 2019/12/01. doi: 10.1002/ajpa.23978. PubMed PMID: 31785113.
104. Mann AE, Sabin S, Ziesemer K, Vagene AJ, Schroeder H, Ozga AT, Sankaranarayanan K, Hofman CA, Fellows Yates JA, Salazar-Garcia DC, Frohlich B, Aldenderfer M, Hoogland M, Read C, Milner GR, Stone AC, Lewis CM, Jr., Krause J, Hofman C, Bos KI, Warinner C. Differential preservation of endogenous human and microbial DNA in dental calculus and dentin. Sci Rep. 2018;8(1):9822. Epub 2018/07/01. doi: 10.1038/s41598-018-28091-9. PubMed PMID: 29959351; PMCID: PMC6026117.
105. Adler CJ, Dobney K, Weyrich LS, Kaidonis J, Walker AW, Haak W, Bradshaw CJ, Townsend G, Sołtysiak A, Alt KW. Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. Nature genetics. 2013;45(4):450.
106. Ziesemer KA, Mann AE, Sankaranarayanan K, Schroeder H, Ozga AT, Brandt BW, Zaura E, Waters-Rist A, Hoogland M, Salazar-Garcia DC. Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification. Scientific Reports. 2015;5:16498.
107. Velsko IM, Yates JAF, Aron F, Hagan RW, Frantz LA, Loe L, Martinez JBR, Chaves E, Gosden C, Larson G. Microbial differences between dental plaque and historic dental calculus are related to oral biofilm maturation stage. Microbiome. 2019;7(1):102.
108. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. PLoS Comput Biol. 2012;8(9):e1002687. Epub 2012/10/03. doi: 10.1371/journal.pcbi.1002687. PubMed PMID: 23028285; PMCID: PMC3447976.
109. Layeghifard M, Hwang DM, Guttman DS. Constructing and Analyzing Microbiome Networks in R. Microbiome Analysis: Springer; 2018. p. 243-66.
110. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods. 2015;12(10):902-3. Epub 2015/09/30. doi: 10.1038/nmeth.3589. PubMed PMID: 26418763.

111. Page L, Brin S, Motwani R, Winograd T. PageRank: Bringing order to the web. Stanford Digital Libraries Working Paper, 1997.
112. Xing W, Ghorbani A, editors. Weighted pagerank algorithm. Proceedings Second Annual Conference on Communication Networks and Services Research, 2004; 2004: IEEE.
113. Bi D, Ning H, Liu S, Que X, Ding K. Gene expression patterns combined with network analysis identify hub genes associated with bladder cancer. Comput Biol Chem. 2015;56:71-83. Epub 2015/04/19. doi: 10.1016/j.compbiolchem.2015.04.001. PubMed PMID: 25889321.
114. Sporns O, Honey CJ, Kotter R. Identification and classification of hubs in brain networks. PLoS One. 2007;2(10):e1049. Epub 2007/10/18. doi: 10.1371/journal.pone.0001049. PubMed PMID: 17940613; PMCID: PMC2013941.
115. Brandes U, Borgatti SP, Freeman LC. Maintaining the duality of closeness and betweenness centrality. Social Networks. 2016;44:153-9.
116. Rochat Y. Closeness centrality extended to unconnected graphs: The harmonic centrality index. 2009.
117. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, Lipson KS, Knight R, Caporaso JG, Segata N. Species-level functional profiling of metagenomes and metatranscriptomes. Nature methods. 2018;15(11):962-8.
118. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics. 2007;23(10):1282-8. Epub 2007/03/24. doi: 10.1093/bioinformatics/btm098. PubMed PMID: 17379688.
119. Webster D. The fall of the ancient Mayasolving the mystery of the Maya collapse2002.
120. Booher AM. Assessing the form and function of the Sacbeob and associated structures at Chan Chich, Belize 2016.
121. Houk BA. The Chan Chich Archaeological Project: 1996 to 2019 Project Lists. In: Houk BA, editor. THE 2019 SEASONS OF THE BELIZE ESTATES ARCHAEOLOGICAL SURVEY TEAM. Papers of the Chan Chich Archaeological Project, Number 14: Department of Sociology, Anthropology, and Social Work, Texas Tech University, Lubbock. p. 169-208.
122. Kosakowsky LJ, Robin C. Ceramics and chronology at Chan. Chan: an ancient Maya farming community. 2012:42-70.
123. Novotny A, Robin C. The Chan community: A bioarchaeological perspective. Chan: An ancient Maya farming community. 2012:231-52.
124. Fernandes DM, Mittnik A, Olalde I, Lazaridis I, Cheronet O, Rohland N, Mallick S, Bernardos R, Broomandkhoshbacht N, Carlsson J, Culleton BJ, Ferry M, Gamarra B, Lari M, Mah M, Michel M, Modi A, Novak M, Oppenheimer J, Sirak KA, Stewardson K, Mandl K, Schattke C, Ozdogan KT, Lucci M, Gasperetti G, Candilio F, Salis G, Vai S, Camaros E, Calo C, Catalano G, Cueto M, Forgia V, Lozano M, Marini E, Micheletti M, Micciche RM, Palombo MR, Ramis D, Schimmenti V, Sureda P, Teira L, Teschler-Nicola M, Kennett DJ, Lalueza-Fox C, Patterson N, Sineo L, Coppa A, Caramelli D, Pinhasi R, Reich D. The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean. Nat Ecol Evol. 2020;4(3):334-45. Epub 2020/02/26. doi: 10.1038/s41559-020-1102-0. PubMed PMID: 32094539; PMCID: PMC7080320.
125. Ucchesu M, Peña-Chocarro L, Sabato D, Tanda G. Bronze Age subsistence in Sardinia, Italy: cultivated plants and wild resources. Vegetation history and archaeobotany. 2015;24(2):343-55.

126. Portas L, Bagella S, Farina V, Carcupino M, Cacchioli A, Gazza F, Zedda M. Study of animal remains dug out during the excavations of a Nuragic village in Sardinia. Journal of Biological Research-Bollettino della Società Italiana di Biologia Sperimentale. 2015.
127. Lai L, Tykot RH, Usai E, Beckett JF, Floris R, Fonzo O, Goddard E, Hollander D, Manunza MR, Usai A. Diet in the Sardinian Bronze Age: models, collagen isotopic data, issues and perspectives. Préhistoires Méditerranéennes. 2013(4).
128. Sarigu M, Floris GU, Floris R, Pusceddu V. The Osteological Collection of the University of Cagliari: From Early Neolithic to Modern Age. Homo. 2016;67(3):216-25. Epub 2016/03/28. doi: 10.1016/j.jchb.2016.03.001. PubMed PMID: 27017155.
129. Cooper A, Poinar HN. Ancient DNA: do it right or not at all. Science. 2000;289(5482):1139-.
130. Ozga AT, Nieves-Colon MA, Honap TP, Sankaranarayanan K, Hofman CA, Milner GR, Lewis CM, Jr., Stone AC, Warinner C. Successful enrichment and recovery of whole mitochondrial genomes from ancient human dental calculus. Am J Phys Anthropol. 2016;160(2):220-8. Epub 2016/03/19. doi: 10.1002/ajpa.22960. PubMed PMID: 26989998; PMCID: PMC4866892.
131. Rohland N, Harney E, Mallick S, Nordenfelt S, Reich D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. Philos Trans R Soc Lond B Biol Sci. 2015;370(1660):20130624. Epub 2014/12/10. doi: 10.1098/rstb.2013.0624. PubMed PMID: 25487342; PMCID: PMC4275898.
132. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes. 2016;9(1):88. Epub 2016/02/13. doi: 10.1186/s13104-016-1900-2. PubMed PMID: 26868221; PMCID: PMC4751634.
133. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72(7):5069-72. Epub 2006/07/06. doi: 10.1128/AEM.03006-05. PubMed PMID: 16820507; PMCID: PMC1489311.
134. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9. Epub 2012/03/06. doi: 10.1038/nmeth.1923. PubMed PMID: 22388286; PMCID: PMC3322381.
135. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. Epub 2009/06/10. doi: 10.1093/bioinformatics/btp352. PubMed PMID: 19505943; PMCID: PMC2723002.
136. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335-6. Epub 2010/04/13. doi: 10.1038/nmeth.f.303. PubMed PMID: 20383131; PMCID: PMC3156573.
137. Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, Bushman FD, Knight R, Kelley ST. Bayesian community-wide culture-independent microbial source tracking. Nat Methods. 2011;8(9):761-3. Epub 2011/07/19. doi: 10.1038/nmeth.1650. PubMed PMID: 21765408; PMCID: PMC3791591.

138. Jonsson H, Ginolhac A, Schubert M, Johnson PL, Orlando L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics. 2013;29(13):1682-4. Epub 2013/04/25. doi: 10.1093/bioinformatics/btt193. PubMed PMID: 23613487; PMCID: PMC3694634.
139. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60. Epub 2009/05/20. doi: 10.1093/bioinformatics/btp324. PubMed PMID: 19451168; PMCID: PMC2705234.
140. Schubert M, Ginolhac A, Lindgreen S, Thompson JF, Al-Rasheid KA, Willerslev E, Krogh A, Orlando L. Improving ancient DNA read mapping against modern reference genomes. BMC Genomics. 2012;13(1):178. Epub 2012/05/12. doi: 10.1186/1471-2164-13-178. PubMed PMID: 22574660; PMCID: PMC3468387.
141. Peltzer A, Jager G, Herbig A, Seitz A, Kniep C, Krause J, Nieselt K. EAGER: efficient ancient genome reconstruction. Genome Biol. 2016;17(1):60. Epub 2016/04/03. doi: 10.1186/s13059-016-0918-z. PubMed PMID: 27036623; PMCID: PMC4815194.
142. Team RC. R: A language and environment for statistical computing2013.
143. Kurtz Z, Mueller C, Miraldi E, Bonneau R. SpiecEasi: Sparse Inverse Covariance for Ecological Statistical Inference. R package version. 2017;1(2).
144. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal, complex systems. 2006;1695(5):1-9.
145. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, Chun J. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol. 2017;67(5):1613-7. Epub 2016/12/23. doi: 10.1099/ijsem.0.001755. PubMed PMID: 28005526; PMCID: PMC5563544.
146. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-80. Epub 2013/01/19. doi: 10.1093/molbev/mst010. PubMed PMID: 23329690; PMCID: PMC3603318.
147. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5(3):e9490. Epub 2010/03/13. doi: 10.1371/journal.pone.0009490. PubMed PMID: 20224823; PMCID: PMC2835736.
148. Oksanen J, Blanchet FG, Kindt R, Legendre P, O'hara R, Simpson GL, Solymos P, Stevens MHH, Wagner H. Vegan: community ecology package. R package version 1.17-4. URL http://CRAN R-project org/package= vegan. 2010.
149. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO. Picante: R tools for integrating phylogenies and ecology. Bioinformatics. 2010;26(11):1463-4. Epub 2010/04/17. doi: 10.1093/bioinformatics/btq166. PubMed PMID: 20395285.
150. Wickham H. ggplot2: elegant graphics for data analysis: Springer; 2016.
151. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological). 1995;57(1):289-300.
152. Eisenhofer R, Cooper A, Weyrich LS. Reply to Santiago-Rodriguez et al.: proper authentication of ancient DNA is essential. FEMS microbiology ecology. 2017;93(5).
153. Chambers ES, Preston T, Frost G, Morrison DJ. Role of Gut Microbiota-Generated Short-Chain Fatty Acids in Metabolic and Cardiovascular Health. Curr Nutr Rep. 2018;7(4):198-206.

Epub 2018/09/29. doi: 10.1007/s13668-018-0248-8. PubMed PMID: 30264354; PMCID: PMC6244749.
154. van de Wouw M, Boehme M, Lyte JM, Wiley N, Strain C, O'Sullivan O, Clarke G, Stanton C, Dinan TG, Cryan JF. Short-chain fatty acids: microbial metabolites that alleviate stress-induced brain–gut axis alterations. The Journal of physiology. 2018;596(20):4923-44.
155. Venegas DP, Marjorie K, Landskron G, González MJ, Quera R, Dijkstra G, Harmsen HJ, Faber KN, Hermoso MA. Short chain fatty acids (SCFAs)-mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. Frontiers in immunology. 2019;10.
156. Wade WG. The oral microbiome in health and disease. Pharmacol Res. 2013;69(1):137-43. Epub 2012/12/04. doi: 10.1016/j.phrs.2012.11.006. PubMed PMID: 23201354.
157. Colombo APV, Magalhães CB, Hartenbach FARR, do Souto RM, da Silva-Boghossian CM. Periodontal-disease-associated biofilm: A reservoir for pathogens of medical importance. Microbial pathogenesis. 2016;94:27-34.
158. Reyes L, Herrera D, Kozarov E, Roldán S, Progulske-Fox A. Periodontal bacterial invasion and infection: contribution to atherosclerotic pathology. Journal of clinical periodontology. 2013;40:S30-S50.
159. Suzuki N, Yoneda M, Hirofuji T. Mixed red-complex bacterial infection in periodontitis. Int J Dent. 2013;2013:587279. Epub 2013/03/28. doi: 10.1155/2013/587279. PubMed PMID: 23533413; PMCID: PMC3606728.
160. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012;490(7418):55-60.
161. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. Nature. 2007;449(7164):804-10. Epub 2007/10/19. doi: 10.1038/nature06244. PubMed PMID: 17943116; PMCID: PMC3709439.
162. Brewster R, Tamburini FB, Asiimwe E, Oduaran O, Hazelhurst S, Bhatt AS. Surveying gut microbiome research in Africans: toward improved diversity and representation. Trends in microbiology. 2019.
163. Radini A, Nikita E, Buckley S, Copeland L, Hardy K. Beyond food: The multiple pathways for inclusion of materials into ancient dental calculus. American journal of physical anthropology. 2017;162:71-83.
164. Louis P, Young P, Holtrop G, Flint HJ. Diversity of human colonic butyrate-producing bacteria revealed by analysis of the butyryl-CoA:acetate CoA-transferase gene. Environ Microbiol. 2010;12(2):304-14. Epub 2009/10/08. doi: 10.1111/j.1462-2920.2009.02066.x. PubMed PMID: 19807780.
165. Reichardt N, Duncan SH, Young P, Belenguer A, Leitch CM, Scott KP, Flint HJ, Louis P. Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. The ISME journal. 2014;8(6):1323-35.
166. Gao YD, Zhao Y, Huang J. Metabolic modeling of common Escherichia coli strains in human gut microbiome. Biomed Res Int. 2014;2014:694967. Epub 2014/08/16. doi: 10.1155/2014/694967. PubMed PMID: 25126572; PMCID: PMC4122010.
167. Mirajkar NS, Davies PR, Gebhart CJ. Antimicrobial Susceptibility Patterns of Brachyspira Species Isolated from Swine Herds in the United States. J Clin Microbiol. 2016;54(8):2109-19.

Epub 2016/06/03. doi: 10.1128/JCM.00834-16. PubMed PMID: 27252458; PMCID: PMC4963479.
168. Feberwee A, Hampson DJ, Phillips ND, La T, Van Der Heijden HM, Wellenberg GJ, Dwars RM, Landman WJ. Identification of Brachyspira hyodysenteriae and other pathogenic Brachyspira species in chickens from laying flocks with diarrhea or reduced production or both. Journal of clinical microbiology. 2008;46(2):593-600.
169. Hampson DJ. The Spirochete Brachyspira pilosicoli, Enteric Pathogen of Animals and Humans. Clin Microbiol Rev. 2018;31(1):e00087-17. Epub 2017/12/01. doi: 10.1128/CMR.00087-17. PubMed PMID: 29187397; PMCID: PMC5740978.
170. Ogata S, Shimizu K, Tominaga S, Nakanishi K. Immunohistochemical study of mucins in human intestinal spirochetosis. Hum Pathol. 2017;62:126-33. Epub 2017/02/12. doi: 10.1016/j.humpath.2017.01.013. PubMed PMID: 28188751.
171. SeÑa AC, Pillay A, Cox DL, Radolf JD. TreponemaandBrachyspira, Human Host-Associated Spirochetes. Manual of Clinical Microbiology, Eleventh Edition: American Society of Microbiology; 2015. p. 1055-81.
172. Boye M, Baloda SB, Leser TD, Moller K. Survival of Brachyspira hyodysenteriae and B. pilosicoli in terrestrial microcosms. Vet Microbiol. 2001;81(1):33-40. Epub 2001/05/18. doi: 10.1016/s0378-1135(01)00328-5. PubMed PMID: 11356316.
173. Fukuda S, Toh H, Taylor TD, Ohno H, Hattori M. Acetate-producing bifidobacteria protect the host from enteropathogenic infection via carbohydrate transporters. Gut Microbes. 2012;3(5):449-54. Epub 2012/07/25. doi: 10.4161/gmic.21214. PubMed PMID: 22825494.
174. Holt SC, Ebersole JL. Porphyromonas gingivalis, Treponema denticola, and Tannerella forsythia: the 'red complex', a prototype polybacterial pathogenic consortium in periodontitis. Periodontology 2000. 2005;38(1):72-122.
175. Kolenbrander PE, Palmer RJ, Jr., Rickard AH, Jakubovics NS, Chalmers NI, Diaz PI. Bacterial interactions and successions during plaque development. Periodontol 2000. 2006;42(1):47-79. Epub 2006/08/26. doi: 10.1111/j.1600-0757.2006.00187.x. PubMed PMID: 16930306.
176. Bothelo E, Gouriet F, Fournier PE, Roux V, Habib G, Thuny F, Metras D, Raoult D, Casalta JP. Endocarditis caused by Cardiobacterium valvarum. J Clin Microbiol. 2006;44(2):657-8. Epub 2006/02/04. doi: 10.1128/JCM.44.2.657-658.2006. PubMed PMID: 16455940; PMCID: PMC1392641.
177. Han XY, Falsen E. Characterization of oral strains of Cardiobacterium valvarum and emended description of the organism. J Clin Microbiol. 2005;43(5):2370-4. Epub 2005/05/06. doi: 10.1128/JCM.43.5.2370-2374.2005. PubMed PMID: 15872268; PMCID: PMC1153756.
178. Han XY, Falsen E. Characterization of oral strains of Cardiobacterium valvarum and emended description of the organism. Journal of clinical microbiology. 2005;43(5):2370-4.
179. Caramelli D, Vernesi C, Sanna S, Sampietro L, Lari M, Castri L, Vona G, Floris R, Francalacci P, Tykot R, Casoli A, Bertranpetit J, Lalueza-Fox C, Bertorelle G, Barbujani G. Genetic variation in prehistoric Sardinia. Hum Genet. 2007;122(3-4):327-36. Epub 2007/07/17. doi: 10.1007/s00439-007-0403-6. PubMed PMID: 17629747.
180. Gomez A, Espinoza JL, Harkins DM, Leong P, Saffery R, Bockmann M, Torralba M, Kuelbs C, Kodukula R, Inman J, Hughes T, Craig JM, Highlander SK, Jones MB, Dupont CL, Nelson KE. Host Genetic Control of the Oral Microbiome in Health and Disease. Cell Host Microbe.

2017;22(3):269-78 e3. Epub 2017/09/15. doi: 10.1016/j.chom.2017.08.013. PubMed PMID: 28910633; PMCID: PMC5733791.
181. Begemann F, Schmitt-Strecker S, Pernicka E, Schiavo FL. Chemical composition and lead isotopy of copper and bronze from Nuragic Sardinia. European Journal of Archaeology. 2001;4(1):43-85.
182. Sabatini S, Lo Schiavo F. Late Bronze Age Metal Exploitation and Trade: Sardinia and Cyprus. Materials and Manufacturing Processes. 2020:1-18.
183. Grass G, Rensing C, Solioz M. Metallic copper as an antimicrobial surface. Appl Environ Microbiol. 2011;77(5):1541-7.
184. Scott DA. Copper and bronze in art: corrosion, colorants, conservation: Getty publications; 2002.
185. Grazzi F, Brunetti A, Scherillo A, Minoja ME, Salis G, Orrù S, Depalmas A. Non-destructive compositional and microstructural characterization of Sardinian Bronze Age swords through Neutron Diffraction. Materials Characterization. 2018;144:387-92.
186. Khan ST, Ahamed M, Al-Khedhairy A, Musarrat J. Biocidal effect of copper and zinc oxide nanoparticles on human oral microbiome and biofilm formation. Materials Letters. 2013;97:67-70.
187. Haile J, Holdaway R, Oliver K, Bunce M, Gilbert MTP, Nielsen R, Munch K, Ho SY, Shapiro B, Willerslev E. Ancient DNA chronology within sediment deposits: Are paleobiological reconstructions possible and is DNA leaching a factor? Molecular biology and evolution. 2007;24(4):982-9.
188. Borry M, Cordova B, Perri A, Wibowo M, Honap TP, Ko J, Yu J, Britton K, Girdland-Flink L, Power RC. CoproID predicts the source of coprolites and paleofeces using microbiome composition and host DNA content. PeerJ. 2020;8:e9001.
189. Appelt S, Armougom F, Le Bailly M, Robert C, Drancourt M. Polyphasic analysis of a middle ages coprolite microbiota, Belgium. PLoS One. 2014;9(2):e88376. Epub 2014/03/04. doi: 10.1371/journal.pone.0088376. PubMed PMID: 24586319; PMCID: PMC3938422.
190. Cano RJ, Rivera-Perez J, Toranzos GA, Santiago-Rodriguez TM, Narganes-Storde YM, Chanlatte-Baik L, Garcia-Roldan E, Bunkley-Williams L, Massey SE. Paleomicrobiology: revealing fecal microbiomes of ancient indigenous cultures. PLoS One. 2014;9(9):e106833. Epub 2014/09/11. doi: 10.1371/journal.pone.0106833. PubMed PMID: 25207979; PMCID: PMC4160228.
191. Poinar H, Kuch M, McDonald G, Martin P, Paabo S. Nuclear gene sequences from a late pleistocene sloth coprolite. Curr Biol. 2003;13(13):1150-2. Epub 2003/07/05. doi: 10.1016/s0960-9822(03)00450-0. PubMed PMID: 12842016.
192. Dabney J, Meyer M, Paabo S. Ancient DNA damage. Cold Spring Harb Perspect Biol. 2013;5(7):a012567. Epub 2013/06/05. doi: 10.1101/cshperspect.a012567. PubMed PMID: 23729639; PMCID: PMC3685887.
193. Harper K, Armelagos G. The changing disease-scape in the third epidemiological transition. Int J Environ Res Public Health. 2010;7(2):675-97. Epub 2010/07/10. doi: 10.3390/ijerph7020675. PubMed PMID: 20616997; PMCID: PMC2872288.
194. Ou J, Carbonero F, Zoetendal EG, DeLany JP, Wang M, Newton K, Gaskins HR, O'Keefe SJ. Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African

Americans. Am J Clin Nutr. 2013;98(1):111-20. Epub 2013/05/31. doi: 10.3945/ajcn.112.056689. PubMed PMID: 23719549; PMCID: PMC3683814.
195. Schnorr SL. The diverse microbiome of the hunter-gatherer. Nature. 2015;518(7540):S14-5. Epub 2015/02/26. doi: 10.1038/518S14a. PubMed PMID: 25715276.
196. Ríos-Covián D, Ruas-Madiedo P, Margolles A, Gueimonde M, de los Reyes-Gavilán CG, Salazar N. Intestinal short chain fatty acids and their link with diet and human health. Frontiers in microbiology. 2016;7:185.
197. White D. Physiology and biochemistry of prokaryotes: Oxford University Press; 2000.
198. Louis P, Flint HJ. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. FEMS Microbiol Lett. 2009;294(1):1-8. Epub 2009/02/19. doi: 10.1111/j.1574-6968.2009.01514.x. PubMed PMID: 19222573.
199. Vital M, Howe AC, Tiedje JM. Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. mBio. 2014;5(2):e00889. Epub 2014/04/24. doi: 10.1128/mBio.00889-14. PubMed PMID: 24757212; PMCID: PMC3994512.
200. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, Lipson KS, Knight R, Caporaso JG, Segata N, Huttenhower C. Species-level functional profiling of metagenomes and metatranscriptomes. Nat Methods. 2018;15(11):962-8. Epub 2018/11/01. doi: 10.1038/s41592-018-0176-y. PubMed PMID: 30377376; PMCID: PMC6235447.
201. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. Nature. 2007;449(7164):804-10.
202. Karlsson FH, Tremaroli V, Nookaew I, Bergstrom G, Behre CJ, Fagerberg B, Nielsen J, Backhed F. Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature. 2013;498(7452):99-103. Epub 2013/05/31. doi: 10.1038/nature12198. PubMed PMID: 23719380.
203. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D, Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai Y, Sun X, Li Z, Tang A, Zhong S, Li X, Chen W, Xu R, Wang M, Feng Q, Gong M, Yu J, Zhang Y, Zhang M, Hansen T, Sanchez G, Raes J, Falony G, Okuda S, Almeida M, LeChatelier E, Renault P, Pons N, Batto JM, Zhang Z, Chen H, Yang R, Zheng W, Li S, Yang H, Wang J, Ehrlich SD, Nielsen R, Pedersen O, Kristiansen K, Wang J. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012;490(7418):55-60. Epub 2012/10/02. doi: 10.1038/nature11450. PubMed PMID: 23023125.
204. Nishijima S, Suda W, Oshima K, Kim SW, Hirose Y, Morita H, Hattori M. The gut microbiome of healthy Japanese and its microbial and functional uniqueness. DNA Res. 2016;23(2):125-33. Epub 2016/03/10. doi: 10.1093/dnares/dsw002. PubMed PMID: 26951067; PMCID: PMC4833420.
205. Kushugulova A, Forslund SK, Costea PI, Kozhakhmetov S, Khassenbekova Z, Urazova M, Nurgozhin T, Zhumadilov Z, Benberin V, Driessen M, Hercog R, Voigt AY, Benes V, Kandels-Lewis S, Sunagawa S, Letunic I, Bork P. Metagenomic analysis of gut microbial communities from a Central Asian population. BMJ Open. 2018;8(7):e021682. Epub 2018/07/30. doi: 10.1136/bmjopen-2018-021682. PubMed PMID: 30056386; PMCID: PMC6067398.
206. Zhang J, Guo Z, Lim AA, Zheng Y, Koh EY, Ho D, Qiao J, Huo D, Hou Q, Huang W, Wang L, Javzandulam C, Narangerel C, Jirimutu, Menghebilige, Lee YK, Zhang H. Mongolians core gut

microbiota and its correlation with seasonal dietary changes. Sci Rep. 2014;4:5001. Epub 2014/05/17. doi: 10.1038/srep05001. PubMed PMID: 24833488; PMCID: PMC4023135.
207. Hill MO. Diversity and evenness: a unifying notation and its consequences. Ecology. 1973;54(2):427-32.
208. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P, Tett A, Ghensi P, Collado MC, Rice BL, DuLong C, Morgan XC, Golden CD, Quince C, Huttenhower C, Segata N. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell. 2019;176(3):649-62 e20. Epub 2019/01/22. doi: 10.1016/j.cell.2019.01.001. PubMed PMID: 30661755; PMCID: PMC6349461.
209. Meier-Kolthoff JP, Goker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun. 2019;10(1):2182. Epub 2019/05/18. doi: 10.1038/s41467-019-10210-3. PubMed PMID: 31097708; PMCID: PMC6522516.
210. Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC. New insights from uncultivated genomes of the global human gut microbiome. Nature. 2019;568(7753):505-10. Epub 2019/03/15. doi: 10.1038/s41586-019-1058-x. PubMed PMID: 30867587; PMCID: PMC6784871.
211. Sirugo G, Williams SM, Tishkoff SA. The missing diversity in human genetic studies. Cell. 2019;177(1):26-31.
212. Claw KG, Henderson LM, Burke W, Thummel KE. Pharmacogenomics in Indigenous Populations. The FASEB Journal. 2019;33(1_supplement):217.2-.2.
213. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019;69(1):7-34. Epub 2019/01/09. doi: 10.3322/caac.21551. PubMed PMID: 30620402.
214. Torre LA, Trabert B, DeSantis CE, Miller KD, Samimi G, Runowicz CD, Gaudet MM, Jemal A, Siegel RL. Ovarian cancer statistics, 2018. CA Cancer J Clin. 2018;68(4):284-96. Epub 2018/05/29. doi: 10.3322/caac.21456. PubMed PMID: 29809280; PMCID: PMC6621554.
215. Jacobs IJ, Menon U. Progress and challenges in screening for early detection of ovarian cancer. Mol Cell Proteomics. 2004;3(4):355-66. Epub 2004/02/07. doi: 10.1074/mcp.R400006-MCP200. PubMed PMID: 14764655.
216. Baldwin LA, Huang B, Miller RW, Tucker T, Goodrich ST, Podzielinski I, DeSimone CP, Ueland FR, van Nagell JR, Seamon LG. Ten-year relative survival for epithelial ovarian cancer. Obstet Gynecol. 2012;120(3):612-8. Epub 2012/08/24. doi: 10.1097/AOG.0b013e318264f794. PubMed PMID: 22914471.
217. Raja FA, Chopra N, Ledermann JA. Optimal first-line treatment in ovarian cancer. Ann Oncol. 2012;23 Suppl 10(suppl_10):x118-27. Epub 2012/09/26. doi: 10.1093/annonc/mds315. PubMed PMID: 22987945.
218. Dasari S, Tchounwou PB. Cisplatin in cancer therapy: molecular mechanisms of action. European journal of pharmacology. 2014;740:364-78.
219. Chang S-J, Hodeib M, Chang J, Bristow RE. Survival impact of complete cytoreduction to no gross residual disease for advanced-stage ovarian cancer: a meta-analysis. Gynecologic oncology. 2013;130(3):493-8.
220. Ushijima K. Treatment for recurrent ovarian cancer-at first relapse. J Oncol. 2010;2010:497429. Epub 2010/01/13. doi: 10.1155/2010/497429. PubMed PMID: 20066162; PMCID: PMC2801501.

221. Davis A, Tinker AV, Friedlander M. “Platinum resistant” ovarian cancer: what is it, who to treat and how to measure benefit? Gynecologic oncology. 2014;133(3):624-31.
222. Bookman MA. Extending the platinum-free interval in recurrent ovarian cancer: the role of topotecan in second-line chemotherapy. Oncologist. 1999;4(2):87-94. Epub 1999/05/25. PubMed PMID: 10337378.
223. Gore M, Fryatt I, Wiltshaw E, Dawson T. Treatment of relapsed carcinoma of the ovary with cisplatin or carboplatin following initial treatment with these compounds. Gynecologic oncology. 1990;36(2):207-11.
224. Markman M, Rothman R, Hakes T, Reichman B, Hoskins W, Rubin S, Jones W, Almadrones L, Lewis JL, Jr. Second-line platinum therapy in patients with ovarian cancer previously treated with cisplatin. J Clin Oncol. 1991;9(3):389-93. Epub 1991/03/11. doi: 10.1200/JCO.1991.9.3.389. PubMed PMID: 1999708.
225. Pfisterer J, Ledermann JA, editors. Management of platinum-sensitive recurrent ovarian cancer. Seminars in oncology; 2006: Elsevier.
226. Bossuet-Greif N, Vignard J, Taieb F, Mirey G, Dubois D, Petit C, Oswald E, Nougayrede JP. The Colibactin Genotoxin Generates DNA Interstrand Cross-Links in Infected Cells. mBio. 2018;9(2):e02393-17. Epub 2018/03/22. doi: 10.1128/mBio.02393-17. PubMed PMID: 29559578; PMCID: PMC5874909.
227. Iida N, Dzutsev A, Stewart CA, Smith L, Bouladoux N, Weingarten RA, Molina DA, Salcedo R, Back T, Cramer S, Dai RM, Kiu H, Cardone M, Naik S, Patri AK, Wang E, Marincola FM, Frank KM, Belkaid Y, Trinchieri G, Goldszmid RS. Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. Science. 2013;342(6161):967-70. Epub 2013/11/23. doi: 10.1126/science.1240527. PubMed PMID: 24264989; PMCID: PMC6709532.
228. Perez-Chanona E, Trinchieri G. The role of microbiota in cancer therapy. Curr Opin Immunol. 2016;39:75-81. Epub 2016/01/29. doi: 10.1016/j.coi.2016.01.003. PubMed PMID: 26820225; PMCID: PMC4801762.
229. Crusz SM, Balkwill FR. Inflammation and cancer: advances and new agents. Nat Rev Clin Oncol. 2015;12(10):584-96. Epub 2015/07/01. doi: 10.1038/nrclinonc.2015.105. PubMed PMID: 26122183.
230. Zackular JP, Baxter NT, Iverson KD, Sadler WD, Petrosino JF, Chen GY, Schloss PD. The gut microbiome modulates colon tumorigenesis. mBio. 2013;4(6):e00692-13. Epub 2013/11/07. doi: 10.1128/mBio.00692-13. PubMed PMID: 24194538; PMCID: PMC3892781.
231. Zackular JP, Rogers MA, Ruffin MT, Schloss PD. The human gut microbiome as a screening tool for colorectal cancer. Cancer prevention research. 2014.
232. Champer M, Wong AM, Champer J, Brito IL, Messer PW, Hou JY, Wright JD. The role of the vaginal microbiome in gynaecological cancer. BJOG. 2018;125(3):309-15. Epub 2017/03/10. doi: 10.1111/1471-0528.14631. PubMed PMID: 28278350.
233. Chase D, Goulder A, Zenhausern F, Monk B, Herbst-Kralovetz M. The vaginal and gastrointestinal microbiomes in gynecologic cancers: a review of applications in etiology, symptoms and treatment. Gynecol Oncol. 2015;138(1):190-200. Epub 2015/05/10. doi: 10.1016/j.ygyno.2015.04.036. PubMed PMID: 25957158.
234. Muls A, Andreyev J, Lalondrelle S, Taylor A, Norton C, Hart A. Systematic Review: The Impact of Cancer Treatment on the Gut and Vaginal Microbiome in Women With a

Gynecological Malignancy. Int J Gynecol Cancer. 2017;27(7):1550-9. Epub 2017/06/08. doi: 10.1097/IGC.0000000000000999. PubMed PMID: 28590950; PMCID: PMC5571893.
235. Sharma H, Tal R, Clark NA, Segars JH, editors. Microbiota and pelvic inflammatory disease. Seminars in reproductive medicine; 2014: Thieme Medical Publishers.
236. Lin HW, Tu YY, Lin SY, Su WJ, Lin WL, Lin WZ, Wu SC, Lai YL. Risk of ovarian cancer in women with pelvic inflammatory disease: a population-based study. Lancet Oncol. 2011;12(9):900-4. Epub 2011/08/13. doi: 10.1016/S1470-2045(11)70165-6. PubMed PMID: 21835693.
237. Xu J, Peng JJ, Yang W, Fu K, Zhang Y. Vaginal microbiomes and ovarian cancer: a review. Am J Cancer Res. 2020;10(3):743-56. Epub 2020/04/09. PubMed PMID: 32266088; PMCID: PMC7136922.
238. Shi N, Li N, Duan X, Niu H. Interaction between the gut microbiome and mucosal immune system. Military Medical Research. 2017;4(1):1-7.
239. Thaiss CA, Zmora N, Levy M, Elinav E. The microbiome and innate immunity. Nature. 2016;535(7610):65-74. Epub 2016/07/08. doi: 10.1038/nature18847. PubMed PMID: 27383981.
240. Ozben T. Oxidative stress and apoptosis: impact on cancer therapy. J Pharm Sci. 2007;96(9):2181-96. Epub 2007/06/27. doi: 10.1002/jps.20874. PubMed PMID: 17593552.
241. McInnes P, Cutting M. Manual of Procedures for Human Microbiome Project, Core Microbiome Sampling, Protocol A, HMP Protocol #07-001. 2010.
242. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the national academy of sciences. 2011;108(Supplement 1):4516-22.
243. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26(19):2460-1.
244. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol. 2005;71(12):8228-35. Epub 2005/12/08. doi: 10.1128/AEM.71.12.8228-8235.2005. PubMed PMID: 16332807; PMCID: PMC1317376.
245. Ward Jr JH. Hierarchical grouping to optimize an objective function. Journal of the American statistical association. 1963;58(301):236-44.
246. Warnes MGR, Bolker B, Bonebakker L, Gentleman R, Huber W. Package 'gplots'. Various R Programming Tools for Plotting Data. 2016.
247. Aragon T. epitools: Epidemiology Tools R package version 0.5-7. Berkeley, CA: University of California. Retrieved from http://CRAN. R …; 2012.
248. Harrower M, Brewer CA. ColorBrewer. org: an online tool for selecting colour schemes for maps. The Cartographic Journal. 2003;40(1):27-37.
249. Brotman RM, Shardell MD, Gajer P, Fadrosh D, Chang K, Silver MI, Viscidi RP, Burke AE, Ravel J, Gravitt PE. Association between the vaginal microbiota, menopause status, and signs of vulvovaginal atrophy. Menopause. 2014;21(5):450-8. Epub 2013/10/02. doi: 10.1097/GME.0b013e3182a4690b. PubMed PMID: 24080849; PMCID: PMC3994184.
250. Nené NR, Reisel D, Leimbach A, Franchi D, Jones A, Evans I, Knapp S, Ryan A, Ghazali S, Timms JF. Association between the cervicovaginal microbiome, BRCA1 mutation status, and risk of ovarian cancer: a case-control study. The Lancet Oncology. 2019;20(8):1171-82.

251. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, Brotman RM, Davis CC, Ault K, Peralta L, Forney LJ. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A. 2011;108 Suppl 1(Supplement 1):4680-7. Epub 2010/06/11. doi: 10.1073/pnas.1002611107. PubMed PMID: 20534435; PMCID: PMC3063603.
252. Lamont RF, Sobel JD, Akins RA, Hassan SS, Chaiworapongsa T, Kusanovic JP, Romero R. The vaginal microbiome: new information about genital tract flora using molecular based techniques. BJOG: An International Journal of Obstetrics & Gynaecology. 2011;118(5):533-49.
253. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Dore J, Meta HITC, Antolin M, Artiguenave F, Blottiere HM, Almeida M, Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariaz G, Dervyn R, Foerstner KU, Friss C, van de Guchte M, Guedon E, Haimet F, Huber W, van Hylckama-Vlieg J, Jamet A, Juste C, Kaci G, Knol J, Lakhdari O, Layec S, Le Roux K, Maguin E, Merieux A, Melo Minardi R, M'Rini C, Muller J, Oozeer R, Parkhill J, Renault P, Rescigno M, Sanchez N, Sunagawa S, Torrejon A, Turner K, Vandemeulebrouck G, Varela E, Winogradsky Y, Zeller G, Weissenbach J, Ehrlich SD, Bork P. Enterotypes of the human gut microbiome. Nature. 2011;473(7346):174-80. Epub 2011/04/22. doi: 10.1038/nature09944. PubMed PMID: 21508958; PMCID: PMC3728647.
254. Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A, Waller A, Mende DR, Kultima JR, Martin J, Kota K, Sunyaev SR, Weinstock GM, Bork P. Genomic variation landscape of the human gut microbiome. Nature. 2013;493(7430):45-50. Epub 2012/12/12. doi: 10.1038/nature11711. PubMed PMID: 23222524; PMCID: PMC3536929.
255. Mirmonsef P, Hotton AL, Gilbert D, Burgad D, Landay A, Weber KM, Cohen M, Ravel J, Spear GT. Free glycogen in vaginal fluids is associated with Lactobacillus colonization and low vaginal pH. PloS one. 2014;9(7):e102467.
256. Mirmonsef P, Modur S, Burgad D, Gilbert D, Golub ET, French AL, McCotter K, Landay AL, Spear GT. An exploratory comparison of vaginal glycogen and Lactobacillus levels in pre-and post-menopausal women. Menopause (New York, NY). 2015;22(7):702.
257. Melkumyan A, Priputnevich T, Ankirskaya A, Murav'eva V, Lubasovskaya L. Effects of antibiotic treatment on the lactobacillus composition of vaginal microbiota. Bulletin of experimental biology and medicine. 2015;158(6):766-8.
258. Witkin SS, Linhares IM. Why do lactobacilli dominate the human vaginal microbiota? BJOG: An International Journal of Obstetrics & Gynaecology. 2017;124(4):606-11.
259. Mirmonsef P, Hotton AL, Gilbert D, Gioia CJ, Maric D, Hope TJ, Landay AL, Spear GT. Glycogen levels in undiluted genital fluid and their relationship to vaginal pH, estrogen, and progesterone. PloS one. 2016;11(4):e0153553.
260. Hemalatha R, Mastromarino P, Ramalaxmi BA, Balakrishna NV, Sesikeran B. Effectiveness of vaginal tablets containing lactobacilli versus pH tablets on vaginal health and inflammatory cytokines: a randomized, double-blind study. Eur J Clin Microbiol Infect Dis. 2012;31(11):3097-105. Epub 2012/07/11. doi: 10.1007/s10096-012-1671-1. PubMed PMID: 22777592.

261. Bachmann GA, Nevadunsky NS. Diagnosis and treatment of atrophic vaginitis. Am Fam Physician. 2000;61(10):3090-6. Epub 2000/06/06. PubMed PMID: 10839558.
262. Muhleisen AL, Herbst-Kralovetz MM. Menopause and the vaginal microbiome. Maturitas. 2016;91:42-50. Epub 2016/07/28. doi: 10.1016/j.maturitas.2016.05.015. PubMed PMID: 27451320.
263. Lewis FM, Bernstein KT, Aral SO. Vaginal Microbiome and Its Relationship to Behavior, Sexual Health, and Sexually Transmitted Diseases. Obstet Gynecol. 2017;129(4):643-54. Epub 2017/03/10. doi: 10.1097/AOG.0000000000001932. PubMed PMID: 28277350; PMCID: PMC6743080.
264. Younes JA, Lievens E, Hummelen R, van der Westen R, Reid G, Petrova MI. Women and their microbes: the unexpected friendship. Trends in microbiology. 2017.
265. Audirac-Chalifour A, Torres-Poveda K, Bahena-Román M, Téllez-Sosa J, Martínez-Barnetche J, Cortina-Ceballos B, López-Estrada G, Delgado-Romero K, Burguete-García AI, Cantú D. Cervical microbiome and cytokine profile at various stages of cervical cancer: a pilot study. PloS one. 2016;11(4):e0153274.
266. Oh HY, Kim BS, Seo SS, Kong JS, Lee JK, Park SY, Hong KM, Kim HK, Kim MK. The association of uterine cervical microbiota with an increased risk for cervical intraepithelial neoplasia in Korea. Clin Microbiol Infect. 2015;21(7):674 e1-9. Epub 2015/03/11. doi: 10.1016/j.cmi.2015.02.026. PubMed PMID: 25752224.
267. Seo SS, Oh HY, Lee JK, Kong JS, Lee DO, Kim MK. Combined effect of diet and cervical microbiome on the risk of cervical intraepithelial neoplasia. Clin Nutr. 2016;35(6):1434-41. Epub 2016/04/15. doi: 10.1016/j.clnu.2016.03.019. PubMed PMID: 27075319.
268. Brotman RM, Shardell MD, Gajer P, Tracy JK, Zenilman JM, Ravel J, Gravitt PE. Interplay between the temporal dynamics of the vaginal microbiota and human papillomavirus detection. The Journal of infectious diseases. 2014;210(11):1723-33.
269. Delley M, Bruttin A, Richard M, Affolter M, Rezzonico E, Bruck WM. In vitro activity of commercial probiotic Lactobacillus strains against uropathogenic Escherichia coli. FEMS Microbiol Lett. 2015;362(13):fnv096. Epub 2015/06/17. doi: 10.1093/femsle/fnv096. PubMed PMID: 26078118.
270. Gupta K, Stapleton AE, Hooton TM, Roberts PL, Fennell CL, Stamm WE. Inverse association of H2O2-producing lactobacilli and vaginal Escherichia coli colonization in women with recurrent urinary tract infections. J Infect Dis. 1998;178(2):446-50. Epub 1998/08/11. doi: 10.1086/515635. PubMed PMID: 9697725.
271. Heinonen PK, Miettinen A. Laparoscopic study on the microbiology and severity of acute pelvic inflammatory disease. Eur J Obstet Gynecol Reprod Biol. 1994;57(2):85-9. Epub 1994/11/01. doi: 10.1016/0028-2243(94)90048-5. PubMed PMID: 7859910.
272. Youssef O, Lahti L, Kokkola A, Karla T, Tikkanen M, Ehsan H, Carpelan-Holmstrom M, Koskensalo S, Bohling T, Rautelin H, Puolakkainen P, Knuutila S, Sarhadi V. Stool Microbiota Composition Differs in Patients with Stomach, Colon, and Rectal Neoplasms. Dig Dis Sci. 2018;63(11):2950-8. Epub 2018/07/12. doi: 10.1007/s10620-018-5190-5. PubMed PMID: 29995183; PMCID: PMC6182444.
273. de Groot PF, Belzer C, Aydin O, Levin E, Levels JH, Aalvink S, Boot F, Holleman F, van Raalte DH, Scheithauer TP, Simsek S, Schaap FG, Olde Damink SWM, Roep BO, Hoekstra JB, de Vos WM, Nieuwdorp M. Distinct fecal and oral microbiota composition in human type 1

diabetes, an observational study. PLoS One. 2017;12(12):e0188475. Epub 2017/12/07. doi: 10.1371/journal.pone.0188475. PubMed PMID: 29211757; PMCID: PMC5718513.
274. Vacca M, Celano G, Calabrese FM, Portincasa P, Gobbetti M, De Angelis M. The Controversial Role of Human Gut Lachnospiraceae. Microorganisms. 2020;8(4):573. Epub 2020/04/25. doi: 10.3390/microorganisms8040573. PubMed PMID: 32326636; PMCID: PMC7232163.
275. Lun H, Yang W, Zhao S, Jiang M, Xu M, Liu F, Wang Y. Altered gut microbiota and microbial biomarkers associated with chronic kidney disease. Microbiologyopen. 2019;8(4):e00678. Epub 2018/08/09. doi: 10.1002/mbo3.678. PubMed PMID: 30088332; PMCID: PMC6460263.
276. Montassier E, Gastinne T, Vangay P, Al-Ghalith GA, Bruley des Varannes S, Massart S, Moreau P, Potel G, de La Cochetiere MF, Batard E, Knights D. Chemotherapy-driven dysbiosis in the intestinal microbiome. Aliment Pharmacol Ther. 2015;42(5):515-28. Epub 2015/07/07. doi: 10.1111/apt.13302. PubMed PMID: 26147207.
277. Hakim H, Dallas R, Wolf J, Tang L, Schultz-Cherry S, Darling V, Johnson C, Karlsson EA, Chang TC, Jeha S, Pui CH, Sun Y, Pounds S, Hayden RT, Tuomanen E, Rosch JW. Gut Microbiome Composition Predicts Infection Risk During Chemotherapy in Children With Acute Lymphoblastic Leukemia. Clin Infect Dis. 2018;67(4):541-8. Epub 2018/03/09. doi: 10.1093/cid/ciy153. PubMed PMID: 29518185; PMCID: PMC6070042.
278. Alexander JL, Wilson ID, Teare J, Marchesi JR, Nicholson JK, Kinross JM. Gut microbiota modulation of chemotherapy efficacy and toxicity. Nat Rev Gastroenterol Hepatol. 2017;14(6):356-65. Epub 2017/03/09. doi: 10.1038/nrgastro.2017.20. PubMed PMID: 28270698.
279. Menni C, Jackson MA, Pallister T, Steves CJ, Spector TD, Valdes AM. Gut microbiome diversity and high-fibre intake are related to lower long-term weight gain. Int J Obes (Lond). 2017;41(7):1099-105. Epub 2017/03/14. doi: 10.1038/ijo.2017.66. PubMed PMID: 28286339; PMCID: PMC5500185.
280. Reese AT, Dunn RR. Drivers of Microbiome Biodiversity: A Review of General Rules, Feces, and Ignorance. mBio. 2018;9(4):e01294-18. Epub 2018/08/02. doi: 10.1128/mBio.01294-18. PubMed PMID: 30065092; PMCID: PMC6069118.
281. Ley RE. Gut microbiota in 2015: Prevotella in the gut: choose carefully. Nat Rev Gastroenterol Hepatol. 2016;13(2):69-70. Epub 2016/02/02. doi: 10.1038/nrgastro.2016.4. PubMed PMID: 26828918.
282. Larsen JM. The immune response to Prevotella bacteria in chronic inflammatory disease. Immunology. 2017;151(4):363-74. Epub 2017/05/26. doi: 10.1111/imm.12760. PubMed PMID: 28542929; PMCID: PMC5506432.
283. Scher JU, Sczesnak A, Longman RS, Segata N, Ubeda C, Bielski C, Rostron T, Cerundolo V, Pamer EG, Abramson SB, Huttenhower C, Littman DR. Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. Elife. 2013;2:e01202. Epub 2013/11/07. doi: 10.7554/eLife.01202. PubMed PMID: 24192039; PMCID: PMC3816614.
284. Tian L, Bashan A, Shi DN, Liu YY. Articulation points in complex networks. Nat Commun. 2017;8(1):14223. Epub 2017/02/01. doi: 10.1038/ncomms14223. PubMed PMID: 28139697; PMCID: PMC5290321.

285. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. Frontiers in microbiology. 2017;8:2224.
286. Tsilimigras MC, Fodor AA. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. Ann Epidemiol. 2016;26(5):330-5. Epub 2016/06/04. doi: 10.1016/j.annepidem.2016.03.002. PubMed PMID: 27255738.
287. Fang H, Huang C, Zhao H, Deng M. gCoda: Conditional Dependence Network Inference for Compositional Data. J Comput Biol. 2017;24(7):699-708. Epub 2017/05/11. doi: 10.1089/cmb.2017.0054. PubMed PMID: 28489411; PMCID: PMC5510714.
288. Kurtz ZD, Muller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and compositionally robust inference of microbial ecological networks. PLoS Comput Biol. 2015;11(5):e1004226. Epub 2015/05/08. doi: 10.1371/journal.pcbi.1004226. PubMed PMID: 25950956; PMCID: PMC4423992.
289. Faust K, Bauchinger F, Laroche B, de Buyl S, Lahti L, Washburne AD, Gonze D, Widder S. Signatures of ecological processes in microbial community time series. Microbiome. 2018;6(1):120. Epub 2018/06/30. doi: 10.1186/s40168-018-0496-2. PubMed PMID: 29954432; PMCID: PMC6022718.
290. Faust K, Raes J. Microbial interactions: from networks to models. Nat Rev Microbiol. 2012;10(8):538-50. Epub 2012/07/17. doi: 10.1038/nrmicro2832. PubMed PMID: 22796884.
291. Barber MJ. Modularity and community detection in bipartite networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2007;76(6 Pt 2):066102. Epub 2008/02/01. doi: 10.1103/PhysRevE.76.066102. PubMed PMID: 18233893.
292. Newman ME. Modularity and community structure in networks. Proc Natl Acad Sci U S A. 2006;103(23):8577-82. Epub 2006/05/26. doi: 10.1073/pnas.0601602103. PubMed PMID: 16723398; PMCID: PMC1482622.
293. Fang X, Monk JM, Nurk S, Akseshina M, Zhu Q, Gemmell C, Gianetto-Hill C, Leung N, Szubin R, Sanders J. Metagenomics-based, strain-level analysis of Escherichia coli from a time-series of microbiome samples from a Crohn's disease patient. Frontiers in microbiology. 2018;9:2559.
294. Gao Y-D, Zhao Y, Huang J. Metabolic modeling of common Escherichia coli strains in human gut microbiome. BioMed research international. 2014;2014.
295. Million M, Thuny F, Angelakis E, Casalta J, Giorgi R, Habib G, Raoult D. Lactobacillus reuteri and Escherichia coli in the human gut microbiota may predict weight gain associated with vancomycin treatment. Nutrition & diabetes. 2013;3(9):e87-e.

Supplementary Material A

Co-Authors, affiliations, contributions, supplementary figures, and supplementary tables for Chapter 2: Functional Diversity of Microbial Ecologies Estimated from Ancient Human Coprolites and Dental Calculus; adapted from Jacobson et al. in press. Functional Diversity of

Microbial Ecologies Estimated from Ancient Human Coprolites and Dental Calculus.

Philosophical Transactions of the Royal Society B.

Author List and Affiliations

David K. Jacobson[1,2], Tanvi P. Honap[1,2], Cara Monroe[1], Justin Lund[1,2], Brett A. Houk[3], Anna C. Novotny[3], Cynthia Robin[4], Elisabetta Marini[5], and Cecil M. Lewis, Jr.[1,2]

[1] Laboratories of Molecular Anthropology and Microbiome Research (LMAMR), University of Oklahoma, Norman, Oklahoma, USA

[2] Department of Anthropology, University of Oklahoma, Norman, Oklahoma, USA

[3] Department of Sociology, Anthropology, and Social Work, Texas Tech University, Lubbock, Texas, USA

[4] Department of Anthropology, Northwestern University, Evanston, Illinois, USA

[5] Department of Life and Environmental Sciences, University of Cagliari, Cagliari, Sardinia, Italy

Authors' Contributions

BAH, ACN, and CR collected and processed the Maya dental calculus samples prior to DNA extraction. EM collected and processed the Nuragic dental calculus samples prior to DNA extraction. JRL and CM performed wet lab work for the Maya and Nuragic dental calculus samples. DKJ and TPH performed data analysis. DKJ, TPH, and CML conceived and wrote the manuscript.

Microbiome Network Analysis

Co-occurring bacteria, keystone taxa, and community structure can be represented as an ecological network. Network analysis takes advantage of elements of graph theory to uncover relationships between members of complex communities, whether it be infrastructure networks, social networks, or biological ecosystems (39, 284). Generating microbiome networks requires special attention due to the nature of sequencing data. Microbiome data are compositional due to

an arbitrary maximum number of sequencing reads that can be generated on a NGS sequencing instrument (285) and this compositionality can lead to spurious relationships if it is not taken into account (285, 286). Various methods have been developed to address compositionality in microbiome networks (108, 143, 287, 288) but in general, they use log-ratio transformations prior to downstream analysis (108, 143, 285, 286, 288). In microbiome network analysis, each bacterial taxon forms an individual circular node and straight-line edges connect nodes, which represents correlations between two taxa and various cutoffs can be used to depict strength of correlations. In Supplementary Figure A - 9A, we present an example of a network with edges representing positive correlations >0.3. A node's degree is the number of edges connected to that node and clusters are groups of nodes that share a high number of connections within the cluster and fewer connections outside the cluster. Identifying **nodes, edges, and clusters** are key tenets of network analysis (39, 109, 289, 290). Network theory has applications in many fields of study, and is a broad and growing field, so here we focus on only a few important aspects that are easily translated to the characterization of a microbial ecology.

Co-occurring clusters of bacteria not only inform about which bacteria are co-dependent and interact with each other, but also inform about the structure of the microbial community. Knowing the structure of the microbiome is important ecologically because communities are made up of more than just singular interactions - there are niches and subgroups within larger communities. The taxonomic and functional nature of clusters can highlight how taxa are partitioned within the microbiome and potentially suggest spatial or niche segregation. Various methods have been created to detect clusters within a network but in general they work by simulating network clusters and choosing a network topology that optimizes modularity (strength

of division into clusters) and transitivity (a measure of connectivity) of the network. Clusters are typically connected to other clusters in the network but can be isolated from the rest of the network**. Isolated clusters** may indicate a highly specialized set of taxa or functions.

Modularity (291, 292) is an important feature in uncovering microbiome structure as high modularity indicates there are many interactions within any given cluster, but few interactions between clusters, while low modularity indicates that there are many connections between clusters (Supplementary Figure A - 9B). In microbiomes with highly segregated functional groups and niches, we would expect high modularity as taxa within any given niche have few interactions with bacteria outside the niche. Similarly, we can interpret structure from the number and size of clusters within a microbiome. If there are few clusters with many bacteria, the community may have less specialized structuring and thus more fluidity in ecological function. **Transitivity** provides an indication of whether nodes already connected through a central node are likely to be connected to each other independent of that central node (39, 109). High transitivity means many connections between bacteria and many routes to connect bacteria, suggesting a microbiome community with many different layers of interactions (Supplementary Figure A - 9C). Modularity and transitivity typically have an inverse relationship to each other, as networks with low modularity and high transitivity both signal high numbers connections between nodes.

Modularity, transitivity, and co-occurring clusters present a holistic view of the network structure; in other words, they are focused on emergent network properties. There is additional benefit to focusing on individual nodes in the network. A primary interest for microbiome

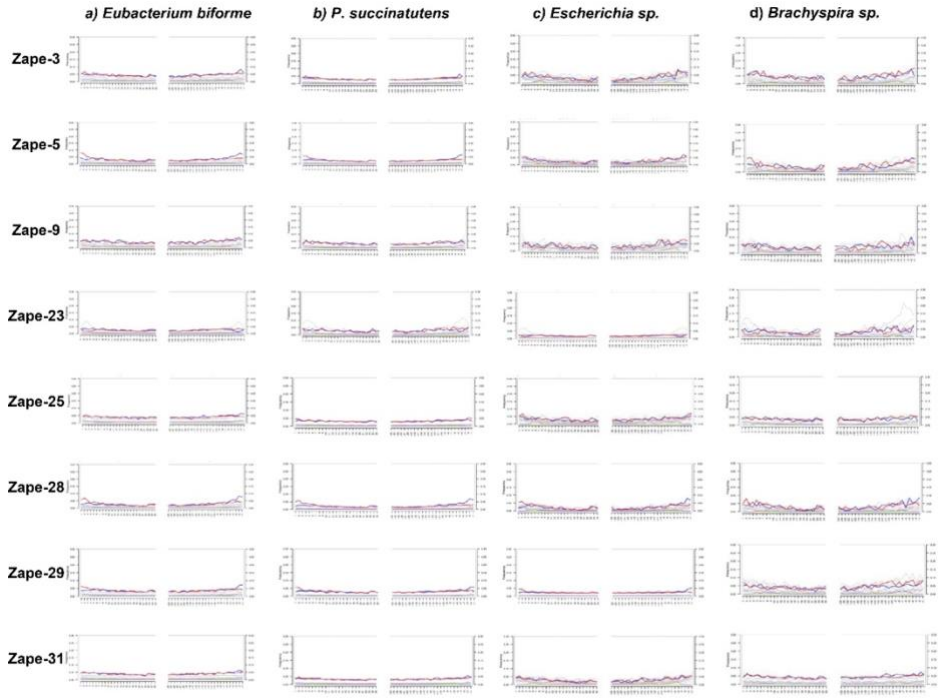
ecology research is how individual taxa are connected to the rest of the network: are they at the center of the whole network, do they connect different clusters together, or are they isolated from taxa across the entire network? There are a multitude of approaches for identifying taxa central to the network and the most straightforward mechanism is by looking at its **degree**. Degree is simply a measure of how many connections a given node (bacterial taxon) maintains and it provides a quick way to identify highly and sparsely connected bacteria in the network. In our analysis, **Hub Score** can be thought of as an analogue to degree (39). A slightly more nuanced approach is to use the **PageRank** algorithm, in which each node is given a weight depending on the quantity and quality of connections for each node but ultimately provides a similar interpretation as using degrees (39, 111, 112). **Centrality** is a further method for determining how an individual node interacts with the remainder of the network (115). **Closeness Centrality** depends on the use of paths, which trace the number of edges needed to connect any two nodes. High closeness nodes are those nodes that have the shortest average path length between itself and all other nodes in the network, meaning that it is central to the full network.

The connectivity of individual nodes in the network uncovers keystone taxa and taxa important for ecological stability. **Keystone taxa** have large influences on the microbiome community, independent of their relative abundance (38). These bacteria may produce specific nutrients that are metabolized by other microbes or provide protection against environmental stressors. Keystone taxa are at the center of a network because they are not just important for one group of bacteria, but rather are important for the entire community to function. In microbiome networks, keystone taxa should be thought about as hubs for the network and thus are central to the full network. Therefore, we can determine keystone taxa using Hub Score, PageRank, and Closeness

Centrality. There is a possibility that keystone identified by Hub Score and PageRank will differ than those identified through Closeness Centrality because the former two methods rely on number and quality of connections to each node, while the latter relies on tracing paths throughout the network. However, in most cases the keystone taxa identified by HubScore and PageRank are the same as those identified by using Closeness Centrality.

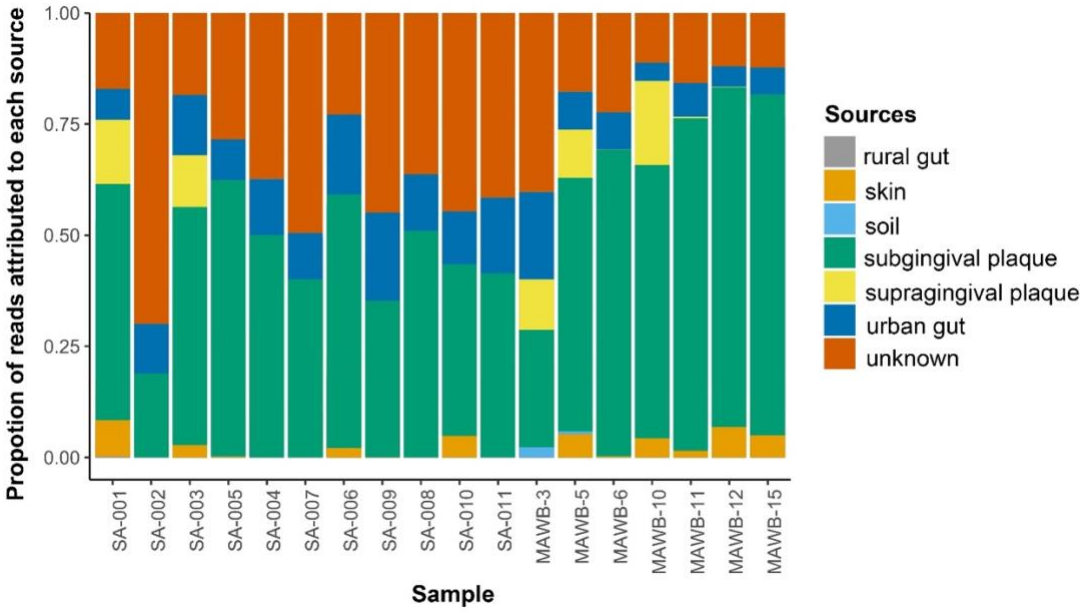
Outside of keystone taxa, other taxa can also have a large impact on the network by serving as **articulation points** (284). Articulation points are nodes that connect two different clusters together and they are the only node that connect those two clusters. Articulation points are important for maintaining integrity in the network. Such taxa are often different from keystone species because often articulation points will have a small degree and be found more towards the periphery of the network; however, it is possible for an articulation point to have high centrality and serve as a keystone. Peripheral articulation points can be just as important as keystone taxa because removal of an articulation point can lead to disconnected clusters in the network (284). Disconnected clusters may lead to instability and potentially a loss of resilience in the community.

Supplementary Figures A: 1-9



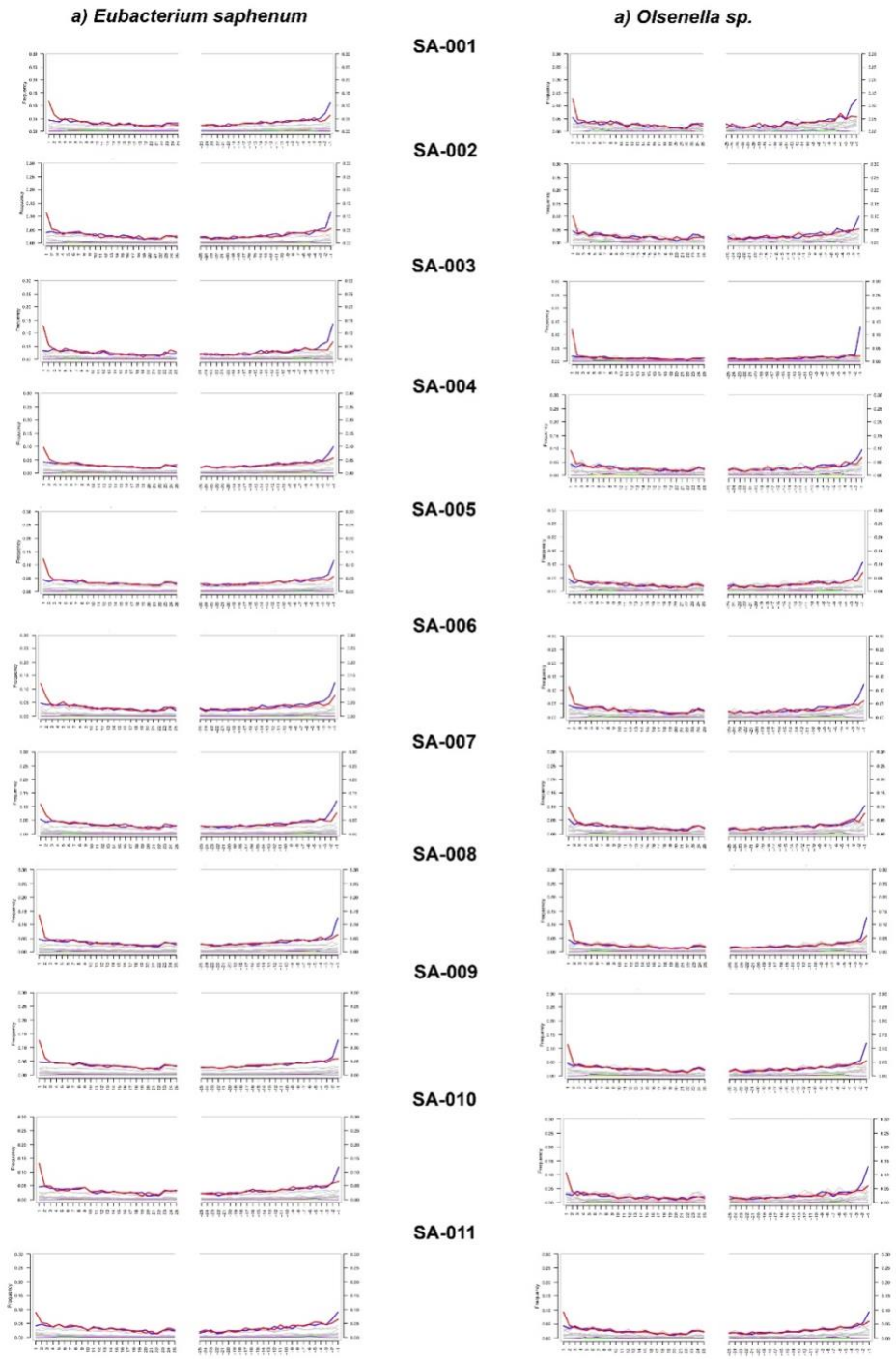
Supplementary Figure A - 1A-D: Rio Zape coprolite DNA damage patterns for keystone taxa

Damage plots generated using MapDamage for the keystones identified in the Rio Zape coprolites. Red indicates C to T transitions and blue indicates G to A transitions. The Y axis shows the proportion of sites containing the nucleotide change and the X axis shows the position along the DNA fragment. Escherichia and Brachyspira keystones were not identified at the species level; however, E. coli and B. pilosicoli were both identified in the coprolite samples and were thus used as references in MapDamage. Damage patterns are consistent with ancient DNA for the Rio Zape coprolite keystones.



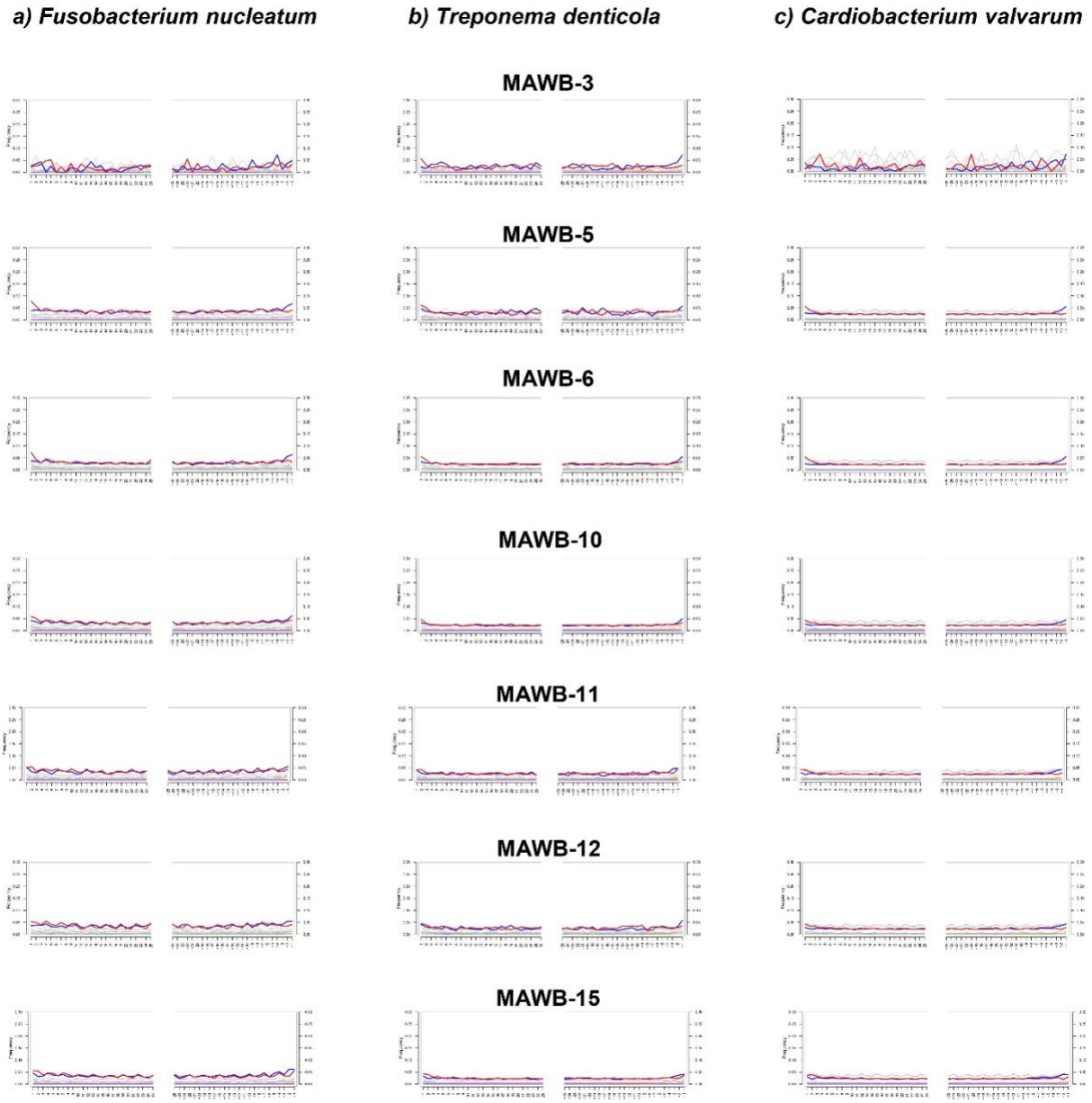
Supplementary Figure A - 2: SourceTracker results for novel ancient dental calculus samples

Stacked bar plots of Bayesian SourceTracker results for the dental calculus samples from the Nuragic and Maya individuals. The Y axis shows estimated proportions of source contribution at the genus level, using modern subgingival and supragingival plaque, urban and rural gut, skin, and soil datasets as model sources.



Supplementary Figure A - 3A-B: Nuragic dental calculus DNA damage plots for keystone taxa.

Damage plots generated using MapDamage for the keystones identified in Nuragic dental calculus. Olsenella was not identified at the species level as a keystone; however, O. uli was identified in the Nuragic samples and was thus used as the reference for MapDamage. A) E. saphenum and B) O. uli show damage patterns that are consistent with ancient DNA.



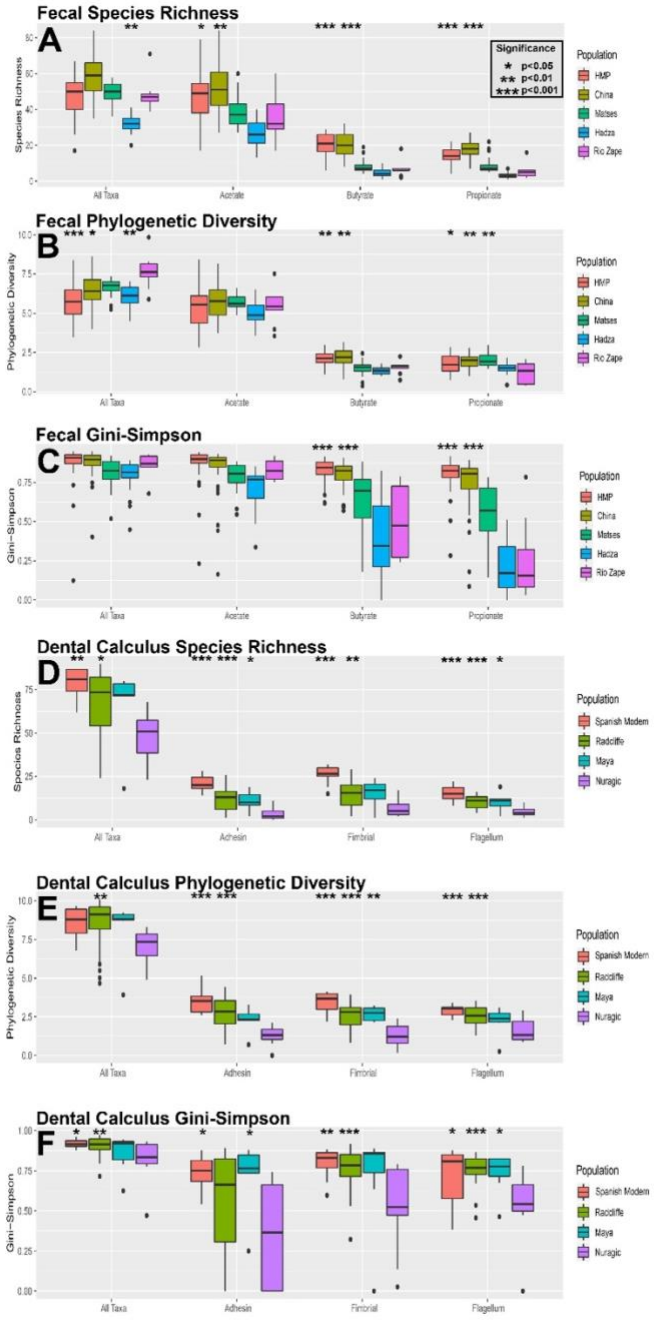
Supplementary Figure A - 4A-C: Maya dental calculus DNA damage plots for keystone taxa.

Damage plots generated using MapDamage for the keystones identified in Maya dental calculus. Damage patterns are consistent with ancient DNA for the Maya dental calculus keystones



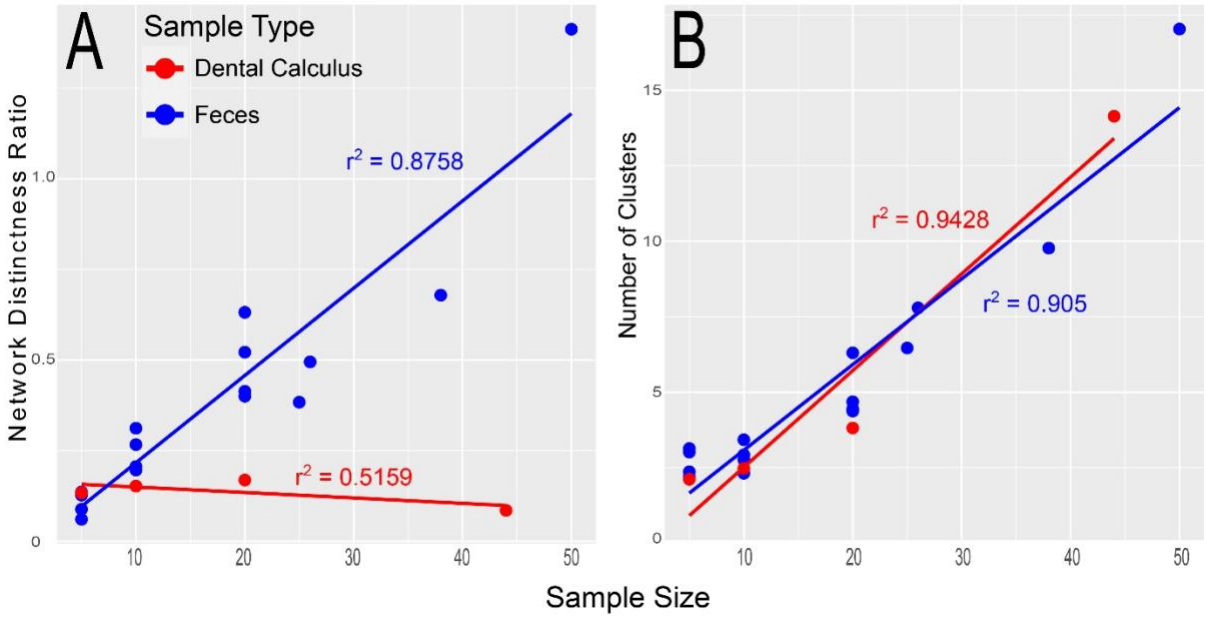
Supplementary Figure A - 5: Co-occurrence heatmap of oral taxa of interest

Heatmap demonstrating co-occurrence of early colonizing oral taxa and selected periodontitis-associated bacteria (across top of table) with common oral taxa. Networks were generated 100 times and the shade of red represents the number of network iterations that the selected bacteria localize in same cluster as taxa along the y-axis. Cells in gray represent taxa that were not found in each respective dataset. In the Maya and Radcliffe datasets, A. naeslundii has a distinct clustering pattern compared to other early colonizing/red complex bacteria, while in the Nuragic population, S. gordonii shows a unique clustering pattern.



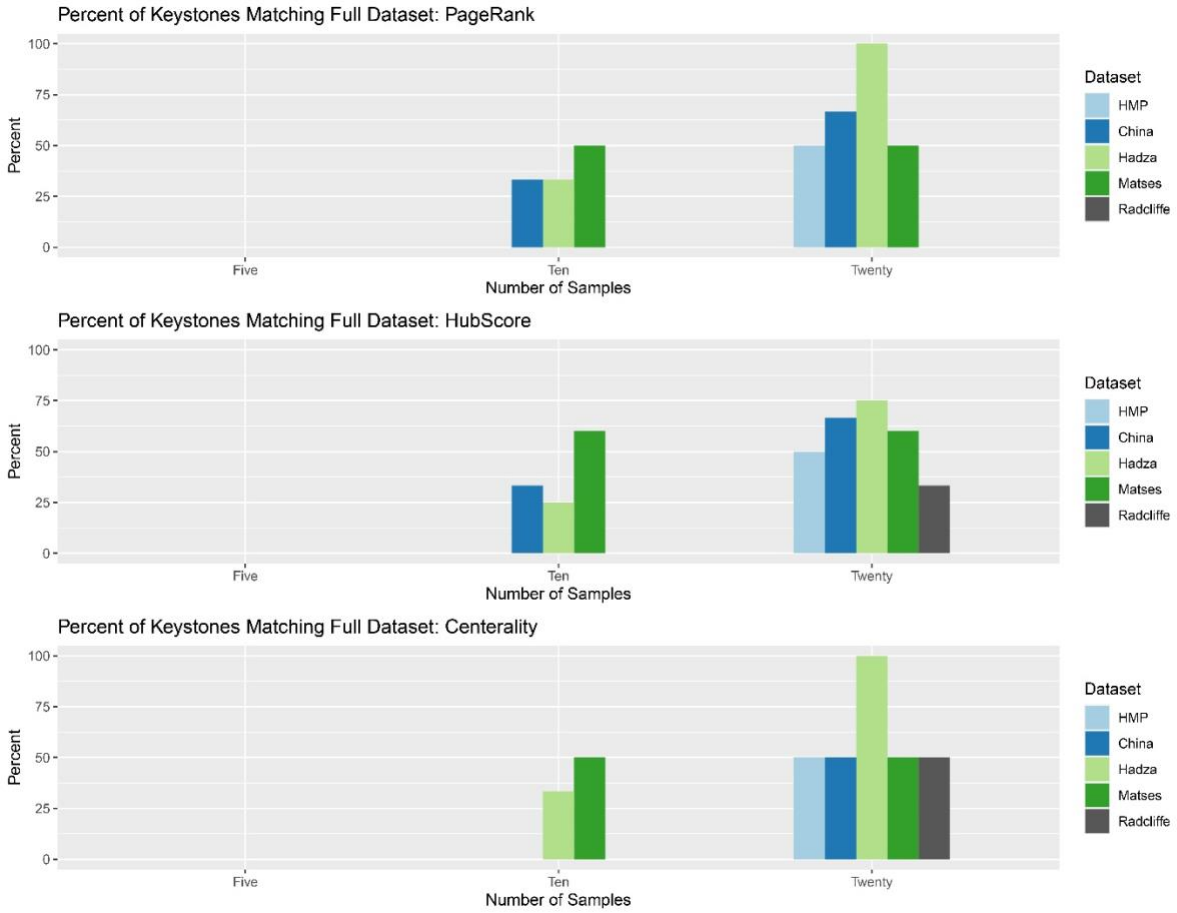
Supplementary Figure A - 6A-F: Functional diversity in ancient and modern microbiome samples

Significant p-values are given in reference to the Rio Zape coprolites in A-C, while significant p-values are given in reference to the Nuragic dataset in D-F. A-C) Modern non-industrial gut microbiomes are similar to the Rio Zape coprolites, while the modern industrial datasets are more diverse than the coprolites. Increased functional diversity in modern industrial gut microbiomes may be driven by database bias. D) The Nuragic dataset is an outlier for functional diversity compared to the modern and other ancient dental calculus datasets.



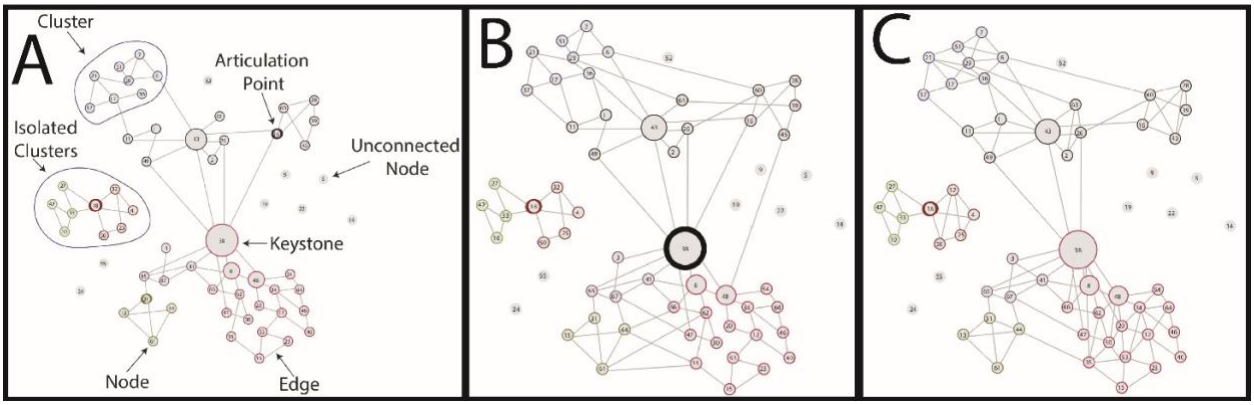
Supplementary Figure A - 7A-B: Small sample size effect network properties

Effect of small sample size on network properties for both gut microbiomes and dental calculus. A) The number of clusters increases directly with sample size in both sample types. B)The network distinctness ratio is modularity divided by transitivity and this ratio helps operationalize the interconnectivity of the network. Higher network distinctness is found in networks with clusters and nodes that are distinct from each other. Network distinctness increases with sample size for gut microbiomes but not for ancient dental calculus.



Supplementary Figure A - 8: Small sample size hinders identification of keystone taxa

Small sample size results in poor recovery of keystones from full dataset. The y-axis represents the percent of keystones from the full dataset that were found in each of the small sample size datasets. Five and Ten sample networks have few keystones matching the keystones from the full dataset.



Supplementary Figure A - 9A-C: Visual representation of network properties

Supplementary Figure 9: A) Visual representation of network properties. Each number represents an individual taxon as a node, each color represents a cluster, each line represents connected nodes, and nodes with thick borders are keystone taxa. B) Network with low modularity (i.e. clusters are highly connected to each other). C) Network with high transitivity (i.e. nodes are connected to other nodes without needing to be connected through a central or keystone taxa). High transitivity and low modularity are often found in the same network.

Supplementary Tables A: 1-8

Supplementary Table A - 1: Archaeological and anatomical context for Nuragic and Maya samples

LMAMR Sample ID	Population	Archaeological ID	Individual ID	Archaeological Site
SA-001	Nuragic (Sardinia)	LMC1: MSAE 6506	6506	Lu Maccioni
SA-002	Nuragic (Sardinia)	LMC2: MSAE 6525	6525	Lu Maccioni
SA-003	Nuragic (Sardinia)	LMC3: MSAE 6515	6515	Lu Maccioni
SA-004	Nuragic (Sardinia)	LMC4: MSAE 6507	6507	Lu Maccioni
SA-005	Nuragic (Sardinia)	CPP1: MSAE 6151	6151	Capo Pecora
SA-006	Nuragic (Sardinia)	CPP2: MSAE 6153	6153	Capo Pecora
SA-007	Nuragic (Sardinia)	CPP3: MSAE 6176	6176	Capo Pecora
SA-008	Nuragic (Sardinia)	CPP4: MSAE 6120	6120	Capo Pecora
SA-009	Nuragic (Sardinia)	SRD1: MSAE 6626	6626	Perdalba
SA-010	Nuragic (Sardinia)	SRD2: MSAE 6623	6623	Perdalba

SA-011	Nuragic (Sardinia)	SRD3: MSAE 6612	6612	Perdalba
MAWB-3	Maya (Belize)	BRV-CH19	CH19	Chan
MAWB-5	Maya (Belize)	BRV-CH6	CH6	Chan
MAWB-6	Maya (Belize)	CC-B12	1	Chan Chich
MAWB-10	Maya (Belize)	CC-B14	1	Chan Chich
MAWB-11	Maya (Belize)	CC-B14	1	Chan Chich
MAWB-12	Maya (Belize)	CC-B14	1	Chan Chich
MAWB-15	Maya (Belize)	CC-B14	1	Chan Chich

LMAMR Sample ID	**Date**	**Radiocarbon or Archaeological**	**Tooth sampled**	**Total Raw Reads**
SA-001	1126-825 calBCE	Radiocarbon (from the same layer)	Mandibular second left molar	16,317,989
SA-002	1126-825 calBCE	Radiocarbon (from the same layer)	Mandibular second right molar	12,375,354
SA-003	1126-825 calBCE	Radiocarbon (from the same layer)	Mandibular third right molar	11,376,242
SA-004	1126-825 calBCE	Radiocarbon (from the same layer)	Mandibular third left molar	13,545,219
SA-005	1384-936 calBCE	Radiocarbon (from the same layer)	Maxillary left incisor	15,259,048
SA-006	1384-936 calBCE	Radiocarbon (from the same layer)	Maxillary third second molar	11,169,001
SA-007	1384-936 calBCE	Radiocarbon (from the same layer)	Mandibular left first incisor	12,895,505
SA-008	1384-936 calBCE	Radiocarbon (from the same layer)	Mandibular third right molar	15,430,462
SA-009	1900-1300 BCE (Middle Bronze Age)	Archaeological	Mandibular second left molar	14,265,263
SA-010	1900-1300 BCE (Middle Bronze Age)	Archaeological	Mandibular first right molar	10,724,679
SA-011	1900-1300 BCE (Middle Bronze Age)	Archaeological	Maxillary first left premolar	14,632,189
MAWB-3	2-sigma cal.BCE 170-CE 50*	Radiocarbon	Maxillary right lateral incisor	2,730,685
MAWB-5	2-sigma cal. 570-660 CE*	Radiocarbon	Mandibular right lateral incisor	9,631,144
MAWB-6	2-sigma cal. 713–885 CE**	Radiocarbon	Maxillary right central incisor	32,402,210
MAWB-10	830-1000 CE (Late-to-Terminal Classic)	Archaeological	Maxillary canine	51,541,754
MAWB-11	830-1000 CE (Late-to-Terminal Classic)	Archaeological	Maxillary canine	11,999,444
MAWB-12	830-1000 CE (Late-to-Terminal Classic)	Archaeological	Mandibular incisor	8,652,497
MAWB-15	830-1000 CE (Late-to-Terminal Classic)	Archaeological	Mandibular incisor	35,754,125

Supplementary Table A - 2: Metagenome samples used in this study that were downloaded from NCBI

Sample Name	Dataset	Sample Type	Modern or Ancient	Article	Analysis Ready Reads
Zape23	RioZape	Coprolites	Ancient	Hagan et al. 2020	9047526
Zape25	RioZape	Coprolites	Ancient	Hagan et al. 2020	21126417
Zape28	RioZape	Coprolites	Ancient	Hagan et al. 2020	6942671
Zape29	RioZape	Coprolites	Ancient	Hagan et al. 2020	21134414
Zape31	RioZape	Coprolites	Ancient	Hagan et al. 2020	24219550
Zape31	RioZape	Coprolites	Ancient	Hagan et al. 2020	15571845
Zape9	RioZape	Coprolites	Ancient	Hagan et al. 2020	9473478
ERR3003613	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	7432088
ERR3003614	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	9224977
ERR3003615	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	11888299
ERR3003616	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	11147785
ERR3003617	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	11375775
ERR3003618	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	8856397
ERR3003619	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	8443593
ERR3003620	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	13444340
ERR3003621	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	15568455
ERR3003622	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	18881786
ERR3003623	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	6976790
ERR3003624	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	14396865
ERR3003625	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	1636949
ERR3003626	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	12959172
ERR3003627	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	10735389
ERR3003628	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	13671623
ERR3003629	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	5619963
ERR3003630	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	6892173

ERR3003631	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	18613560
ERR3003632	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	16636369
ERR3003633	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	16505193
ERR3003634	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	25529619
ERR3003635	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	7250609
ERR3003636	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	5037146
ERR3003637	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	544226
ERR3003638	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	13049172
ERR3003639	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	20570653
ERR3003640	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	5655198
ERR3003641	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	21869403
ERR3003642	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	29512892
ERR3003643	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	11130624
ERR3003644	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	8832898
ERR3003645	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	27353094
ERR3003646	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	7181939
ERR3003647	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	22769227
ERR3003648	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	14199232
ERR3003649	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	9306260
ERR3003650	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	4837871
ERR3003651	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	5594629
ERR3003652	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	8611932
ERR3003653	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	10701129
ERR3003654	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	4350720
ERR3003655	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	9494263
ERR3003656	Radcliffe	DentalCalculus	Ancient	Velsko et al. 2019	29865341
ERR3307045	Spanish	DentalCalculus	Modern	Velsko et al. 2019	532661
ERR3307046	Spanish	DentalCalculus	Modern	Velsko et al. 2019	66116509

ERR3307047	Spanish	DentalCalculus	Modern	Velsko et al. 2019	50448700
ERR3307048	Spanish	DentalCalculus	Modern	Velsko et al. 2019	47951659
ERR3307049	Spanish	DentalCalculus	Modern	Velsko et al. 2019	49725788
ERR3307050	Spanish	DentalCalculus	Modern	Velsko et al. 2019	46646925
ERR3307051	Spanish	DentalCalculus	Modern	Velsko et al. 2019	44934750
ERR3307052	Spanish	DentalCalculus	Modern	Velsko et al. 2019	29167396
ERR3307053	Spanish	DentalCalculus	Modern	Velsko et al. 2019	49779231
ERR3307054	Spanish	DentalCalculus	Modern	Velsko et al. 2019	42609453
HMP_70001371	Human Microbiome Project	Feces	Modern	Methe et al. 2012	69559884
HMP_70001456	Human Microbiome Project	Feces	Modern	Methe et al. 2012	60328206
HMP_70001472	Human Microbiome Project	Feces	Modern	Methe et al. 2012	65768913
HMP_70001483	Human Microbiome Project	Feces	Modern	Methe et al. 2012	124239150
HMP_70001511	Human Microbiome Project	Feces	Modern	Methe et al. 2012	62790792
HMP_70001518	Human Microbiome Project	Feces	Modern	Methe et al. 2012	55171741
HMP_70001525	Human Microbiome Project	Feces	Modern	Methe et al. 2012	67801058
HMP_70001541	Human Microbiome Project	Feces	Modern	Methe et al. 2012	68904447
HMP_70001585	Human Microbiome Project	Feces	Modern	Methe et al. 2012	68373859
HMP_70001592	Human Microbiome Project	Feces	Modern	Methe et al. 2012	63686163
HMP_70001598	Human Microbiome Project	Feces	Modern	Methe et al. 2012	61297782
HMP_70001614	Human Microbiome Project	Feces	Modern	Methe et al. 2012	67429644
HMP_70001645	Human Microbiome Project	Feces	Modern	Methe et al. 2012	92884679

HMP_70001654	Human Microbiome Project	Feces	Modern	Methe et al. 2012	60661296
HMP_70001661	Human Microbiome Project	Feces	Modern	Methe et al. 2012	56801679
HMP_70001676	Human Microbiome Project	Feces	Modern	Methe et al. 2012	68981342
HMP_70001696	Human Microbiome Project	Feces	Modern	Methe et al. 2012	61064234
HMP_70002130	Human Microbiome Project	Feces	Modern	Methe et al. 2012	60102109
HMP_70002182	Human Microbiome Project	Feces	Modern	Methe et al. 2012	49572783
HMP_70002187	Human Microbiome Project	Feces	Modern	Methe et al. 2012	53859064
HMP_70002190	Human Microbiome Project	Feces	Modern	Methe et al. 2012	46993323
HMP_70002311	Human Microbiome Project	Feces	Modern	Methe et al. 2012	55727490
HMP_70002326	Human Microbiome Project	Feces	Modern	Methe et al. 2012	45389755
HMP_70002333	Human Microbiome Project	Feces	Modern	Methe et al. 2012	56015647
HMP_70002357	Human Microbiome Project	Feces	Modern	Methe et al. 2012	46214451
HMP_70002363	Human Microbiome Project	Feces	Modern	Methe et al. 2012	59541250
HMP_70002372	Human Microbiome Project	Feces	Modern	Methe et al. 2012	55382577
HMP_70002384	Human Microbiome Project	Feces	Modern	Methe et al. 2012	49526090
HMP_70002387	Human Microbiome Project	Feces	Modern	Methe et al. 2012	70233273
HMP_70002391	Human Microbiome Project	Feces	Modern	Methe et al. 2012	55612409
HMP_70002402	Human Microbiome Project	Feces	Modern	Methe et al. 2012	35149975

HMP_70002423	Human Microbiome Project	Feces	Modern	Methe et al. 2012	89685829
HMP_70002431	Human Microbiome Project	Feces	Modern	Methe et al. 2012	61627637
HMP_70002443	Human Microbiome Project	Feces	Modern	Methe et al. 2012	57012863
HMP_70002444	Human Microbiome Project	Feces	Modern	Methe et al. 2012	55469673
HMP_70002450	Human Microbiome Project	Feces	Modern	Methe et al. 2012	63665931
HMP_70002461	Human Microbiome Project	Feces	Modern	Methe et al. 2012	55080220
HMP_70002467	Human Microbiome Project	Feces	Modern	Methe et al. 2012	49800536
HMP_70002471	Human Microbiome Project	Feces	Modern	Methe et al. 2012	80926958
HMP_70002475	Human Microbiome Project	Feces	Modern	Methe et al. 2012	46935506
HMP_70002486	Human Microbiome Project	Feces	Modern	Methe et al. 2012	58765847
HMP_70002493	Human Microbiome Project	Feces	Modern	Methe et al. 2012	83543419
HMP_70002499	Human Microbiome Project	Feces	Modern	Methe et al. 2012	55518468
HMP_70003222	Human Microbiome Project	Feces	Modern	Methe et al. 2012	65325909
HMP_70003224	Human Microbiome Project	Feces	Modern	Methe et al. 2012	57218206
HMP_70003233	Human Microbiome Project	Feces	Modern	Methe et al. 2012	63596310
HMP_70003294	Human Microbiome Project	Feces	Modern	Methe et al. 2012	57182192
HMP_70003315	Human Microbiome Project	Feces	Modern	Methe et al. 2012	55130349
HMP_70003343	Human Microbiome Project	Feces	Modern	Methe et al. 2012	68709617

HMP_700033502	Human Microbiome Project	Feces	Modern	Methe et al. 2012	64224547
HMP_700033665	Human Microbiome Project	Feces	Modern	Methe et al. 2012	67213936
Had1929408	Hadza	Feces	Modern	Rampelli et al. 2015	31667557
Had1929484	Hadza	Feces	Modern	Rampelli et al. 2015	8023250
Had1929485	Hadza	Feces	Modern	Rampelli et al. 2015	4582373
Had1929563	Hadza	Feces	Modern	Rampelli et al. 2015	13965176
Had1929574	Hadza	Feces	Modern	Rampelli et al. 2015	10368513
Had1930121	Hadza	Feces	Modern	Rampelli et al. 2015	34860393
Had1930122	Hadza	Feces	Modern	Rampelli et al. 2015	15283833
Had1930123	Hadza	Feces	Modern	Rampelli et al. 2015	36600411
Had1930128	Hadza	Feces	Modern	Rampelli et al. 2015	14926939
Had1930132	Hadza	Feces	Modern	Rampelli et al. 2015	4278117
Had1930133	Hadza	Feces	Modern	Rampelli et al. 2015	4896636
Had1930134	Hadza	Feces	Modern	Rampelli et al. 2015	10999317
Had1930136	Hadza	Feces	Modern	Rampelli et al. 2015	13262019
Had1930138	Hadza	Feces	Modern	Rampelli et al. 2015	4698201
Had1930140	Hadza	Feces	Modern	Rampelli et al. 2015	7972145
Had1930141	Hadza	Feces	Modern	Rampelli et al. 2015	32205660
Had1930142	Hadza	Feces	Modern	Rampelli et al. 2015	5258350
Had1930143	Hadza	Feces	Modern	Rampelli et al. 2015	6562608
Had1930144	Hadza	Feces	Modern	Rampelli et al. 2015	4848670
Had1930145	Hadza	Feces	Modern	Rampelli et al. 2015	16560525
Had1930149	Hadza	Feces	Modern	Rampelli et al. 2015	4073672
Had1930176	Hadza	Feces	Modern	Rampelli et al. 2015	5041521
Had1930177	Hadza	Feces	Modern	Rampelli et al. 2015	5364707
Had1930179	Hadza	Feces	Modern	Rampelli et al. 2015	4013392
Had1930187	Hadza	Feces	Modern	Rampelli et al. 2015	3114848

Had1930244	Hadza	Feces	Modern	Rampelli et al. 2015	7743506
bgi-N075A	China	Feces	Modern	Qin et al. 2012	36105871
bgi-NLF002	China	Feces	Modern	Qin et al. 2012	16058868
bgi-NLF005	China	Feces	Modern	Qin et al. 2012	19796667
bgi-NLF006	China	Feces	Modern	Qin et al. 2012	13659087
bgi-NLF007	China	Feces	Modern	Qin et al. 2012	22421737
bgi-NLF009	China	Feces	Modern	Qin et al. 2012	17280426
bgi-NLF010	China	Feces	Modern	Qin et al. 2012	16371986
bgi-NLF011	China	Feces	Modern	Qin et al. 2012	16745859
bgi-NLF014	China	Feces	Modern	Qin et al. 2012	17711182
bgi-NLF015	China	Feces	Modern	Qin et al. 2012	21249687
bgi-NLM006	China	Feces	Modern	Qin et al. 2012	25751198
bgi-NLM010	China	Feces	Modern	Qin et al. 2012	24699774
bgi-NLM015	China	Feces	Modern	Qin et al. 2012	14397119
bgi-NLM016	China	Feces	Modern	Qin et al. 2012	13553365
bgi-NLM022	China	Feces	Modern	Qin et al. 2012	21914212
bgi-NLM023	China	Feces	Modern	Qin et al. 2012	27269972
bgi-NLM027	China	Feces	Modern	Qin et al. 2012	23760705
bgi-NLM028	China	Feces	Modern	Qin et al. 2012	23235883
bgi-NLM029	China	Feces	Modern	Qin et al. 2012	23089593
bgi-NLM031	China	Feces	Modern	Qin et al. 2012	23091231
bgi-NOF002	China	Feces	Modern	Qin et al. 2012	20696915
bgi-NOF005	China	Feces	Modern	Qin et al. 2012	22765953
bgi-NOF008	China	Feces	Modern	Qin et al. 2012	24152020
bgi-NOF009	China	Feces	Modern	Qin et al. 2012	21438264
bgi-NOF012	China	Feces	Modern	Qin et al. 2012	19714306
bgi-NOF013	China	Feces	Modern	Qin et al. 2012	21137523
bgi-NOF014	China	Feces	Modern	Qin et al. 2012	17546823
bgi-NOM001	China	Feces	Modern	Qin et al. 2012	14058605
bgi-NOM004	China	Feces	Modern	Qin et al. 2012	20126357
bgi-NOM007	China	Feces	Modern	Qin et al. 2012	14905722
bgi-NOM009	China	Feces	Modern	Qin et al. 2012	22994893
bgi-NOM017	China	Feces	Modern	Qin et al. 2012	17256657
bgi-NOM018	China	Feces	Modern	Qin et al. 2012	21390737
bgi-NOM019	China	Feces	Modern	Qin et al. 2012	15779040
bgi-NOM020	China	Feces	Modern	Qin et al. 2012	20163678

bgi-NOM023	China	Feces	Modern	Qin et al. 2012	17845086
bgi-NOM027	China	Feces	Modern	Qin et al. 2012	17335223
bgi-NOM028	China	Feces	Modern	Qin et al. 2012	19197467
SM01	Matses	Feces	Modern	Obregon-Tito et al. 2015	36243326
SM02	Matses	Feces	Modern	Obregon-Tito et al. 2015	46085816
SM03	Matses	Feces	Modern	Obregon-Tito et al. 2015	22843103
SM05	Matses	Feces	Modern	Obregon-Tito et al. 2015	23906323
SM10	Matses	Feces	Modern	Obregon-Tito et al. 2015	27723664
SM11	Matses	Feces	Modern	Obregon-Tito et al. 2015	33348829
SM18	Matses	Feces	Modern	Obregon-Tito et al. 2015	32026148
SM20	Matses	Feces	Modern	Obregon-Tito et al. 2015	30099313
SM23	Matses	Feces	Modern	Obregon-Tito et al. 2015	30146410
SM24	Matses	Feces	Modern	Obregon-Tito et al. 2015	32991287
SM25	Matses	Feces	Modern	Obregon-Tito et al. 2015	31026199
SM28	Matses	Feces	Modern	Obregon-Tito et al. 2015	27389507
SM29	Matses	Feces	Modern	Obregon-Tito et al. 2015	30151973
SM30	Matses	Feces	Modern	Obregon-Tito et al. 2015	27805060
SM31	Matses	Feces	Modern	Obregon-Tito et al. 2015	30133715
SM32	Matses	Feces	Modern	Obregon-Tito et al. 2015	35061331
SM33	Matses	Feces	Modern	Obregon-Tito et al. 2015	28563888
SM34	Matses	Feces	Modern	Obregon-Tito et al. 2015	26961514
SM37	Matses	Feces	Modern	Obregon-Tito et al. 2015	27433220
SM39	Matses	Feces	Modern	Obregon-Tito et al. 2015	33510372
SM40	Matses	Feces	Modern	Obregon-Tito et al. 2015	33098422
SM41	Matses	Feces	Modern	Obregon-Tito et al. 2015	30877682
SM42	Matses	Feces	Modern	Obregon-Tito et al. 2015	28698309
SM43	Matses	Feces	Modern	Obregon-Tito et al. 2015	31343591
SM44	Matses	Feces	Modern	Obregon-Tito et al. 2015	31275968

Supplementary Table A - 3A-D: Top 50 genes in keystone taxa - Rio Zape Coprolites

Top 50 genes by gene abundance from HUMAnN2 for each keystone taxa found in the Rio Zape coprolites. Abundance is gene copies per 1 million gene copies, with the mean value across the dataset reported for each gene. Antibiotic resistance genes are in bold.

a) Escherichia	
Gene Name	meanAbund
Escherichia coli IMT2125 genomic chromosome, IMT2125	157.21948
hypothetical protein	147.55349
Escherichia coli IMT2125 genomic chromosome, IMT2125|unclassified	104.17424
hypothetical protein, partial	68.28250
Putative membrane protein	37.22662
Membrane protein	32.14988
Predicted protein	28.65321
Escherichia coli 1540 plasmid pIP1206 complete genome	23.98338
Putative membrane protein (Fragment)	23.44465
Escherichia coli 1540 plasmid pIP1206 complete genome|unclassified	19.47129
Mannitol-1-phosphate 5-dehydrogenase	18.06854
Conserved domain protein	14.58802
TTG start codon	13.55998
Transposase	10.47736
Transposase family protein	9.33703
Protein SrnB	8.47705
Ornithine carbamoyltransferase 1	7.97669
RepA3	7.73112
Multiple stress resistance protein BhsA domain protein	7.55945
Thioredoxin reductase	7.21115
Transposase, IS605 family	7.18980
PyrBI operon leader peptide	7.08167
Putative asparagine synthetase B	6.95930
Acetyltransferase	6.77352
Enterobactin synthase	6.35817
Ribonucleoside-diphosphate reductase 1, beta subunit, B2	6.34910
MalG gene 3-flanking DNA	6.09923
Transport of hexuronates	5.96863
Transcriptional regulator	5.90242
Hemolysin E, chromosomal domain protein	5.84491
Cellulose synthase catalytic subunit [UDP-forming]	5.79237
Glutamyl-tRNA synthetase domain protein	5.74202
Phage recombination protein Bet (Fragment)	5.71533
RpmH ribosomal protein L34	5.61438

Tryptophan permease	5.53631
cell division protein FtsL	5.44006
Phosphatidylserine decarboxylase	5.41644
Single-stranded DNA-binding protein	5.38347
Toxin SymE, type I toxin-antitoxin system family protein	5.30155
Abc transport membrane permease	5.24969
Putative transposase	5.21585
Aldo/keto reductase	5.16524
Protein rof	5.08851
Putative IS1 encoded protein	5.05463
Transcription elongation factor	4.96258
Putative HTH-type transcriptional regulator YneL	4.96033
Ybl54	4.75796
N-acetyl-gamma-glutamyl-phosphate reductase	4.67759
Sulfate transport system permease protein CysT	4.64893
Thiamine import ATP-binding protein ThiQ	4.50032

b) Brachyspira

Gene Name	meanAbund
hypothetical protein	6.42956
TPR domain-containing protein	6.07162
Lipoprotein	3.65874
TPR repeat-containing protein	3.44506
Glycosyl transferase family 2	3.23542
Ankyrin repeat-containing protein	3.08389
Extracellular solute-binding protein, family 5	2.91664
Acriflavin resistance protein	**2.73465**
Pseudouridine synthase	2.71509
Pyruvate phosphate dikinase	2.68776
Serpulina hyodysenteriae variable surface protein	2.48640
Ankyrin	2.25450
3-deoxy-7-phosphoheptulonate synthase	2.18766
Methyltransferase	2.17242
ABC transporter related protein	2.13497
Outer membrane protein	2.01585
Galactose-1-phosphate uridylyltransferase	2.00448
Appr-1-p processing domain protein	1.99894
Transporter	1.75380
Cytidylate kinase	1.67780
SAM-dependent methyltransferase	1.67613
hypothetical protein, partial	1.63529
Radical SAM domain protein	1.61393

Inner-membrane translocator	1.55803
Flavodoxin	1.51300
Extracellular solute-binding protein family 1	1.48899
Glycosyltransferase	1.43515
Peptidase M23	1.39032
Tetratricopeptide TPR_2 repeat protein	1.37358
50S ribosomal protein L3 (Fragment)	1.34407
RNA polymerase sigma factor	1.32794
Phosphopentomutase	1.30253
Adenine specific DNA methyltransferase	1.28853
D-3-phosphoglycerate dehydrogenase	1.28524
Thioredoxin reductase	1.25088
Putative reductase BN758_00609	1.24574
Unclassified	1.23447
5-methylcytosine restriction system component-like protein	1.19024
MATE efflux family protein	**1.15107**
Uridine phosphorylase	1.14636
50S ribosomal protein L1	1.14591
50S ribosomal protein L11	1.09380
CheW protein	1.09207
N-acetylmuramoyl-L-alanine amidase	1.08677
N-acylglucosamine 2-epimerase	1.07755
Chemotaxis protein methyltransferase CheR	1.06766
TatD protein	1.06152
Carbohydrate kinase, PfkB family	1.04430
Putative K(+)-stimulated pyrophosphate-energized sodium pump	1.03691
Methyltransferase type 11	1.02666

c) Eubacterium biforme

Gene Name	meanAbund
Transposase	24.28352
Putative transposase DNA-binding domain protein (Fragment)	20.74955
Transposase-like protein	13.41998
ABC transporter, ATP-binding protein	12.67177
MATE efflux family protein	**12.28083**
50S ribosomal protein L36	10.65286
50S ribosomal protein L34	10.16945
Transposase, IS116/IS110/IS902 family	8.23550
ATP synthase subunit c	8.19592
ABC transporter, substrate-binding protein, family 5	8.06157
HAD hydrolase, family IA, variant 3	7.81502
30S ribosomal protein S13	7.02504

ABC transporter permease protein	6.96366
30S ribosomal protein S11	6.76403
Transcriptional regulator, TetR family	6.07580
ATP synthase subunit b	5.89358
30S ribosomal protein S14 type Z	5.89029
Cof-like hydrolase	5.84100
ATPase/histidine kinase/DNA gyrase B/HSP90 domain protein	5.83098
Pseudouridine synthase	5.66666
Single-stranded DNA-binding protein	5.65279
Amidohydrolase	5.44604
ATP synthase gamma chain	5.44145
Addiction module toxin, RelE/StbE family	5.37597
ATP synthase epsilon chain	5.36676
50S ribosomal protein L32	5.34519
ATP-dependent zinc metalloprotease FtsH	5.01661
Diguanylate cyclase (GGDEF) domain protein	4.97768
DNA-directed RNA polymerase subunit alpha	4.97711
Amidophosphoribosyltransferase	4.97575
Transcriptional regulator	4.92216
50S ribosomal protein L29	4.77886
Elongation factor P	4.72858
PTS family mannose porter, IIC component	4.66532
SIS domain protein	4.66105
30S ribosomal protein S18	4.61437
Transporter	4.60889
UDP-glucose 4-epimerase	4.60178
50S ribosomal protein L33 1	4.60120
IS66 family element, transposase	4.59927
50S ribosomal protein L31	4.52545
30S ribosomal protein S10	4.46549
ATP synthase subunit beta	4.44541
Peptide deformylase	4.43869
Putative transposase	4.36506
PTS system mannose/fructose/sorbose family IIB component	4.34517
Transcriptional regulator, XRE family	4.34135
Thioredoxin	4.32730
Transcriptional regulator, MarR family	4.32331
Serine carboxypeptidase	4.31232

d) Phascolarctobacterium succinatutens

Gene Name	meanAbund
50S ribosomal protein L33 1	7.80331

DNA-binding helix-turn-helix protein	6.50082
ATPase/histidine kinase/DNA gyrase B/HSP90 domain protein	6.17674
Transporter, DASS family	5.64666
50S ribosomal protein L30	5.55721
Periplasmic binding protein	4.91736
ABC transporter ATP-binding protein	4.70193
4Fe-4S binding domain protein	4.66279
MATE efflux family protein	**4.64157**
Elongation factor Tu, apicoplast	4.59695
Pyridine nucleotide-disulfide oxidoreductase	3.82089
Nucleoside diphosphate kinase	3.78357
FAD linked oxidase domain protein	3.47685
Peptidyl-prolyl cis-trans isomerase	3.41422
F420-non-reducing hydrogenase iron-sulfur subunit D	3.36498
30S ribosomal protein S18	3.30487
Oxidoreductase, nitrogenase component 1	3.09705
ABC polar amino acid transporter	3.01383
Outer membrane protein	2.97801
Transposase	2.96064
Cobalt transporter	2.89418
Lipoprotein	2.89072
50S ribosomal protein L21	2.84896
Cysteine--tRNA ligase	2.84868
ATPase	2.81229
TonB-dependent receptor	2.69838
Monovalent cation/H+ antiporter subunit B domain protein	2.69413
30S ribosomal protein S21	2.68228
30S ribosomal protein S13	2.67281
DNA-directed RNA polymerase subunit beta	2.67030
FAD dependent oxidoreductase	2.65715
Radical SAM domain protein	2.65164
2-nitropropane dioxygenase NPD	2.64955
50S ribosomal protein L16, chloroplastic	2.64650
50S ribosomal protein L35	2.61008
Transporter	2.60551
50S ribosomal protein L20	2.59246
Transposase (Fragment)	2.58272
50S ribosomal protein L29	2.56885
RNA binding S1 domain protein	2.51467
30S ribosomal protein S9	2.48475
Extracellular ligand-binding receptor	2.47446

50S ribosomal protein L27	2.47359
30S ribosomal protein S15	2.44770
Glutamate--tRNA ligase	2.43361
Acriflavin resistance protein	**2.43237**
Response regulator receiver domain protein	2.40683
Rubrerythrin (RR)	2.40118
Amidohydrolase	2.36100
ATP synthase subunit b	2.35636

Supplementary Table A - 4A-B: Top 50 genes in keystone taxa - Nuragic dental calculus

Top 50 genes by gene abundance from HUMAnN2 for each keystone taxa found in the Nuragic dental calculus samples. Abundance is gene copies per 1 million gene copies, with the mean value across the dataset reported for each gene. Antibiotic resistance genes are in bold.

a) Eubacterium saphenum	
Gene Name	meanAbund
Bacterial surface protein 26-residue PARCEL repeat (3 repeats)	14.95073
50S ribosomal protein L31	12.79862
YibE/F-like protein	12.63424
Efflux ABC transporter, permease protein	11.61885
Flavodoxin	10.75676
50S ribosomal protein L34	8.60214
TIGR02185 family protein	8.56613
DNA-damage-inducible protein D family protein	8.36224
NA+/H+ antiporter NHAC	8.27645
Elongation factor Tu, apicoplast	7.88859
Repeat protein	7.62742
30S ribosomal protein S21	7.57228
Fic family protein	6.80146
Phenazine biosynthesis protein, PhzF family	6.67937
CoA-binding domain protein	6.61519
AMP-binding enzyme	6.53746
Translation initiation factor IF-1	6.49625
Hypothetical bacterial integral membrane protein (Trep_Strep)	6.44447
30S ribosomal protein S17	6.14516
LPXTG-motif cell wall anchor domain protein	6.11107
Polysaccharide deacetylase	6.01976
Bacterial group 2 Ig-like protein	6.01798
Biotin synthase	5.97151
30S ribosomal protein S13	5.87702
50S ribosomal protein L18	5.79206
Papain family cysteine protease	5.78749
50S ribosomal protein L30	5.72232
50S ribosomal protein L29	5.57128
Bacteriocin-associated integral membrane protein	5.43963
Putative septation protein SpoVG	5.39673
DNA-binding protein HU	5.39548
RIP metalloprotease RseP	5.31936
50S ribosomal protein L35	5.31727
30S ribosomal protein S7	5.26371

50S ribosomal protein L5	5.19890
30S ribosomal protein S14 type Z	5.17576
30S ribosomal protein S8	5.16821
Serine-type D-Ala-D-Ala carboxypeptidase	5.15686
50S ribosomal protein L14	5.06668
Phage major tail protein, phi13 family	4.75000
Bacteriocin, lactococcin 972 family	4.70212
Amino acid permease-associated region	4.63399
FMN-binding domain protein	4.60785
DNA repair protein RecO	4.56762
30S ribosomal protein S18	4.52065
50S ribosomal protein L16	4.46216
Tryptophanase	4.44278
30S ribosomal protein S15	4.38201
30S ribosomal protein S6	4.30991
V-type sodium ATPase K subunit	4.25547

b) Olsenella

Gene Name	meanAbund
DNA-binding helix-turn-helix protein	26.92384
ABC1 family protein	14.98862
ABC transporter, ATP-binding protein	11.84994
ABC-2 family transporter protein	10.22147
MacB-like periplasmic core domain protein	10.01913
Transcriptional regulator, DeoR family	9.55413
Acetolactate synthase, small subunit	8.56009
Glycoside hydrolase, family 25	8.43790
PF14335 domain protein	8.24566
Transcriptional regulator, ArsR family	8.03606
Transcriptional regulator, AbrB family	7.30884
PF13635 domain protein	7.30504
ATP-dependent DNA helicase RecG C-terminal domain protein	7.14551
Haloacid dehalogenase-like hydrolase	6.91339
Ketopantoate reductase ApbA/PanE domain protein	6.72580
Nucleotidyl transferase, PF08843 domain protein	6.65711
Haloacid dehalogenase-like hydrolase domain protein	6.63774
Site-specific recombinase, phage integrase domain protein	6.41275
Toxin-antitoxin system, toxin component, Fic domain protein	6.29031
Site-specific recombinase, phage integrase family	6.28408
GHKL domain protein	6.26230
HAD-superfamily hydrolase, subfamily IIB	6.10579
Alpha/beta hydrolase fold-3 domain protein	5.97270

DNA methylase family protein	5.73054
Cinnamoyl ester hydrolase	5.65442
HAD hydrolase, family IA, variant 3	5.42257
Putative major cell-binding factor	5.02274
Histidinol phosphate phosphatase HisJ family	4.97715
Cell division protein FtsZ	4.96390
Zinc-finger of transposase IS204/IS1001/IS1096/IS1165 (Fragment)	4.89528
Integral membrane sensor signal transduction histidine kinase	4.80561
3-isopropylmalate dehydrogenase	4.66355
SCP-2 sterol transfer family protein	4.55784
Short-chain dehydrogenase/reductase SDR	4.53504
Acetylornithine aminotransferase	4.38097
Virulence activator alpha C-terminal family protein	4.35334
Acetylglutamate kinase	4.32056
Calcineurin-like phosphoesterase family protein	4.26203
Transposase domain protein (Fragment)	4.17989
Fic/DOC family protein	4.07258
TatD-related deoxyribonuclease	4.06937
Galactokinase galactose-binding signature	3.99070
Peptidase, S9A/B/C family, catalytic domain protein	3.88265
Basic membrane domain protein	3.87643
PF06115 domain protein	3.86580
Small molecule-binding regulator domain protein	3.85141
PF14014 family protein	3.80053
Sortase, SrtB family	3.73665
ATPase	3.73579
N-acetylmuramoyl-L-alanine amidase domain protein	3.70939

Supplementary Table A - 5A-C: Top 50 genes in keystone taxa - Maya dental calculus

Top 50 genes by gene abundance from HUMAnN2 for each keystone taxa found in the Mayan dental calculus samples. Abundance is gene copies per 1 million gene copies, with the mean value across the dataset reported for each gene. Antibiotic resistance genes are in bold

Fusobacterium nucleatum	
Gene Name	meanAbund
Cell wall-associated hydrolase	79.57285
Transposase	20.11884153
Hypothetical cytosolic protein	9.835568
Flavodoxin	5.583699738
Transporter	5.241214675
Transposase, IS605 OrfB family	5.16738375
Integral membrane protein	4.89230775
Transposase, IS605 OrfB family, central region	4.47566375
Possible transcriptional regulator	3.867855963
Hypothetical Cytosolic Protein	3.838350738
MATE efflux family protein	**3.603400888**
ISChy9, transposase OrfB	3.3649925
Transcriptional regulator	3.321478463
Peptidyl-prolyl cis-trans isomerase	3.317424
Methyltransferase	3.10504935
Ethanolamine utilization protein	2.8974015
Transposase-like protein B	2.85549875
MORN repeat protein	2.7722654
Cysteine synthase	2.760272625
Tetratricopeptide repeat family protein	2.759906438
Lipoprotein	2.634973
Predicted protein	2.536824525
Transcriptional regulator, TetR family	2.480689325
Acetyltransferase	2.47860535
Radical SAM domain protein	2.45034175
Outer membrane protein	2.432701088
Possible transposase	2.364140413
Pseudouridine synthase	2.265692875
Hemolysin	2.24385648
Transposase IS116/IS110/IS902 family protein	2.22705
Thioredoxin reductase	2.127039413
IS1296 transposase protein B	2.076046
Conserved protein	2.015840725
50S ribosomal protein L34	1.992396

VWA containing CoxE family protein	1.98906125
DNA-directed RNA polymerase subunit beta	1.96381375
NA+/H+ antiporter NHAC	1.935227938
Zinc finger SWIM domain protein	1.8977675
ATPase	1.85072545
Thymidylate synthase	1.84568525
Amidohydrolase	1.83815185
Transcriptional regulator, DeoR family	1.820777925
Possible tyrosine transporter P-protein	1.76818
Anthranilate synthase component II	1.7671905
GCN5-related N-acetyltransferase	1.761517788
RfaE bifunctional protein	1.731037375
Cobalt-precorrin-4 C(11)-methyltransferase	1.72463875
Guanine-hypoxanthine permease	1.711039625
Manganese-binding protein	1.694659
Cobyric acid synthase	1.676362788
Treponema denticola	
Gene Name	meanAbund
Transcriptional regulator, TetR family	6.25098
ABC transporter, ATP-binding protein	5.42128
Ankyrin repeat protein	3.06271
Site-specific recombinases, DNA invertase Pin homologs	3.01433
Lipoprotein	2.40664
Thioredoxin	2.28700
Methyl-accepting chemotaxis protein	2.18394
ABC transporter ATP-binding protein	1.74008
MATE efflux family protein	**1.42679**
Glutathione peroxidase	1.38907
Membrane protein, putative	1.35796
ABC transporter ATP-binding protein/permease	1.29008
DNA mismatch endonuclease Vsr	1.05253
Xenobiotic-transporting ATPase	0.99221
Ribonuclease VapC	0.95592
TPR protein	0.94999
Oligopeptide/dipeptide ABC transporter, ATP-binding protein	0.91104
ABC-type multidrug transport system, ATPase and permease component	0.88981
Diguanylate cyclase (GGDEF) domain-containing protein	0.84254
Pseudouridine synthase	0.84222
YcfA family protein	0.81817
50S ribosomal protein L32	0.76883
ABC transporter related protein	0.76582

ABC transporter	0.74570
RluA family pseudouridine synthase	0.73355
Heavy metal translocating P-type ATPase	0.73148
Metallo-beta-lactamase family protein	0.72542
Histidine kinase	0.71756
Conserved domain protein	0.71313
Addiction module antitoxin, RelB/DinJ family	0.67645
RNA polymerase sigma factor	0.63557
L-lactate dehydrogenase	0.62536
RelB/DinJ family addiction module antitoxin	0.59393
Integrase catalytic region	0.58639
Chorismate mutase	0.58630
Oligopeptide transport ATP-binding protein AppD	0.58623
RpiB/LacA/LacB family sugar-phosphate isomerase	0.58315
Possible dnaK suppressor	0.57961
Cobalt transport protein	0.57135
LysM/M23/M37 peptidase	0.54945
ATPase AAA	0.54873
Prevent-host-death family protein	0.54788
Flagellar hook-basal body complex protein FliE	0.54580
MATE family transporter	**0.53897**
Signal peptidase I	0.53443
Peptidase M42 family protein	0.53376
RNA methyltransferase	0.53039
Peptidyl-prolyl cis-trans isomerase	0.52088
RHS repeat-associated core domain-containing protein	0.51748
DNA polymerase III	0.50657
Cardiobacterium valvarum	
Gene Name	meanAbund
Transposase	53.58075
Sel1 repeat protein	30.17798
Helix-turn-helix domain of resolvase (Fragment)	21.38863
IS1480b transposase	15.72224
Tetratricopeptide repeat protein	13.64956
Acetyltransferase, GNAT family	13.18181
IS5 family transposase,Transposase DDE domain	12.22235
Tat pathway signal sequence domain protein	12.04198
Peptidyl-prolyl cis-trans isomerase	11.38757
Ser/Thr phosphatase family protein	10.17345
ATPase/histidine kinase/DNA gyrase B/HSP90 domain protein	7.98308
Membrane protein	7.92403

ABC transporter, ATP-binding protein	7.59536
Acyl carrier protein	7.50314
Addiction module antitoxin, RelB/DinJ family	7.48685
SMI1 / KNR4 family protein	7.48015
Pseudouridine synthase	6.95228
DNA-binding helix-turn-helix protein	6.69018
Transglycosylase SLT domain protein	6.68673
OmpA family protein	6.57143
Efflux transporter, RND family, MFP subunit	**6.56882**
NlpC/P60 family protein	6.33083
Glyoxalase/bleomycin resistance protein/dioxygenase	6.22122
Transcriptional regulator, AraC family	6.22077
Peptidase, M48 family	6.19822
Acyltransferase	6.14974
HAD hydrolase, family IA, variant 3	5.91526
S4 domain protein	5.88368
Phosphoglycerate mutase	5.85409
CRISPR-associated endoribonuclease Cas2	5.84203
Transcriptional regulator, DeoR family	5.80208
Spermidine N(1)-acetyltransferase	5.77363
NAD dependent epimerase/dehydratase family protein	5.65739
Hydrolase, TatD family	5.64212
DnaJ domain protein	5.63168
Cof-like hydrolase	5.48318
Response regulator receiver domain protein	5.42215
LysR substrate binding domain protein	5.37501
ABC transporter ATP-binding protein	5.16074
Endonuclease III	5.10799
Band 7 protein	5.08845
Two component transcriptional regulator, winged helix family	5.02484
Exodeoxyribonuclease III	4.92089
ADP-ribose pyrophosphatase	4.91616
Peptidase propeptide and YPEB domain protein	4.86149
Carbamate kinase	4.83808
Peptidase, S54 family	4.73912
3-oxoacyl-[acyl-carrier-protein] reductase FabG	4.65901
ABC-2 type transporter	4.63549
Bacterioferritin	4.62869

Supplementary Table A - 6A-B: Top 50 genes in keystone taxa - Radcliffe museum dental calculus

Top 50 genes by gene abundance from HUMAnN2 for each keystone taxa found in the Radcliffe dental calculus samples. Abundance is gene copies per 1 million gene copies, with the mean value across the dataset reported for each gene. Antibiotic resistance genes are in bold

Treponema socranskii	
Gene Name	meanAbund
ABC-type transporter, integral membrane subunit.	18.17577
Binding-protein-dependent transport systems inner membrane component.	16.40026
ABC transporter, permease protein.	15.47453
DNA-binding helix-turn-helix protein.	15.08302
ABC transporter, solute-binding protein.	14.59235
Flavocytochrome c.	12.64163
Transcriptional regulator, TetR family.	11.83416
Tetratricopeptide repeat protein.	11.82341
MATE efflux family protein.	**11.47407**
Extracellular solute-binding protein family 1.	10.95863
Pseudouridine synthase.	10.13271
ABC transporter, ATP-binding protein.	8.95850
ABC transporter related protein.	8.76818
Methyltransferase domain protein.	8.17172
Methyl-accepting chemotaxis protein.	7.97500
Ferredoxin.	7.59735
Putative lipoprotein.	7.34340
Tripartite tricarboxylate transporter TctB family protein.	7.15727
Transcriptional regulator, DeoR family.	7.02176
Inner-membrane translocator.	6.78443
Glycosyl transferase group 1.	6.75112
Transcriptional regulator, LacI family.	6.60438
Transcriptional regulator.	6.33526
TRAP transporter, DctQ-like membrane protein.	6.03125
Response regulator receiver domain protein.	5.98242
Tripartite tricarboxylate transporter family receptor.	5.89056
Ribose import ATP-binding protein RbsA.	5.54881
FMN-binding domain protein.	5.25295
RNA polymerase sigma factor.	5.05100
ABC transporter ATP-binding protein.	4.98376
Tetratricopeptide TPR_2 repeat-containing protein.	4.96469
Beta-lactamase domain protein.	4.94554
Phosphonate-transporting ATPase.	4.78914

DEAD/DEAH box helicase domain protein.	4.78738
Elongation factor G.	4.68901
Ketose-bisphosphate aldolase.	4.67159
Amidohydrolase family protein.	4.60343
AAA domain protein.	4.60341
Na+/H+ antiporter family protein.	4.58608
PF03382 family protein.	4.53589
ABC transporter permease protein.	4.47630
OmpA family protein.	4.46184
HRDC domain protein.	4.45457
tRNA/rRNA methyltransferase (SpoU).	4.44506
Seryl-tRNA synthetase.	4.43007
ABC-3 protein.	4.30286
Transposase IS4 family protein.	4.29741
Acyl carrier protein.	4.19149
Peptidase, M23 family.	4.18883
PF04365 family protein.	4.16101
Tannerella forsythia	
Gene Name	meanAbund
Putative lipoprotein	180.16561
TonB-linked outer membrane protein, SusC/RagA family	90.41168
SusD family protein	75.76149
TonB-dependent receptor	73.71884
Tetratricopeptide repeat protein	68.89489
Sigma factor regulatory protein, FecR/PupR family	63.44114
Transposase, IS116/IS110/IS902 family	62.89150
Bacterial group 2 Ig-like protein	53.26293
RNA polymerase sigma-70 factor	52.97999
Radical SAM domain protein	49.85693
Glycosyltransferase, group 1 family protein	49.10539
Putative membrane protein	43.28465
ABC transporter ATP-binding protein	35.57676
Peptidyl-prolyl cis-trans isomerase	32.97369
Efflux ABC transporter, permease protein	32.82248
Peptidase, S41 family	30.92127
Transposase, IS4 family	30.56842
ATPase/histidine kinase/DNA gyrase B/HSP90 domain protein	30.47830
Methyltransferase domain protein	29.27074
Transposase	27.68744
Outer membrane efflux protein	25.75341
Response regulator receiver domain protein	24.45379

Transcriptional regulator, LuxR family	23.92552
Sigma-70 region 2	23.36289
Tat pathway signal sequence domain protein	21.25475
Efflux transporter, RND family, MFP subunit	**21.22168**
Transcriptional regulator, TetR family	20.84302
Outer membrane protein	20.78598
MATE efflux family protein	**20.08566**
Acyl carrier protein	19.64659
Signal peptidase I	19.55459
Arylsulfatase	18.74932
Peptidase, S9A/B/C family, catalytic domain protein	18.31288
Endonuclease/exonuclease/phosphatase family protein	17.78770
Acyltransferase	17.70165
Polysaccharide biosynthesis protein	17.67078
Antioxidant, AhpC/TSA family	17.18099
PAP2 family protein	16.63161
RHS repeat-associated core domain protein	16.38480
Polysaccharide deacetylase	15.64873
TonB-dependent receptor plug domain protein	15.60416
Repeat protein	15.01718
RNA polymerase sigma factor, sigma-70 family	14.87160
ATPase	14.72085
TIGR01200 family protein	14.42236
Pseudouridine synthase	14.01312
Glycosyltransferase, group 2 family protein	13.85109
PepSY domain protein	13.59936
Ser/Thr phosphatase family protein	13.44160
Sporulation and cell division repeat protein	13.28620

Supplementary Table A - 7: Network properties change with sample size

Mean total clusters increase with sample size, regardless of sample type.

Sample Type	Population	Mean # of Clusters: 5 Samples	Mean # of Clusters: 10 Samples	Mean # of Clusters: 20 Samples	Mean # of Clusters: All Samples
Feces	Matses (n = 25)	2.19 (sd = 0.51)	2.31 (sd = 0.73)	4.37 (sd = 1.35)	6.46 (sd = 1.50)
	Hadza (n= 26)	3.01 (sd = 1.03)	2.92 (sd = 0.51)	6.3 (sd = 1.69)	7.79 (sd = 1.65)
	China (n = 38)	2.36 (sd = 1.06)	2.76 (sd = 1.18)	4.44 (sd = 0.90)	9.77 (sd = 3.29)
	Hmp (n = 50)	3.13 (sd = 0.66)	3.42 (sd = 0.95)	4.68 (sd = 1.38)	17.03 (sd = 3.23)

Dental Calculus	Radcliffe (n = 44)	2.11 (sd = 0.33)	2.47 (sd = 0.80)	3.81 (sd = 0.62)	14.14 (sd = 3.30)

Supplementary Table A - 8: Keystone identification falters in small sample size.

Keystone found in small sample size simulations do not match the keystones identified at the full sample size. Values in each table represent the number of keystones found in the small sample size datasets that were also found in the full dataset for each method of identifying keystone taxa.

Page Rank					
	Population	n = 5	n = 10	n = 20	Full Dataset
Feces	Matses (n = 25)	0	2	2	4
	Hadza (n= 26)	0	1	3	3
	China (n = 38)	0	1	2	3
	Hmp (n = 50)	0	0	1	2
Dental Calculus	Radcliffe (n = 44)	0	0	0	1
Hub Score					
	Population	n = 5	n = 10	n = 20	Full Dataset
Feces	Matses (n = 25)	0	3	3	5
	Hadza (n= 26)	0	1	3	4
	China (n = 38)	0	1	2	3
	Hmp (n = 50)	0	0	1	2
Dental Calculus	Radcliffe (n = 44)	0	0	1	3
Closeness Centrality					
	Population	n = 5	n = 10	n = 20	Full Dataset
Feces	Matses (n = 25)	0	2	2	4
	Hadza (n= 26)	0	1	3	3
	China (n = 38)	0	0	1	2
	Hmp (n = 50)	0	0	1	2
Dental Calculus	Radcliffe (n = 44)	0	0	1	2

Supplementary Material B

Co-Authors, affiliations, contributions, supplementary figures, and supplementary tables for Chapter 3: Non-Industrial Gut Microbiomes Provide a More Resilient Ecology for Short-Chain Fatty Acid Production; adapted from Jacobson et al. in review. Non-Industrial Gut Microbiomes Provide a More Resilient Ecology for Short-Chain Fatty Acid Production. *Scientific Reports*.

Author List and Affiliations

David K. Jacobson,[1,2] Tanvi P. Honap,[1,2] Andrew T. Ozga,[3] Nicolas Meda,[4] Thérèse S. Kagoné,[5] Hélène Carabin,[6] Paul Spicer,[2,7] Raul Y. Tito,[2] Alexandra J. Obregon-Tito,[2] Luis Marin Reyes,[8] Luzmila Troncoso-Corzo,[9] Emilio Guija-Poma,[10] Krithi Sankaranarayanan,[1,11] and Cecil M. Lewis, Jr.[1,2]

[1]Laboratories of Molecular Anthropology and Microbiome Research, University of Oklahoma, Norman, OK, USA 73019.

[2]Department of Anthropology, University of Oklahoma, Norman, OK, USA 73019.

[3]Halmos College of Natural Sciences and Oceanography, Nova Southeastern University, Fort Lauderdale, FL, USA 33314.

[4]Ministry of Health, Ouagadougou, Burkina Faso.

[5]Centre MURAZ Research Institute, Bobo-Dioulasso, Burkina Faso.

[6]Département de Pathologie et de Microbiologie, Faculté de Médecine vétérinaire-Université de Montréal, Saint-Hyacinthe, Canada, QC J2S 2M2.

[7]Center for Applied Social Research, University of Oklahoma, Norman, OK, USA 73019.

[8] Centro Nacional de Salud Publica, Instituto Nacional de Salud, Lima, Perú.

[9] Facultad de Medicina, Universidad Nacional Mayor de San Marcos, Lima, Perú.

[10]Centro de Investigación de Bioquímica y Nutrición, Facultad de Medicina Humana, Universidad de San Martín de Porres, Lima, Perú.

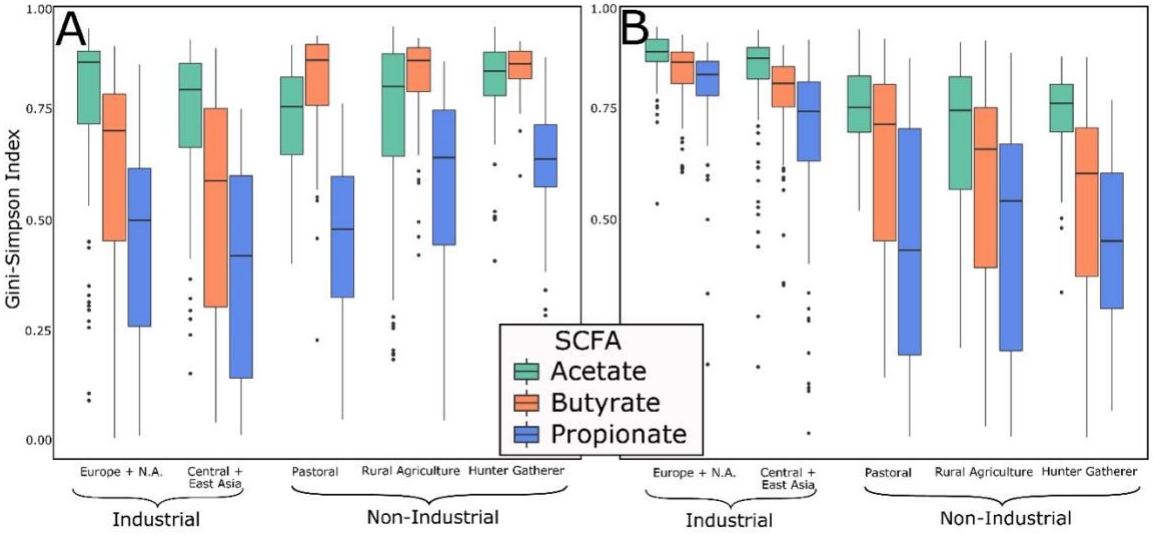
[11] Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019.

Authors' Contributions and Acknowledgements

Author contributions: NM, TSK, HC, PS, LMR, LTC, EGP, and CML conceived the project and provided funding for collection and analysis of the Peruvian and Burkina Faso microbiome samples analyzed in this manuscript. DKJ, ATO, TSK, AJOT, and RYT collected and processed the Burkina Faso and/or Peruvian gut microbiome samples analyzed in this paper. THP and KS bioinformatically processed and prepared the comparative datasets used in this manuscript. DKJ performed statistical analysis and DKJ, THP, and CML conceived and wrote the manuscript.

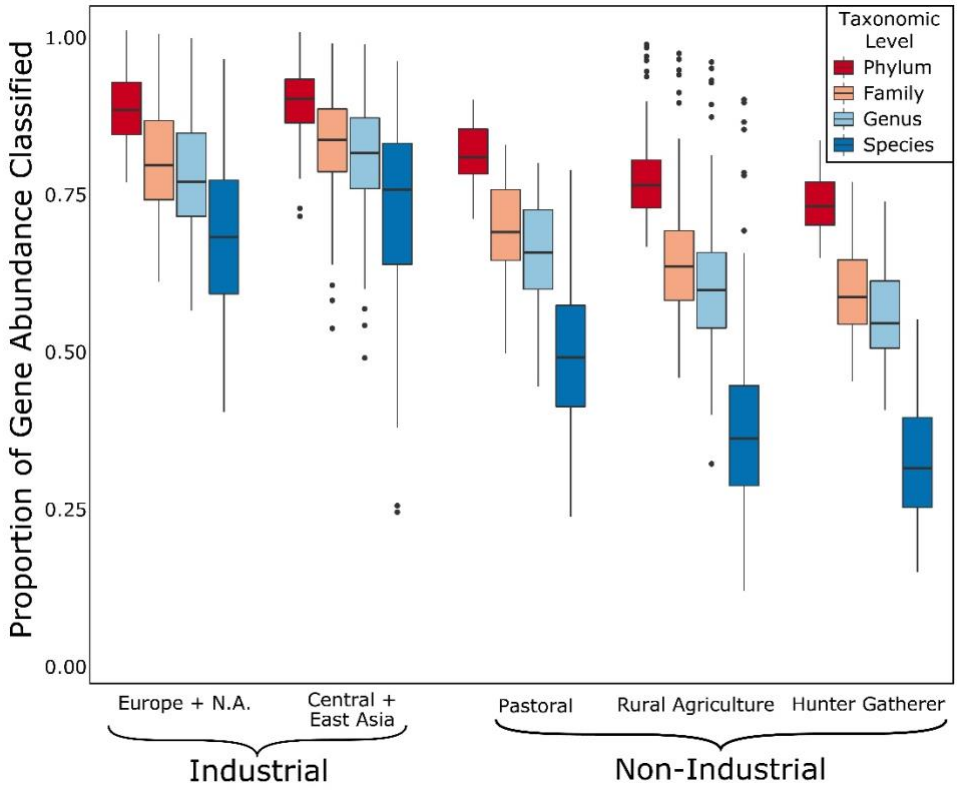
Acknowledgements: We would like to thank all participants who donated samples analyzed in this study. We would also like to thank Dr. Amadou Dicko, Issé Rouamba, Bachirou Tinto, and Alidou Zongo from Centre Muraz Research Institute in Bobo-Dioulasso, Burkina Faso for their assistance in collection and processing of the microbiome samples from Burkina Faso.

Supplementary Figures B: 1-3



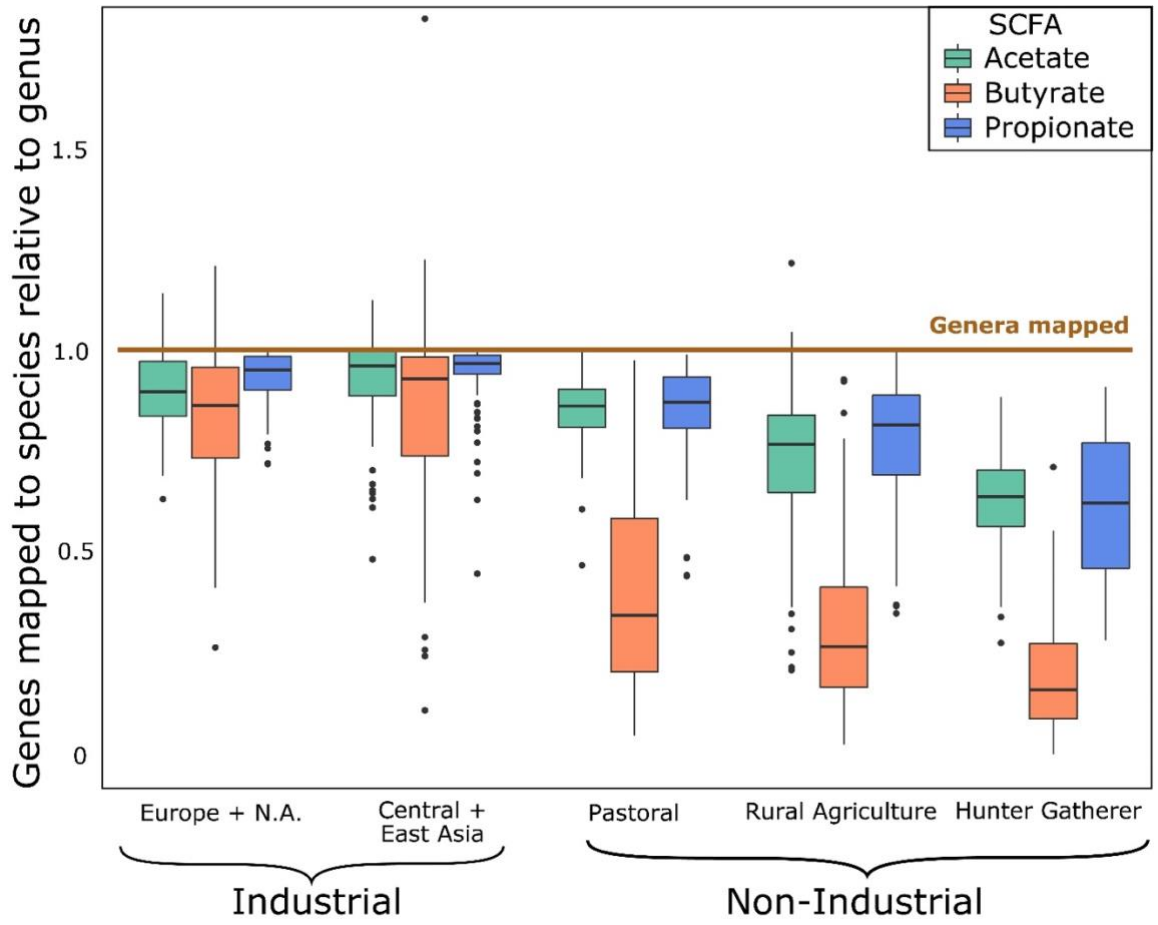
Supplementary Figure B - 1: Gini-Simpson Index Values for Taxa Encoding SCFAs

*Genus (**A**) and species (**B**) level for each SCFA of interest. A) The GS index for butyrate and propionate are higher in the rural agriculturalists and hunter-gatherers compared to the industrial populations at the genus level (FDR-adjusted p-value < 0.003, n = 451). B) Each SCFA has significantly lower GS values at the species level in non-industrial populations (FDR-adjusted p-value < 0.05, n = 451). Statistical comparisons were generated using the Kruskal-Wallis H test and the post-hoc Dunn Test. False discovery rate (FDR) was used to account for multiple testing.*



Supplementary Figure B - 2: Proportion of All Genes Classified to A Taxon at Different Phylogenetic Levels

Classification is significantly worse in the non-industrial populations compared to the industrial populations (FDR-adjusted p-value < 5 x 10^{-5}, n = 451). Statistical comparisons were generated using the Kruskal-Wallis H test and the post-hoc Dunn Test. False discovery rate (FDR) was used to account for multiple testing.



Supplementary Figure B - 3: Genus:Species Relative Mapping Index.

Relative index of genes mapped to a taxon at the species level (each box) normalized to genes mapped to a taxon at the genus level (1.0). In the non-industrial populations, there is a significant drop-off in the genes mapped to a taxon at the species level for each gene (FDR-adjusted p-value < 0.001, n = 451), while in industrial populations the rate of mapping is similar at the genus and species level (FDR-adjusted p-value > 0.05, n = 210). Values above 1.0 are due to genes that map to taxa at the species level but not the genus level. This is the result of a few species that are annotated at the species level but have a missing annotation at the genus level, such as candidate species that do not have a complete phylogeny. In our dataset, this was primarily species belonging to the Lachnospiraceae family that do not have a finalized/approved

genus. This is more common in the industrial populations in our dataset, once again likely due to bias that favors for industrial datasets. Statistical comparisons were generated using the Kruskal-Wallis H test and the post-hoc Dunn Test. False discovery rate (FDR) was used to account for multiple testing.

Supplementary Tables B: 1-5

Supplementary Table B - 1: Genera Involved in SCFA Synthesis

Genera and pathways previously identified in SCFA production.

SCFA	Genus	Pathway (gene)	Study
Acetate	Bacteroides	*ackA*	Rios-Coivan et al. 2016
	Bifidobacterium	*ackA*	Fukuda et al. 2012
	Diverse Sets of Baceria	*ackA*	Rios-Coivan et al. 2016, Venegas et al. 2019
Butyrate	Alistipes	not specified	Vital et al. 2014
	Anaerostipes	*but*	Louis and Flint 2009, 2017
	Bacteroides	*buk*	Louis and Flint 2009, 2017
	Clostridium	*but, buk*	Louis and Flint 2009, Vital et al. 2014
	Coprococcus	*but, buk*	Louis and Flint 2009, 2017
	Eubacterium	*but*	Louis and Flint 2009, 2017
	Faecalibacterium	*but*	Louis and Flint 2009, 2017
	Megasphaera	*but*	Louis and Flint 2009
	Odoribacter	not specified	Vital et al. 2014
	Roseburia	*but*	Louis and Flint 2009, 2017
	Subdoligranulum	*buk*	Louis and Flint 2009, 2017
Propionate	Akkermansia	*mmdA*	Reichardt et al. 2014, Louis and Flint 2017
	Alistipes	*mmdA*	Louis and Flint 2017
	Bacteroides	*mmdA*	Reichardt et al. 2014, Louis and Flint 2017
	Blautia	*pduP*	Louis and Flint 2017
	Clostridium	*mmdA, lcdA, pduP*	Reichardt et al. 2014
	Coprococcus	*lcdA*	Reichardt et al. 2014, Louis and Flint 2017
	Dialister	*mmdA*	Reichardt et al. 2014, Louis and Flint 2017
	Eubacterium	*pduP*	Reichardt et al. 2014, Louis and Flint 2017
	Megasphaera	*mmdA, lcdA*	Reichardt et al. 2014
	Phascolarctobacterium	*mmdA*	Reichardt et al. 2014, Louis and Flint 2017
	Prevotella	*mmdA*	Louis and Flint 2017
	Roseburia	*pduP*	Reichardt et al. 2014
	Ruminococcus	*pduP*	Reichardt et al. 2014
	Selenomonas	*mmdA*	Reichardt et al. 2014
	Veillonella	*mmdA*	Reichardt et al. 2014

Pathway Name	Abbreviated Name
acetate kinase	*ackA*
butyryl-CoA:acetate CoA transferase	*but*
butyrate kinase	*buk*
succinate	*mmdA*
acrylate	*lcdA*
propanediol	*pduP*

Supplementary Table B - 2: Samples Used in SCFA Analysis

SampleName	RunAccession	Population	LifestyleGeneral	LifestyleSpecific	Age	Sex
bftm0101	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	55	male
bftm0102	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	39	female
bftm0103	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	18	male
bftm0201	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	52	male
bftm0202	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	46	female
bftm0203	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	21	male
bftm0301	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	61	male
bftm0302	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	20	female
bftm0303	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	19	male
bftm0401	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	69	male
bftm0402	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	52	female
bftm0403	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	18	male
bftm0501	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	41	male
bftm0502	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	19	male
bftm0504	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	11	female
bftm0601	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	66	male
bftm0602	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	21	male
bftm0604	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	49	female
bftm0701	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	70	male

bftm0703	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	20	male
bftm0704	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	41	female
bftm0801	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	79	male
bftm0802	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	31	female
bftm0803	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	39	male
bftm0902	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	32	female
bftm0903	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	18	male
bftm1001	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	55	male
bftm1002	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	20	female
bftm1003	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	5	male
bftm1004	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	23	male
bftm1101	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	81	male
bftm1102	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	18	male
bftm1104	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	40	female
bftm1201	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	87	male
bftm1202	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	27	female
bftm1204	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	20	male
bftm1301	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	53	male
bftm1303	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	39	female
bftm1401	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	77	male
bftm1402	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	7	female

bftm1403	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	40	female
bftm1404	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	28	male
bftm1501	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	74	male
bftm1502	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	33	male
bftm1503	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	40	female
bftm1601	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	75	male
bftm1602	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	26	female
bftm1603	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	22	male
bftm1701	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	71	male
bftm1702	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	35	female
bftm1802	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	25	female
bftm1803	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	18	male
bftm1901	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	67	male
bftm1902	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	25	female
bftm1903	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	22	male
bftm2001	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	59	male
bftm2002	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	23	male
bftm2003	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	37	female
bftm2101	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	67	male
bftm2102	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	34	female
bftm2103	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	9	female

bftm2104	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	20	male
bftm2201	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	71	male
bftm2202	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	28	female
bftm2203	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	29	male
bftm2204	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	10	male
bftm2301	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	44	male
bftm2302	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	51	female
bftm2303	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	36	male
bftm2304	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	10	male
bftm2401	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	49	male
bftm2501	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	75	male
bftm2503	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	38	female
bftm2504	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	39	male
bftm2601	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	48	male
bftm2602	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	26	female
bftm2604	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	17	male
bftm2701	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	71	male
bftm2702	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	27	female
bftm2704	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	18	male
bftm2801	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	69	male
bftm2803	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	18	male

bftm2804	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	21	female
bftm2901	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	73	male
bftm2902	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	20	male
bftm2904	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	27	female
bftm3001	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	65	male
bftm3002	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	18	male
bftm3003	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	35	female
bftm3004	N/A	BurkinaFaso	nonIndustrial	ruralAgriculture	7	male
bgi.N075A	SRR413615	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NLF002	SRR341617	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NLF005	SRR341618	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NLF006	SRR341619	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NLF007	SRR341620	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NLF009	SRR341621	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NLF010	SRR341622	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NLF011	SRR341623	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NLF014	SRR341624	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NLF015	SRR341693	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NLM006	SRR341696	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NLM010	SRR341698	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NLM015	SRR341630	China	Industrial	Central/East Asian Industrial	N/A	N/A

bgi.NLM016	SRR341631	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NLM022	SRR341700	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NLM023	SRR341633	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NLM027	SRR341703	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NLM028	SRR341704	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NLM029	SRR341705	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NLM031	SRR341706	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NOF002	SRR341636	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NOF005	SRR341708	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NOF008	SRR341709	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NOF009	SRR341640	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NOF012	SRR341711	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NOF013	SRR341642	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NOF014	SRR341643	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NOM001	SRR341712	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NOM004	SRR341644	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NOM007	SRR341645	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NOM009	SRR341715	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NOM017	SRR341720	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NOM018	SRR341649	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NOM019	SRR341721	China	Industrial	Central/East Asian Industrial	N/A	N/A

bgi.NOM020	SRR341722	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NOM023	SRR341651	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NOM027	SRR341724	China	Industrial	Central/East Asian Industrial	N/A	N/A
bgi.NOM028	SRR341725	China	Industrial	Central/East Asian Industrial	N/A	N/A
CNA.NO1	SRR1761676	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO10	SRR1761684	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO11	SRR1761685	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO12	SRR1761686	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO13	SRR1761687	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO14	SRR1761688	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO15	SRR1761689	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO16	SRR1761690	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO17	SRR1761691	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO18	SRR1761692	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO19	SRR1761693	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO2	SRR1761677	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO20	SRR1761694	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO21	SRR1761695	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO22	SRR1761696	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO23	SRR1761697	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO3	SRR1761678	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A

CNA.NO4	SRR1761679	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO5	SRR1761680	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO6	SRR1761681	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
CNA.NO8	SRR1761682	Norman, Oklahoma, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
Had1929408	SRR1929408	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1929484	SRR1929484	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1929485	SRR1929485	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1929563	SRR1929563	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1929574	SRR1929574	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930121	SRR1930121	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930122	SRR1930122	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930123	SRR1930123	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930128	SRR1930128	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930132	SRR1930132	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930133	SRR1930133	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930134	SRR1930134	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930136	SRR1930136	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930138	SRR1930138	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930140	SRR1930140	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930141	SRR1930141	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930142	SRR1930142	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A

Had1930143	SRR1930143	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930144	SRR1930144	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930145	SRR1930145	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930149	SRR1930149	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930176	SRR1930176	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930177	SRR1930177	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930179	SRR1930179	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930187	SRR1930187	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
Had1930244	SRR1930244	Hadza, Tanzania	nonIndustrial	hunterGather	N/A	N/A
HMP_700013715	SRR059412	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700014562	SRR060371	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700014724	SRR062103	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700014837	SRR060443	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700015113	SRR059421	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700015181	SRR059855	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700015250	SRR061170	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700015415	SRR060411	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A

HMP_700 015857	SRR0598 97	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 015922	SRR0593 67	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 015981	SRR0611 64	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 016142	SRR0624 26	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 016456	SRR0624 28	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 016542	SRR0598 13	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 016610	SRR0611 53	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 016765	SRR0603 75	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 016960	SRR0593 39	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 021306	SRR0623 23	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 021824	SRR0615 58	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 021876	SRR0623 88	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 021902	SRR0634 69	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 023113	SRR0611 38	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A

HMP_700 023267	SRR0623 60	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 023337	SRR0611 35	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 023578	SRR0613 68	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 023634	SRR0611 56	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 023720	SRR0611 38	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 023845	SRR0611 47	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 023872	SRR0613 34	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 023919	SRR0634 81	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 024024	SRR0635 89	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 024233	SRR0612 31	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 024318	SRR0612 09	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 024437	SRR0612 08	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 024449	SRR0635 55	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 024509	SRR0614 97	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A

HMP_700 024615	SRR0625 24	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 024673	SRR0611 97	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 024711	SRR0612 00	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 024752	SRR0611 72	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 024866	SRR0625 39	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 024930	SRR0614 95	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 024998	SRR0611 61	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 032222	SRR0600 26	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 032244	SRR0600 25	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 032338	SRR0621 00	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 032944	SRR0598 31	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 033153	SRR0611 45	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 033435	SRR0619 19	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A
HMP_700 033502	SRR0593 73	Human Microbiome Project, USA	Industrial	Europe/N.A. Industrial	N/A	N/A

Ind118	ERR260250	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind121	ERR260251	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind126	ERR260252	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind127	ERR260253	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind137	ERR260268	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind141	ERR260255	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind143	ERR260256	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind146	ERR260258	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind153	ERR260259	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind155	ERR260260	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind187	ERR260263	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind193	ERR260264	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind194	ERR260265	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind195	ERR260266	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind198	ERR260267	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind224	ERR260230	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind225	ERR260231	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind231	ERR260234	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind264	ERR260242	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind268	ERR260243	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind269	ERR260244	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A

Ind288	ERR260246	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind365	ERR260147	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind381	ERR260153	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind424	ERR260163	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind451	ERR260170	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind456	ERR260171	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind463	ERR260175	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind479	ERR260180	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind53	ERR260193	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind58	ERR260204	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind584	ERR260205	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind59	ERR260209	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind614	ERR260215	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind617	ERR260216	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind618	ERR260217	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind628	ERR260218	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind643	ERR260221	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind655	ERR260223	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind77	ERR260224	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind80	ERR260225	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
Ind88	ERR260226	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A

Ind92	ERR260227	Northern European	Industrial	Europe/N.A. Industrial	N/A	N/A
japan_apr01S00	DRR042264	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr02S00	DRR042272	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr03S00	DRR042280	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr05S00	DRR042288	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr09S00	DRR042304	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr10S00	DRR042312	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr11S00	DRR042316	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr12S00	DRR042320	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr15S00	DRR042328	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr16S00	DRR042332	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr17S00	DRR042340	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr18S00	DRR042348	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr19S00	DRR042352	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr21S00	DRR042356	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr22S00	DRR042360	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr23S00	DRR042364	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr36S00	DRR042390	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr37S00	DRR042394	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr38S00	DRR042398	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_apr39S00	DRR042402	Japan	Industrial	Central/East Asian Industrial	N/A	N/A

japan_apr40S00	DRR042410	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_FAKO02	DRR042420	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_FAKO03	DRR042424	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_FAKO05	DRR042428	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_FAKO08	DRR042436	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_FAKO15	DRR042452	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_FAKO19	DRR042462	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_FAKO22	DRR042468	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_FAKO23	DRR042472	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_FAKO27	DRR042482	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_FAKO29	DRR042488	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
japan_FTAG01	DRR042591	Japan	Industrial	Central/East Asian Industrial	N/A	N/A
madSRS3638571	SRR7658688	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638572	SRR7658687	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638573	SRR7658690	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638574	SRR7658689	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638575	SRR7658685	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638576	SRR7658684	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638577	SRR7658682	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638578	SRR7658683	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638579	SRR7658686	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A

madSRS3638580	SRR7658681	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638581	SRR7658679	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638582	SRR7658678	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638583	SRR7658676	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638584	SRR7658677	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638585	SRR7658675	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638586	SRR7658673	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638587	SRR7658672	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638589	SRR7658669	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638591	SRR7658666	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638592	SRR7658664	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638593	SRR7658665	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638632	SRR7658680	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638638	SRR7658674	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638642	SRR7658671	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638652	SRR7658606	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638653	SRR7658668	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638654	SRR7658605	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638655	SRR7658604	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638656	SRR7658603	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638657	SRR7658601	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A

madSRS3638658	SRR7658600	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638659	SRR7658602	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638660	SRR7658598	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638661	SRR7658599	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638662	SRR7658597	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638663	SRR7658596	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638665	SRR7658595	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638666	SRR7658593	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638667	SRR7658594	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638668	SRR7658590	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638669	SRR7658592	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638670	SRR7658591	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638671	SRR7658588	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638672	SRR7658587	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638673	SRR7658586	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638674	SRR7658589	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638675	SRR7658585	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638676	SRR7658583	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638677	SRR7658584	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
madSRS3638678	SRR7658582	Madagascar	nonIndustrial	ruralAgriculture	N/A	N/A
SAMEA4545280	SAMEA4545280	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A

SAMEA4545282	SAMEA4545282	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545284	SAMEA4545284	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545288	SAMEA4545288	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545296	SAMEA4545296	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545308	SAMEA4545308	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545326	SAMEA4545326	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545332	SAMEA4545332	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545340	SAMEA4545340	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545360	SAMEA4545360	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545362	SAMEA4545362	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545364	SAMEA4545364	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545372	SAMEA4545372	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545380	SAMEA4545380	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545388	SAMEA4545388	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545394	SAMEA4545394	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545396	SAMEA4545396	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545400	SAMEA4545400	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545406	SAMEA4545406	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545412	SAMEA4545412	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545420	SAMEA4545420	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545422	SAMEA4545422	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A

SAMEA4545430	SAMEA4545430	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545434	SAMEA4545434	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545446	SAMEA4545446	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SAMEA4545450	SAMEA4545450	Kazakhstan	Industrial	Central/East Asian Industrial	N/A	N/A
SM01	SRR1761698	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM02	SRR1761699	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM03	SRR1761700	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM05	SRR1761701	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM10	N/A	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM11	SRR1761702	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM18	SRR1761703	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM20	SRR1761704	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM23	SRR1761705	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM24	SRR1761706	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM25	SRR1761707	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM28	SRR1761708	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM29	SRR1761709	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM30	SRR1761710	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM31	SRR1761711	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM32	SRR1761712	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM33	SRR1761713	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A

SM34	SRR1761714	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM37	SRR1761715	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM39	SRR1761716	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM40	SRR1761717	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM41	SRR1761718	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM42	SRR1761719	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM43	SRR1761720	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SM44	SRR1761721	Matses, Peru	nonIndustrial	hunterGather	N/A	N/A
SRR3992955	SRR3992955	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992956	SRR3992956	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992957	SRR3992957	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992958	SRR3992958	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992959	SRR3992959	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992960	SRR3992960	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992961	SRR3992961	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992962	SRR3992962	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992963	SRR3992963	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992964	SRR3992964	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992965	SRR3992965	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992966	SRR3992966	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992967	SRR3992967	Mongolia	nonIndustrial	pastoral	N/A	N/A

SRR3992968	SRR3992968	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992969	SRR3992969	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992970	SRR3992970	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992971	SRR3992971	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992972	SRR3992972	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992973	SRR3992973	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992974	SRR3992974	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992975	SRR3992975	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992976	SRR3992976	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992977	SRR3992977	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992978	SRR3992978	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992979	SRR3992979	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992980	SRR3992980	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992981	SRR3992981	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992982	SRR3992982	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992983	SRR3992983	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992984	SRR3992984	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992985	SRR3992985	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992986	SRR3992986	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992987	SRR3992987	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992988	SRR3992988	Mongolia	nonIndustrial	pastoral	N/A	N/A

SRR3992989	SRR3992989	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992990	SRR3992990	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992991	SRR3992991	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992992	SRR3992992	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992993	SRR3992993	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992994	SRR3992994	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992995	SRR3992995	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992996	SRR3992996	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992997	SRR3992997	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992998	SRR3992998	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3992999	SRR3992999	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3993000	SRR3993000	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3993001	SRR3993001	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3993002	SRR3993002	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3993003	SRR3993003	Mongolia	nonIndustrial	pastoral	N/A	N/A
SRR3993004	SRR3993004	Mongolia	nonIndustrial	pastoral	N/A	N/A

Supplementary Table B - 3: Proportional Contribution to Total SCFA Gene Abundance.

The total SCFA abundance is contributed in the following ratio 0.600:0.215:0.184 for Acetate:Butyrate:Propionate. This follows an expected ratio of 0.6:0.2:0.2 reported in previous studies.

	Acetate (proportion of SCFA)	Butyrate (proportion of SCFA)	Propionate (proportion of SCFA)
Europe/N.A Industrial	0.580 (0.001)	0.244 (0.000)	0.176 (0.001)
Central/East Asia Industrial	0.615 (0.003)	0.191 (0.003)	0.195 (0.003)
Pastoral	0.593 (0.001)	0.235 (0.001)	0.172 (0.000)
Rural Agriculture	0.615 (0.002)	0.191 (0.002)	0.195 (0.002)
Hunter Gatherer	0.588 (0.000)	0.244 (0.000)	0.168 (0.000)
Total	0.600 (0.001)	0.215 (0.001)	0.184 (0.001)

Supplementary Table B - 4: Median richness of selected SCFA-producing genera.

Number of species found to encode each respective SCFA within different genera of interest. Bacteroides and Clostridium are at high abundance in industrial gut microbiomes and have many species within each genus, which drives of species richness in industrial populations.

	Population	Bacteroides	Clostridium	Coprococcus	Prevotella	Faecalibacterium	Phascolarctobacterium
Acetate Species Richness	**Europe/N.A Industrial**	9 (sd = 2.99)	3 (2.55)	2 (0.83)	0 (0.57)	2 (0.09)	0 (0.26)
	Central/East Asia Industrial	8.5 (3.83)	4 (2.79)	1 (0.88)	0 (0.79)	1 (0.14)	0 (0.39)
	Pastoral	5.5(3.43)	0 (1.49)	2 (0.75)	2 (0.27)	2 (0.00)	1 (0.46)
	Rural Agriculture	1 (2.98)	1 (1.00)	2 (0.81)	2 (0.32)	2 (0.00)	1 (0.45)
	Hunter Gatherer	1 (2.66)	0 (0.61)	2 (0.83)	2 (0.34)	2 (0.00)	1 (0.24)
Butyrate Species Richness	**Europe/N.A Industrial**	9 (3.00)	1 (1.31)	1 (0.71)	0 (0.28)	0 (0.00)	0 (0.00)
	Central/East Asia Industrial	8 (3.73)	2 (1.49)	1 (0.82)	0 (0.44)	0 (0.00)	0 (0.00)
	Pastoral	6 (3.37)	0 (0.81)	1 (0.58)	1 (0.30)	0 (0.00)	0 (0.00)
	Rural Agriculture	1 (2.89)	1 (0.87)	1 (0.78)	1 (0.35)	0 (0.00)	0 (0.00)
	Hunter Gatherer	0 (2.17)	0 (0.46)	1 (0.83)	1 (0.28)	0 (0.00)	0 (0.00)
Propionate Species Richness	**Europe/N.A Industrial**	9 (2.99)	0 (0.00)	1 (0.47)	0 (0.36)	0 (0.00)	0 (0.00)
	Central/East Asia Industrial	9 (3.96)	0 (0.00)	1 (0.49)	0 (0.49)	0 (0.00)	0 (0.00)
	Pastoral	6 (3.40)	0 (0.00)	1 (0.48)	1 (0.14)	0 (0.00)	0 (0.00)
	Rural Agriculture	1 (3.08)	0 (0.00)	1 (0.32)	1 (0.12)	0 (0.00)	0 (0.00)
	Hunter Gatherer	0 (2.33)	0 (0.00)	1 (0.27)	1 (0.00)	0 (0.00)	0 (0.00)

Supplementary Material C

Co-Authors, affiliations, contributions, supplementary figures, and supplementary tables for Chapter 4: Shifts in Gut and Vaginal Microbiomes Associated with Platinum-Free Interval Length in Women with Ovarian Cancer; adapted from Jacobson et al. (Submitted). Shifts in Gut and Vaginal Microbiomes Associated with Platinum-Free Interval Length in Women with Ovarian Cancer. *PeerJ.*

Author List and Affiliations

David K. Jacobson[1,2], Kathleen N. Moore[3], Camille C. Gunderson[3], Michelle R. Rowland[3,4], Rita M. Austin[1,2], Tanvi P. Honap[1,2], Jiawu Xu[5,6] , Christina G. Warinner[7], Kritivasan Sankaranarayanan[2,8], Cecil M. Lewis, Jr[1,2]

[1] Department of Anthropology, University of Oklahoma, Norman, OK, United States
[2] Laboratories of Molecular Anthropology and Microbiome Research, University of Oklahoma, Norman, OK, United States
[3] Stephenson Cancer Center, University of Oklahoma Health Sciences Center, Oklahoma City, OK, United States
[4] Saint Luke's Hospital of Kansas City, Kansas City, MO, United States
[5] Ragon Institute of MGH, MIT and Harvard, Massachusetts General Hospital, Cambridge, MA, United States
[6] Harvard Medical School, Boston, MA, United States
[7] Department of Anthropology, Harvard University, Cambridge, MA, United States
[8] Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, United States

Authors' Contributions and Acknowledgements

Contributions

KNM, CW, KS, and CML contributed to the conception and design of this study. KNM, CCG, and MRR contributed to the acquisition of samples and metadata. DKJ, RMA, and XJ performed wet lab work. DKJ performed bioinformatic analysis. DKJ, KNM, TH, KS, and CML

contributed to interpretation of data and drafting the article. All authors contributed to revising the article and approved the final version to be published.

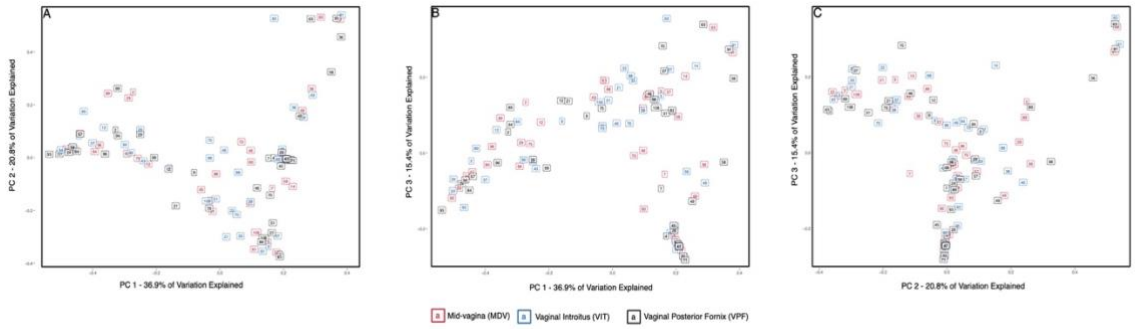
Acknowledgements

The authors would like to thank Sarah Cooper and Cathy Birdsong for their work in coordinating patient recruitment and monitoring sample collection. The authors would also like to thank all individuals who participated in this study.

Escherichia Origin

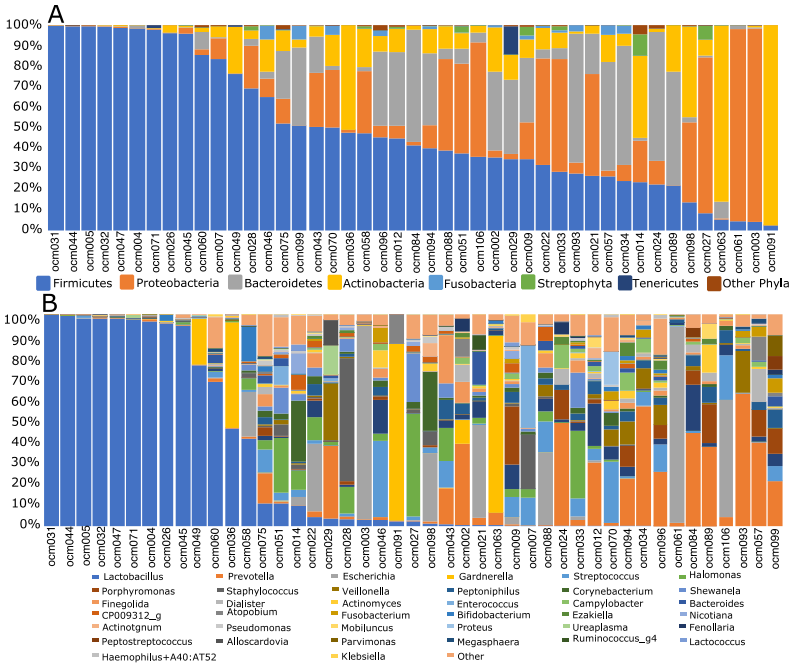
Although *Escherichia* is commonly found in the vaginal microbiome, it has not been reported at such high levels (>50% relative abundance) before. Because *Escherichia* is a common lab-grown bacterium and is often found in the gut microbiome (293-295) we further investigated whether the *Escherichia* abundance in our samples was due to sample contamination. We did not find evidence to support *Escherichia* contamination as the source of *Escherichia* observed in this study. First, our extraction and PCR blanks sequenced in this study report a low number of total reads and the percent of extraction/PCR blank sequencing reads mapping to *Escherichia* compared to the average read depth in this study is minimal (median = 0.0052%, range 0 – 3%, Supplementary Table C - 2). The very low number of *Escherichia* reads in our negatives and blanks indicates any contribution of lab/environmental *Escherichia* to the high relative abundance in our vaginal samples is miniscule. Second, we observe a weak relationship between abundance of *Escherichia* in the gut and vaginal microbiomes ($R^2 = 0.0927$, Supplementary Figure C - 9). If fecal contamination of vaginal swabs was an issue, we would expect high fecal *Escherichia* abundance to co-occur with high vaginal *Escherichia* abundance, but this is not the case.

Supplementary Figures C



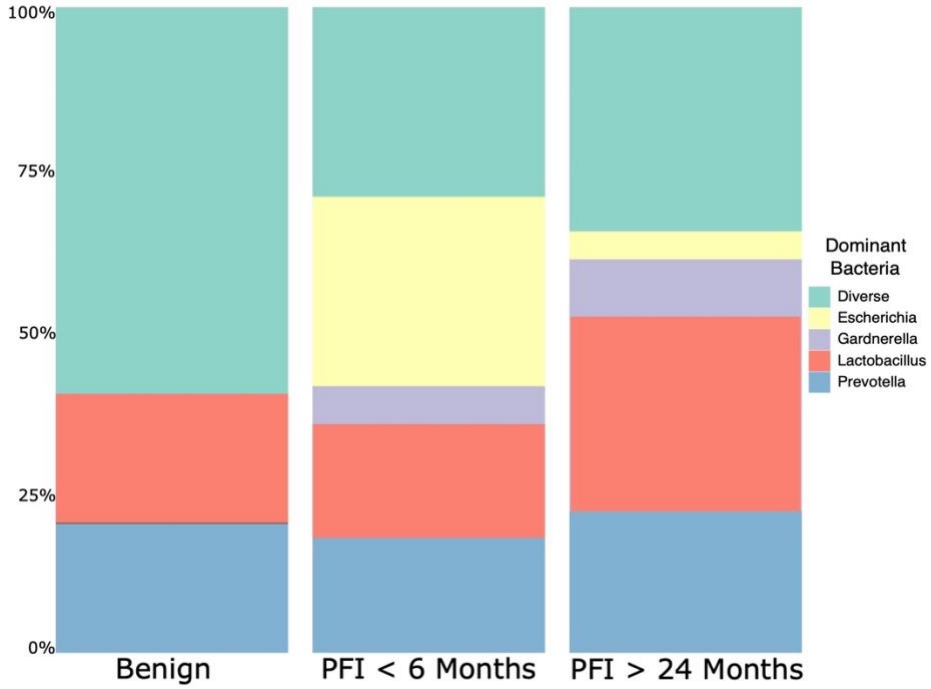
Supplementary Figure C - 1A-C: Weighted UniFrac beta diversity of all vaginal microbiome samples

Samples originating from the same individual had similar taxonomic composition. This informed our decision to combine data from the three vaginal samples per individual into a single sample per individual. Numbers within boxes represent sample ID.



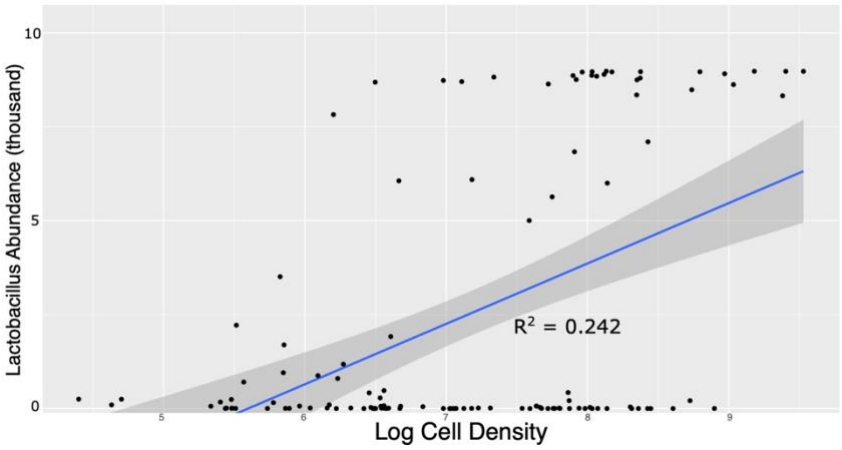
Supplementary Figure C - 2A-B: Proportional contribution of most abundant phyla and genera in the vaginal microbiome in this study.

Contributions from low abundance phyla (A) and genera (B) were combined into 'Other'. Overall, the vaginal microbiome is dominated by the common vaginal bacteria: Firmicutes, Proteobacteria, and Bacteroidetes at the phylum level, and Lactobacillus, Prevotella, Escherichia, Gardnerella, and Streptococcus at the genus level.



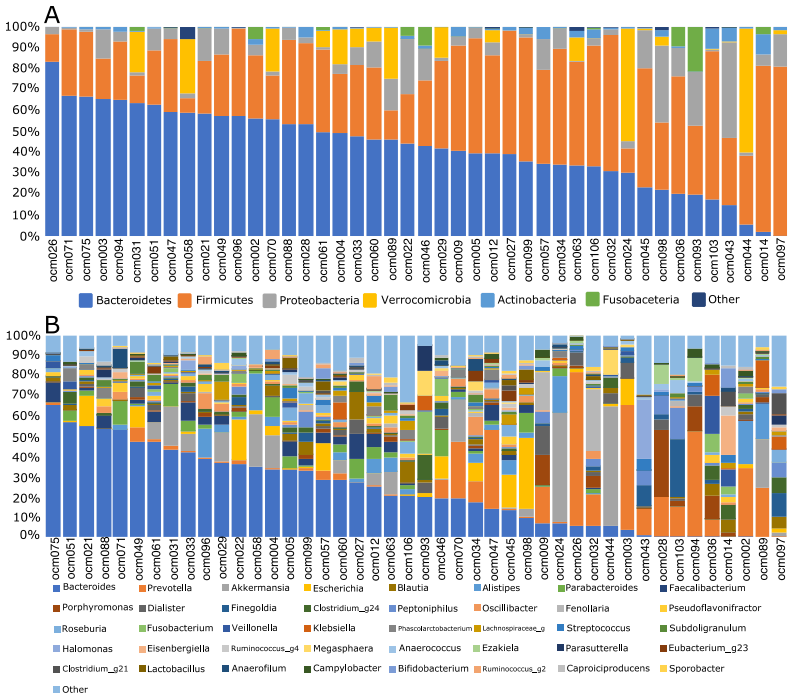
Supplementary Figure C - 3: Stacked bar chart of the proportion of samples within each study group that were dominated by different taxa

Escherichia-dominated communities are overrepresented in patients with PFI < 6 months.



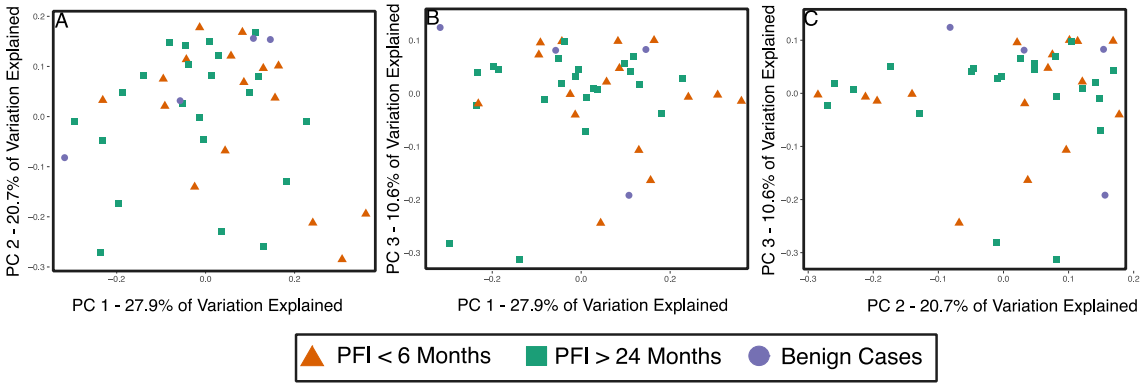
Supplementary Figure C - 4: Lactobacillus abundance has a positive association with log-transformed cell density in each sample.

This means Lactobacillus dominant communities have more bacterial cells in the vaginal ecosystem, compared to microbiomes with less Lactobacillus. Standard error is outlined in grey.

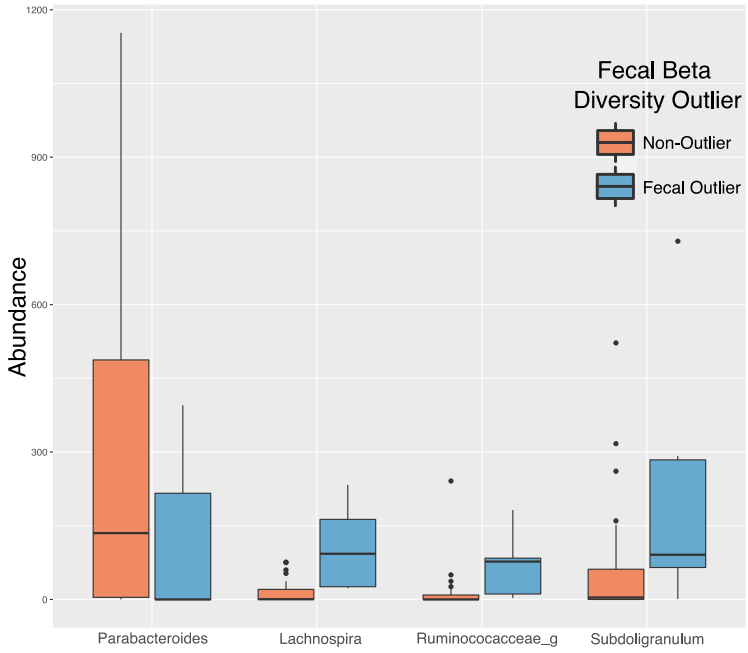


Supplementary Figure C - 5A-B: Proportional contribution of most abundant phyla and genera in the gut microbiome in this study.

Contributions from low abundance phyla (A) and genera (B) were combined into 'Other'. Overall, the gut microbiome is dominated by the expected bacteria: Bacteroidetes, Firmicutes, and Proteobacteria, at the phylum level, and Bacteroides, Prevotella, and Akkermansia at the genus level.

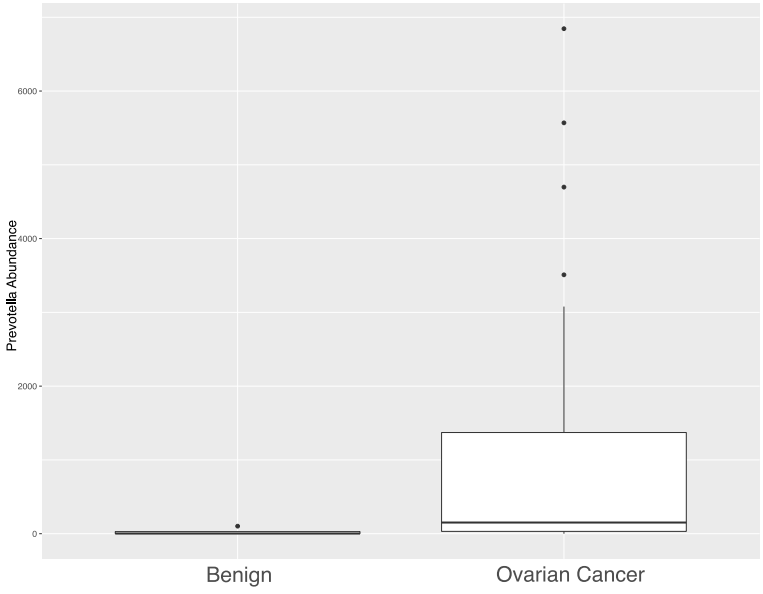


Supplementary Figure C - 6A-C: Weighted UniFrac beta diversity for gut microbiome samples in this study.



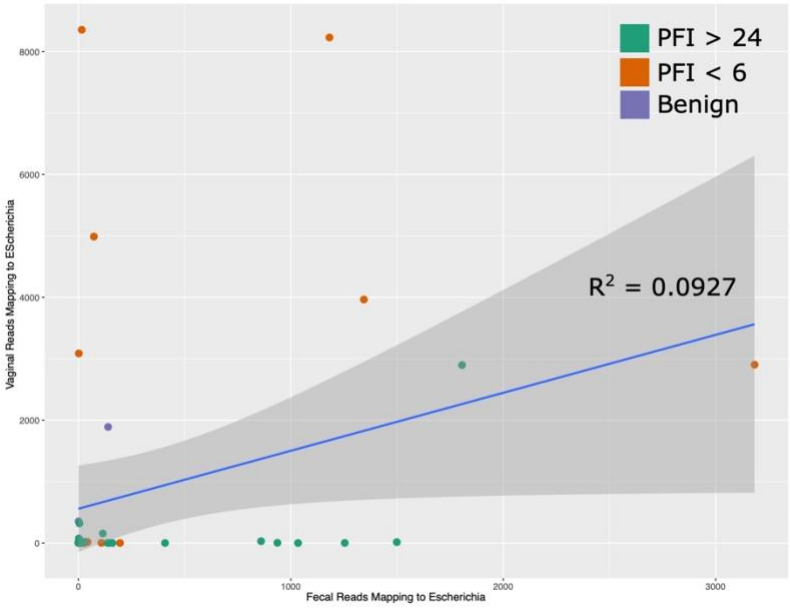
Supplementary Figure C - 7: Genera at high abundance in fecal outlier group

Genera belonging to the Clostridiales order (Lachnospira, Ruminococcaceae, Subdoligranulum) are at higher abundance in the gut beta diversity outlier group.



Supplementary Figure C - 8: Prevotella abundance in gut microbiome of ovarian cancer patients.

Patients with ovarian cancer (both platinum-sensitive and platinum-resistant) have higher levels of Prevotella in the gut microbiome compared to controls.



Supplementary Figure C - 9: Relationship between abundance of Escherichia in the vaginal and gut microbiomes.

These is a very weak relationship to abundance of Escherichia in the gut microbiome. Standard error is outlined in grey.

Supplementary Tables C

Supplementary Table C - 1: Full lifestyle metadata for patients involved in this study

For individuals where medical history or lifestyle metadata was unavailable, UNK was given. N/A is for samples where the metadata category is irrelevant

SampleID	DominantBa	Neuropathy	StudyGroup	age_years	months_last	MultiVitamir	Probiotic	VitaminB_Fc	VitaminD	Other_Dieta	Water_sourc	Pet	Education	Income	TimeOutdoo	Alchohol	Floss_teeth	BowelMover	BowelMover	MostRecentV	WeightChan	TonsilsRemo	AppendixRer	Hormonal_D	Acid_Reflux	Surgery	Menopause	Ethnicity	HormoneRep	Histology	Residual Dis
ocm002	prevotella	none	PFI>24	54.9	29	no	never	daily	daily	yes	bottled	dog	some_colleg	<19,999	<1hr	rarely	rarely	two	normal	6 months	increaseTen	no	yes	yes	yes	iCRS	post_menop	White	no	serous	NGR
ocm003	escherichia	none	PFI<6	69.5	25	no	never	never	never	no	well	cats_dogs	highSchool_	<19,999	<1hr	never	regularly	one	normal	6 months	stable	no	yes	no	yes	pCRS	post_menop	NativeAmeri	no	serous	>1cm
ocm004	lactobacillus	none	PFI>24	52.9	65	no	never	never	never	no	bottled	dog	highSchool_	<19,999	1_3hr	never	regularly	one	diarrhea	>1year	stable	no	no	no	no	pCRS	pre_menopa	White	no	serous	>1cm
ocm005	lactobacillus	none	PFI>24	68.9	82	no	never	never	daily	no	city+bottled	dog	<highSchool	<19,999	<1hr	rarely	daily	four	diarrhea	6 months	stable	no	not sure	yes	no	iCRS	post_menop	White	yes	serous	<1cm
ocm007	diverse	>grade2	PFI>24	68.5	53	no	never	never	never	no	city+filtered	dog	graduate_de	>100,000	1_3hr	never	daily	one	normal	6 months	stable	no	yes	no	no	pCRS	post_menop	NativeAmeri	no	serous	NGR
ocm009	diverse	none	PFI>24	52.8	98	yes	never	never	daily	no	city	cats_dogs	bachelors	>100,000	<1hr	rarely	daily	<1	normal	>1year	stable	no	no	no	no	pCRS	pre_menopa	White	yes	serous	<1cm
ocm012	prevotella	none	PFI>24	59.1	124	yes	never	never	daily	yes	city+filtered	dog	some_colleg	>50000	<1hr	never	daily	one	normal	1 year	stable	no	yes	no	yes	pCRS	post_menop	White	no	serous	NGR
ocm014	diverse	N/A	benign	52.8	N/A	no	daily	never	never	yes	city	no	graduate_de	>100,000	<1hr	never	rarely	one	normal	month	stable	no	no	no	yes	N/A	post_menop	White	no	N/A	>1cm
ocm021	escherichia	UNK	PFI<6	64.8	7	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	iCRS	post_menop	White	no	serous	NGR
ocm022	escherichia	none	PFI>24	77.7	106	no	never	never	daily	yes	city+filtered	no	associates	NA	<1hr	never	daily	one	constipated	1 year	decreaseTen	yes	no	no	yes	pCRS	post_menop	White	no	serous	NGR
ocm024	prevotella	none	PFI>24	76.5	40	yes	never	never	never	no	cit+bottled+f	no	highSchool_	>100,000	1_3hr	never	rarely	one	constipated	>1year	decreaseTen	yes	no	no	yes	pCRS	post_menop	White	no	serous	NGR
ocm026	lactobacillus	UNK	PFI<6	45.5	10	no	never	never	never	no	city	dog	bachelors	>50000	1_3hr	never	rarely	<1	normal	1 year	increaseTen	yes	yes	yes	yes	iCRS	pre_menopa	White	no	serous	NGR
ocm027	diverse	UNK	PFI<6	59.2	7	no	daily	never	never	yes	bottled	cats_dogs	associates	<19,999	>3hr	never	rarely	two	normal	6 months	decreaseTen	no	no	no	no	iCRS	post_menop	White	no	serous	<1cm
ocm028	diverse	none	PFI>24	71.9	45	yes	daily	never	daily	yes	city+filtered	cat	highSchool_	>20,000	<1hr	never	daily	>5	diarrhea	6 months	decreaseTen	yes	yes	yes	yes	pCRS	post_menop	White	no	serous	NGR
ocm029	prevotella	UNK	PFI<6	68.3	4	no	never	never	never	no	well	dog	highSchool_	>50000	1_3hr	never	rarely	three	diarrhea	month	stable	no	no	no	no	iCRS	post_menop	White	no	serous	<1cm
ocm031	lactobacillus	none	PFI<6	47.2	7	no	never	never	never	yes	bottled	dog	some_colleg	>20,000	<1hr	rarely	daily	<1	normal	6 months	stable	yes	yes	no	yes	iCRS	pre_menopa	White	no	serous	<1cm
ocm032	lactobacillus	>grade2	PFI<6	56.9	6	no	daily	daily	daily	yes	bottled	cats_dogs_o	graduate_de	>50000	1_3hr	regularly	daily	one	normal	>1year	increaseTen	yes	no	yes	yes	iCRS	pre_menopa	White	no	serous	>1cm
ocm033	diverse	>grade2	PFI>24	73.66	68	no	never	never	never	no	bottled	cats_dogs	bachelors	<19,999	<1hr	never	daily	four	normal	>1year	stable	no	yes	yes	no	pCRS	post_menop	Black	no	serous	NGR
ocm034	prevotella	>grade2	PFI>24	77.9	61	no	daily	never	daily	yes	cit+bottled+f	dog	some_colleg	>20,000	1_3hr	rarely	never	one	normal	month	increaseTen	no	no	no	yes	pCRS	post_menop	White	no	serous	NGR
ocm036	gardnerella	>grade2	PFI>24	51	39	no	never	never	never	no	filtered	cats_dogs	highSchool_	>20,000	<1hr	never	regularly	>5	normal	6 months	decreaseTen	yes	yes	no	no	pCRS	pre_menopa	White	no	serous	NGR
ocm043	diverse	none	PFI>24	70.2	67	yes	never	daily	daily	yes	city+bottled	no	bachelors	>50000	<1hr	never	never	one	diarrhea	week	UNK	UNK	UNK	UNK	UNK	iCRS	post_menop	White	yes	serous	>1cm
ocm044	lactobacillus	>grade2	PFI>24	58.2	18	no	never	never	never	no	bottled	dog	highSchool_	>20,000	<1hr	never	never	three	diarrhea	1 year	increaseTen	no	yes	no	no	pCRS	post_menop	White	no	serous	NGR
ocm045	lactobacillus	none	PFI>24	67.9	42	yes	never	daily	daily	NA	bottle+filter	cat	highSchool_	>20,000	1_3hr	never	daily	three	normal	6 months	increaseTen	yes	yes	yes	no	pCRS	post_menop	White	no	serous	NGR
ocm046	diverse	none	PFI>24	64.4	54	no	daily	daily	rarely	yes	well	dog	graduate_de	>100,000	1_3hr	regularly	daily	one	normal	6 months	increaseTen	yes	yes	yes	no	pCRS	post_menop	White	no	serous	NGR
ocm047	lactobacillus	none	PFI>24	70.2	87	NA	never	never	daily	yes	well	cats_dogs_o	some_colleg	>20,000	<1hr	never	never	two	normal	6 months	stable	no	yes	yes	yes	pCRS	post_menop	White	no	serous	<1cm
ocm049	lactobacillus	>grade2	PFI>24	47.5	8	no	never	never	never	no	well	cats_dogs_o	some_colleg	<19,999	>3hr	rarely	regularly	one	normal	6 months	decreaseTen	yes	yes	no	yes	iCRS	pre_menopa	White	no	serous	NGR
ocm051	diverse	none	PFI>24	45.5	45	no	never	never	never	no	well+bottled	dog	bachelors	>100,000	<1hr	regularly	regularly	>5	diarrhea	week	stable	no	yes	no	yes	pCRS	pre_menopa	White	no	serous	NGR
ocm057	prevotella	UNK	PFI<6	74.9	18	no	never	never	never	yes	filtered	no	some_colleg	>50000	<1hr	rarely	rarely	two	normal	6 months	stable	no	yes	no	yes	pCRS	post_menop	White	no	UNK	NGR
ocm058	diverse	N/A	benign	79	N/A	yes	never	never	never	yes	well+bottled	dog	graduate_de	NA	<1hr	rarely	rarely	three	diarrhea	1 year	decreaseTen	no	yes	no	no	N/A	post_menop	NativeAmeri	no	N/A	>1cm
ocm060	lactobacillus	none	PFI>24	60.9	150	yes	daily	daily	daily	yes	filtered	no	bachelors	NA	<1hr	never	regularly	two	dont know	1 year	stable	no	no	no	yes	pCRS	pre_menopa	White	no	serous	>1cm
ocm061	escherichia	none	PFI<6	64.6	9	no	never	never	never	no	well	dog	some_colleg	>100,000	NA	never	rarely	one	constipated	month	stable	no	no	no	yes	iCRS	post_menop	White	no	serous	>1cm
ocm063	gardnerella	UNK	PFI>24	58.6	UNK	yes	never	rarely	daily	yes	bottle+filter	dog	some_colleg	NA	<1hr	never	daily	one	normal	6 months	increaseTen	yes	no	no	no	pCRS	pre_menopa	White	yes	serous	NGR
ocm070	diverse	UNK	PFI<6	65.4	38	no	daily	daily	never	yes	city	cats_dogs	graduate_de	>100,000	>3hr	regularly	daily	two	normal	6 months	increaseTen	not sure	yes	yes	no	pCRS	post_menop	White	no	serous	NGR
ocm071	lactobacillus	N/A	benign	57.8	N/A	yes	never	never	never	no	city	cats_dogs_o	bachelors	>50000	1_3hr	rarely	daily	one	normal	month	decreaseTen	no	yes	no	no	N/A	post_menop	White	no	N/A	>1cm
ocm075	diverse	UNK	PFI<6	62.5	6	yes	UNK	never	never	no	bottled	NA	some_colleg	>50000	<1hr	never	rarely	four	normal	month	stable	no	no	no	yes	no surg	post_menop	White	no	serous	>1cm
ocm084	prevotella	N/A	benign	33.7	N/A	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	N/A	pre_menopa	NativeAmeri	no	N/A	>1cm
ocm088	escherichia	none	PFI<6	53.6	1	no	never	never	never	no	bottled	no	graduate_de	>100,000	1_3hr	never	never	two	normal	6 months	increaseTen	no	no	no	no	iCRS	pre_menopa	Black	no	serous	>1cm
ocm089	prevotella	N/A	benign	82.7	6	no	never	daily	never	no	bottled	no	some_colleg	>20,000	<1hr	rarely	never	<1	constipated	month	stable	yes	yes	no	yes	iCRS	post_menop	White	no	serous	<1cm
ocm091	gardnerella	UNK	PFI<6	38.5	35	yes	never	daily	never	yes	well	cats_dogs	associates	>50000	1_3hr	rarely	daily	<1	normal	6 months	stable	no	yes	no	no	iCRS	pre_menopa	White	no	serous	NGR
ocm093	prevotella	>grade2	PFI>24	59.7	72	yes	never	never	never	no	filtered	cat	highSchool_	NA	<1hr	rarely	never	four	diarrhea	week	stable	no	yes	no	no	pCRS	post_menop	White	no	serous	NGR
ocm094	diverse	UNK	PFI<6	73.6	3	yes	never	daily	daily	yes	filtered	no	some_gradu	>50000	<1hr	never	regularly	one	dont Know	6 months	decreaseTen	yes	no	yes	yes	iCRS	post_menop	White	no	serous	>1cm
ocm096	diverse	>grade2	PFI>24	63.6	41	no	never	never	never	yes	city+filtered	dog	bachelors	>20,000	<1hr	never	daily	three	normal	month	stable	no	yes	yes	yes	pCRS	post_menop	White	no	serous	<1cm
ocm098	diverse	>grade2	PFI<6	75.1	15	yes	daily	never	daily	yes	bottled	other_pet	associates	<19,999	<1hr	never	never	two	normal	week	stable	yes	yes	yes	yes	iCRS	post_menop	White	no	serous	<1cm
ocm099	diverse	N/A	benign	61.9	N/A	no	never	never	regularly	no	city+bottled	dog	highSchool_	>100,000	<1hr	regularly	regularly	one	normal	month	decreaseTen	yes	yes	no	no	N/A	post_menop	White	no	N/A	>1cm
ocm106	escherichia	UNK	PFI<6	58	3	no	never	never	rarely	no	filtered	cat	bachelors	>100,000	<1hr	never	regularly	two	diarrhea	6 months	decreaseTen	no	yes	no	yes	iCRS	post_menop	White	no	serous	<1cm

Supplementary Table C - 2: Raw Escherichia reads from extraction negatives and PCR blanks

Number of raw reads mapping to Escherichia from all extraction negatives (EXN) and pcr blanks (pcrBlanks) that were sequenced across the entirity of this analysis. The average unrarefied read depth across all samples in our dataset is 38,818. The percent reads mapping to Escherichia compared to the average total read depth is also provided. The number of reads mapping to Escherichia in our extraction negatives and pcr blanks is very low and indicates lab contamination does not explain Escherichia abundance in our dataset.

originalName	newName	EscherichiaReads	PercentOfAverageUnrarefiedRe
OCM.EXN.EXN.V4R.001	EXN01	70	0.1803
OCM.EXN.EXN.V4R.002	EXN02	634	1.6333
OCM.EXN.EXN.V4R.003	EXN03	55	0.1417
OCM.EXN.EXN.V4R.004	EXN04	24	0.0618
OCM.EXN.EXN.V4R.005	EXN05	12	0.0309
EXN06	EXN06	4	0.0103
EXN07	EXN07	25	0.0644
EXN08	EXN08	0	0.0000
EXN09	EXN09	2	0.0052
EXN10	EXN10	0	0.0000
EXN11	EXN11	4	0.0103
EXN12	EXN12	0	0.0000
EXN13	EXN13	0	0.0000
EXN14	EXN14	333	0.8578
EXN15	EXN15	0	0.0000
EXN16	EXN16	629	1.6204
EXN18	EXN17	0	0.0000
EXN19	EXN18	2	0.0052
EXN20	EXN19	1	0.0026
EXN26	EXN20	22	0.0567
EXN27	EXN21	0	0.0000
EXN28	EXN22	0	0.0000
EXN31	EXN23	1175	3.0269
OCM.EXN.EXN.V4R.EXN	EXN24	0	0.0000
OCM.BLK.XXX.V4R.001	pcrBlank01	18	0.0464
OCM.BLK.XXX.V4R.002	pcrBlank02	10	0.0258
pcr.bl.1	pcrBlank03	3	0.0077
PCR.BL.2	pcrBlank04	0	0.0000
PCR.BL.3	pcrBlank05	0	0.0000
PCR.BL.4	pcrBlank06	1	0.0026
PCR.BL1	pcrBlank07	1	0.0026
PCR.BLK	pcrBlank08	0	0.0000
PCRBL1	pcrBlank09	4	0.0103
PCRBL2	pcrBlank10	2	0.0052