UNIVERSITY OF CALIFORNIA SAN DIEGO

Drool rules!
Novel methods to further knowledge of microbial communities in the human oral cavity

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Clarisse Augusta Marotz

Committee in Charge:

Professor Rob Knight, Chair
Professor Karsten Zengler, Co-Chair
Professor Rachel Dutton
Professor Christian Metallo
Professor Victor Nizet

2020

The dissertation of Clarisse Augusta Marotz is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____
Co-Chair

_____
Chair

University of California San Diego
2020

DEDICATION

To my family, for always having my back, my friends, for making the load a little lighter, and Chase, for making it all worthwhile.

EPIGRAPH

"We are not a collection of individuals,
but a macroorganism living as an ecosystem.
We are completely outside ourselves,
and the world is completely inside us"
-Holly Herndon

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am deeply grateful to have had the opportunity to perform my doctoral research with amazing mentors and colleagues. When I started my doctoral studies in 2015 I was completely new to the field of microbiology. At least once a week I learned something that blew my mind, and this has not changed throughout my entire time as a PhD student.

My mentors Drs. Karsten Zengler and Rob Knight allowed me the flexibility to follow the data and pursue interesting new directions, leading to this dissertation that I could have never dreamed of just four and a half years ago. Their contagious enthusiasm for discovering novel aspects of biology created a wonderful learning environment. They have been great role models for how to successfully perform collaborative research. It was through this collaborative environment that I was able to publish work with scientists across the country including Dr. Ryan Demmer, Dr. Jack Gilbert, Dr. Emiley Eloe-Fadrosh, and many others. In addition to serving as a role model for solid scientific inquiry, I am grateful to Dr. Zengler for showing me that it is possible to maintain a healthy work life balance while producing incredible science. Truly I could not have asked for better co-mentors.

My successes as a graduate student are also due in large part to the network of fabulous colleagues curated by my mentors. In particular, the post-doctoral fellows Livia Zaramela, Cristal Zuniga, and Max Al-bassam have been amazing role models, and I will always be grateful for the time they spent teaching and encouraging me. I am also indebted to Bryn Taylor, who paved the path for a BMS student co-mentored in the Knight lab, and made it look easy.

Although I had no coding background when I began my PhD, long conversations with colleagues turned friends have been a crucial part to my growth as a scientist, and I am grateful to Jamie Morton, Cameron Martino, Marcus Fedarko, George Armstrong, Yoshiki Vasquez-Baeza,

and Daniel McDonald for making math fun. Lastly, I would like to thank Dr. Pedro Belda-Ferre, who I have worked with closely during the past few weeks to scale a research response to the global coronavirus pandemic, for helping me stay calm under pressure and prioritize what's important.

Chapter I, Introduction part 1, in full, is a reprint of previously published material: Marotz, Clarisse A., and Amir Zarrinpar. *Focus: Microbiome: Treating Obesity and Metabolic Syndrome with Fecal Microbiota Transplantation*. The Yale journal of biology and medicine 89.3 (2016): 383. I was the primary investigator and author of this paper. The co-author listed above supervised and provided support for the research and has given permission for the inclusion of the work in this dissertation.

Chapter I, Introduction part 2, in full, is a reprint of previously published material: Knight R, Callewaert C, Marotz C, Hyde ER, Debelius JW, McDonald D, Sogin ML. *The microbiome and human biology*. Annual review of genomics and human genetics. 2017 Aug 31;18:65-86. I was one of the primary investigators and authors of this paper. The co-authors listed above supervised or provided support for the research and have given permission for the inclusion of the work in this dissertation.

Chapter I, Introduction part 3, in full, is a reprint of previously published material: Tripathi A, Marotz C, Gonzalez A, Vázquez-Baeza Y, Song SJ, Bouslimani A, McDonald D, Zhu Q, Sanders JG, Smarr L, Dorrestein PC. *Are microbiome studies ready for hypothesis-driven research?* Current opinion in microbiology. 2018 Aug 1;44:61-9. I was one of the primary investigators and authors of this paper. The co-authors listed above supervised or provided support for the research and have given permission for the inclusion of the work in this dissertation.

Chapter II, part 1, in full, is a reprint of previously published material: Marotz C, Amir A, Humphrey G, Gaffney J, Gogul G, Knight R. *DNA extraction for streamlined metagenomics of diverse environmental samples. Biotechniques*. 2017 Jun;62(6):290-3. I was the primary investigator and author of this paper. The co-authors listed above supervised or provided support for the research and have given permission for the inclusion of the work in this dissertation.

Chapter II, part 2, in full, is a reprint of previously published material: Marotz C, Sharma A, Humphrey G, Gottel N, Daum C, Gilbert JA, Eloe-Fadrosh E, Knight R. *Triplicate PCR reactions for 16S rRNA gene amplicon sequencing are unnecessary*. BioTechniques. 2019 Jul;67(1):29-32. I was the primary investigator and author of this paper. The co-authors listed above supervised or provided support for the research and have given permission for the inclusion of the work in this dissertation.

Chapter II, part 3, in full, is a reprint of previously published material: Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K. *Improving saliva shotgun metagenomics by chemical host DNA depletion*. Microbiome. 2018 Dec;6(1):42. I was the primary investigator and author of this paper. The co-authors listed above supervised or provided support for the research and have given permission for the inclusion of the work in this dissertation.

Chapter III, part 1, in full, is a reprint of previously published material: Morton JT, Marotz C, Washburne A, Silverman J, Zaramela LS, Edlund A, Zengler K, Knight R. *Establishing microbial composition measurement standards with reference frames. Nature communications*. 2019 Jun 20;10(1):1-1. I was one of the primary investigators and authors of this paper. The co-authors listed above supervised or provided support for the research and have given permission for the inclusion of the work in this dissertation.

Chapter IV, part 1, is currently under review for publication: Marotz C, Morton J, Navarro P, Knight R, Zengler K. *Quantifying live microbial load in human saliva samples over time reveals stable composition and dynamic load*. I am the primary investigator and author of this paper. The co-authors listed above supervised or provided support for the research and have given permission for the inclusion of the work in this dissertation.

Chapter IV, part 2, is currently being prepared for submission: Marotz C, Martino C, Knight R, Demmer R. *Early microbial markers of periodontitis in the Oral Infections Glucose Intolerance and Insulin Resistance Study (ORIGINS).* I am the primary investigator and author of this paper. The co-authors listed above supervised or provided support for the research and have given permission for the inclusion of the work in this dissertation.

VITA

2009	Bachelor of Science, University of Washington
2011	Master of Science, Charité Universitätsmedizin
2011-2015	Research assistant, Weill Cornell Medical College
2020	Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Marotz C, Martino C, Knight R, Demmer R. Early microbial markers of periodontitis in the Oral Infections Glucose Intolerance and Insulin Resistance Study (ORIGINS). *In prep*

Marotz C, Morton J, Navarro P, Knight R, Zengler K. Quantifying live microbial load reveals the dynamic ecosystem of human saliva. *In review.*

Fedarko M, Martino C, Morton J, **Marotz** C, Minich J, Knight R. Visualizing 'omic feature rankings in tandem with sample abundances using Qurro. *NAR Genomics and Bioinformatics*. (2020) *in press*

Kohn J, Kosciolek T, Guay-Ross R, Hong S, Hansen S, **Marotz** C, Swafford A, Knight R, Hong S. Oral microbiota associations with psychological distress and peripheral inflammation in adults. *Brain Behavior and Immunity.* (2020) *in press*

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz J, Bittinger K, Brejnrod A, Brislawn C, Titus Brown C, Callahan B, Caraballo-Rodriguez A, Chase J, Cope E, DaSilva R, Diener C, Dorrestein P, Douglas G, Durall D, Duvallet C, Edwardson C, Ernst M, Estaki M, Fouquier J, Gauglitz J, Gibbons S, Gibson D, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley G, Janssen S, Jarmusch A, Jiang L, Kaehler B, Kang K, Keefe C, Keim P, Kelley S, Knights D, Koester I, Kosciolek T, Kreps J, Langille M, Lee J, Liu Y, Loftfield E, Lozupone C, Maher M, **Marotz** C, Martin B, McDonald D, McIver L, Melnik A, Metcalf J, Morgan S, Morton J, Naimey A, Navas-Molina J, Nothias L, Orchanian S, Pearson T, Peoples S, Petras D, Preuss M, Pruesse E, Rasmussen L, Rivers A, Robeson M, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song S, Spear J, Swafford A, Thompson L, Torres P, Trinh P, Tripathi A, Turnbaugh P, Ul-Hasan S, vanderHooft J, Vargas F, Vazquez-Baeza Y, Vogtmann E, Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber K, Williamson C, Willis A, Zu Z, Zaneveld J, Zhang Y, Zhu Q, knight R, and Caporaso JG. Reproducible, interactive, scalable and

extensible microbiome data science using QIIME 2. Nature biotechnology. (2019) Aug;37(8):852-7.

McDonald D, Kaehler B, Gonzalez A, DeReus J, Ackermann G, **Marotz** C, Huttley G, Knight R. redbiom: a Rapid Sample Discovery and Feature Characterization System. *mSystems*. (2019) 4(4):e00215-19.

Morton J*, **Marotz C***, Washburne A, Silverman J, Zaramela L, Edlund A, Zengler K, Knight R. Establishing microbial measurement standards with reference frames. *Nature Communications*. (2019) . **co-first authors*

Marotz, C., Sharma, A., Humphrey, G., Gottel, N., Daum, C., Gilbert, J.A., Eloe-Fadrosh, E. and Knight, R. Triplicate PCR reactions for 16S rRNA gene amplicon sequencing are unnecessary. *BioTechniques*, (2019) *67*(1).

Martino C, Morton J, **Marotz C,** Thompson L, Tripathi A, Knight R, Zengler K. A novel sparse compositional technique reveals microbial perturbations. *mSystems* 4.1 (2019): e00016-19.

Tripathi, A., **Marotz**, C., Gonzalez, A., Vázquez-Baeza, Y., Song, S.J., Bouslimani, A., McDonald, D., Zhu, Q., Sanders, J.G., Smarr, L. and Dorrestein, P.C., Knight, R. Are microbiome studies ready for hypothesis-driven research?. *Current opinion in microbiology*. (2018). *44*, pp.61-69.

Eisenhofer R, Minich J, **Marotz C**, Cooper A, Knight R, Weyric L. Contamination in low-biomass microbiome studies: issues and recommendations. *Trends in Microbiology* (2018).

Zarrinpar A, Chaix A, Xu ZZ, Chang MW, **Marotz CA**, Saghatelian A, Knight R, Panda S. Antibiotic-induced microbiome depletion alters metabolic homeostasis by affecting gut signaling and colonic metabolism. *Nature Communications*. (2018) Jul 20;9(1):2872.

McDonald D, Hyde E, Debelius, J, Morton J, Gonzalez A, Ackermann G, Aksenov A, Behsaz B, Brennan C, Chen Y, DeRight Goldasich L, Dorrestein P, Dunn R, Fahimipour A, Gaffney J, Gilbert J, Gogul G, Green J, Hugenholtz P, Humphrey G, Huttenhower C, Jackson M, Janssen S, Jeste D, Jiang L, Kelley S, Knights D, Kosciolek T, Ladau J, Leach J, **Marotz C**, Meleshko D, Melnik A, Metcalf J, Mohimani H, Montassier, Navas-Molina J, Nguyen T, Peddada S, Pevzner P, Pollard K, Rahnavard G, Robbins-Pianka A, Sangwan N, Shorenstein J, Smarr L, Song SJ, Spector T, Swafford A, Thackray V, Thompson L, Tripathi A, Vázquez-Baeza, Y, Vrbanac A, Wischmeyer P, Wolfe E, Zhu Q. The American Gut Consortium. American Gut: an Open Platform for Citizen Science Microbiome Research. *mSystems* 3.3 (2018): e00031-18.

Marotz, C. A., Sanders, J. G., Zuniga, C., Zaramela, L. S., Knight, R., & Zengler, K. Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome*. (2018) Feb 6(1), 42.

Vázquez-Baeza, Y., Callewaert, C., Debelius, J., Hyde, E., **Marotz, C**., Morton, J. T., ... & Knight, R. (2018). Impacts of the human gut microbiome on therapeutics. *Annual review of pharmacology and toxicology*. (2018) Jan 58, 253-270.

Hemmings S, Malan-Muller S, van den Heuvel L, Demmitt B, Stanislawski M, Smith D, Bohr A, Stamper C, Hyde E, Morton J, **Marotz C**, Siebler P, Hoisington A, Brenner L, Postolache T, Dicks L, McQueen M, Krauter K, Knight R, Seedat S, Lowry C. The Microbiome in Posttraumatic Stress Disorder (PTSD) and Trauma-Exposed Controls: An Exploratory Study. *Psychosomatic Medicine*. (2017) Oct;79(8):936-946.

Marotz C, Amir A, Humphrey G, Gaffney J, Gogul G, Knight R. DNA extraction for streamlined metagenomics of diverse environmental samples. *Biotechniques*. (2017) Jun 1;62(6):290-293

Knight R, Callewaert C, **Marotz C**, Hyde E, Debelius J, McDonald D, Sogin M. The Microbiome and Human Biology. *Annual Review of Genomics and Human Genetics*. (2017) Mar 20.

Marotz C, Knight R. Culturing: looking it up in our gut. *Nature Microbiology*. (2016) Sep 27;1:16169.

Marotz C, Zarrinpar A. Focus: Microbiome: Treating Obesity and Metabolic Syndrome with Fecal Microbiota Transplantation. *The Yale Journal of Biology and Medicine*. (2016) Sep;89(3):383.

Beltran H, Prandi D, Mosquera JM, Benelli M, Puca L, Cyrta J, **Marotz C**, Giannopoulou E, Chakravarthi BV, Varambally S, Tomlins SA. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nature medicine*. (2016) Feb 8.

Boysen G, Barbieri CE, Prandi D, Blattner M, Chae SS, Dahija A, Nataraj S, Huang D, **Marotz C**, Xu L, Huang J. SPOP mutation leads to genomic instability in prostate cancer. *Elife*. (2015) Sep 16;4:e09207.

Busch-Dienstfertig M*, **Roth C***, Stein C. Functional Characteristics of the Naked Mole Rat μ-Opioid Receptor. *PloS one*. (2013) Nov 27;8(11):e79121.**co-first authors*

Lemos JC, **Roth C**, Messinger DI, Gill HK, Phillips PE, Chavkin C. Repeated stress dysregulates κ-opioid receptor signaling in the dorsal raphe through a p38α MAPK-dependent mechanism. *The Journal of Neuroscience*. (2012) Sep 5;32(36):12325-36.

Lemos JC, **Roth C**, Chavkin C. Signaling Events Initiated by Kappa Opioid Receptor Activation: Quantification and Immunocolocalization Using Phospho-Selective KOR, p38 MAPK, and K IR 3.1 Antibodies. *Signal Transduction Immunohistochemistry: Methods and Protocols*. (2011):197-219.

ABSTRACT OF THE DISSERTATION

Drool rules!
Novel methods to further knowledge of microbial communities in the human oral cavity

by

Clarisse Augusta Marotz

Doctor of Philosophy in Biomedical Sciences

University of California San Diego, 2020

Professor Rob Knight, Chair

Professor Karsten Zengler, Co-Chair

As I prepare my dissertation, a global Coronavirus pandemic is spreading--affecting the way we travel, work, and communicate. Transmitted primarily via respiratory droplets, this virus spread across the entire globe in a matter of months, highlighting the interconnectedness of modern life and the microscopic world which invisibly underlies all our interactions.

In fact, the vast majority of life on earth is microbial, and for the vast majority of time microorganisms were the only life present. Animals emerged in a robust microbial world, and our evolution was shaped by and from these microorganisms.

Even during this global pandemic, it is easy to view ourselves as distinct, independent organisms. But recent evidence challenges this notion. Roughly 1% of our body weight is comprised of trillions of microorganisms, collectively referred to as the human microbiome. Our microbiomes are necessary for healthy nutrition, immune function, and even reproduction. How our microbiomes are formed, maintained, shared, and perturbed are active areas of research. My doctoral research focused on understanding and improving the molecular tools used to evaluate microbial communities and applying those tools to better understand human microbiomes.

This dissertation begins with three reviews summarizing what we know about human microbiomes and what is still lacking. Chapters 2 and 3 describe benchtop and computational advances to the human microbiome field, respectively. The final chapter includes two research articles currently under review describing the application of these tools to gain novel insight into human oral microorganisms.

Throughout my research, I found that human saliva served as a valuable tool for developing these novel techniques because it is simple to collect, high biomass, and relatively easy to manipulate. So although it was not my original intention, I have come to appreciate saliva as a valuable resource for pushing the boundary of human microbiome research. It is my hope that this research improves our ability to evaluate microbial communities across a broad range of contexts, allowing for future research on altering microbiomes to improve human and environmental health.

Chapter 1.

Introduction

This introduction is divided ino three sections. The first section describes evidence that the microbial communities in our guts can influence our physiology. Although fecal matter transplantations have an ancient history, their use in modern medicine has only recently become widely accepted, and the data shows that by changing the microbes in an indivdual's gut you can influence immunity and metabolism.

Once the stage has been set for the translational power of understanding human microbiomes, the second section is a general summary of what is known about how microbial communities influence human biology. The final section is a co-authored opinion piece recognizing the current knowledge gaps preventing targeted microbiome perturbation studies and serves as a transition to my research in improving the available molecular tools for evaluating host-associated microbiomes.

1.1

Treating Obesity and Metabolic Syndrome with Fecal Microbiota Transplantation

The worldwide prevalence of metabolic syndrome, which includes obesity and its associated diseases, is rising rapidly. The human gut microbiome is recognized as an independent environmental modulator of host metabolic health and disease. Research in animal models has demonstrated that the gut microbiome has the functional capacity to induce or relieve metabolic syndrome. One way to modify the human gut microbiome is by transplanting fecal matter, which contains an abundance of live microorganisms, from a healthy individual to a diseased one in the hopes of alleviating illness. Here we review recent evidence suggesting efficacy of fecal microbiota transplant (FMT) in animal models and humans for the treatment of obesity and its associated metabolic disorders.

1.1.1 Introduction

Over the past half-century, the prevalence of obesity and its related metabolic disorders, such as type 2 diabetes (T2D), non-alcoholic fatty liver disease, and hypercholesterolemia, have increased dramatically. Collectively, these diseases cause an undue burden on health care costs and significant morbidity and mortality. While these diseases are linked to human genetics and lifestyle changes, the human gut microbiome, or the microorganisms living in the gut and their collective genomes, is now recognized to play an emerging role in metabolic health and disease [1, 2]. Trillions of diverse organisms, including bacteria, fungi, archaea, and viruses, have co-evolved to live in the human gut [3]. These commensal organisms comprise the gut microbiome,

and their collective genome, referred to as the metagenome, contains more than a hundred-fold the number of genes than their host does [4]. Certain metagenomic patterns are associated with obesity, as well as other phenotypes [5]. These patterns are responsive to weight change in individuals [6], suggesting that modulating the gut microbiome is dynamically correlated with the human host's metabolic phenotype.

There are many ways that the gut microbiota can be altered, including probiotics (non-pathogenic organisms beneficial to the host), prebiotics (chemicals that induce growth and/or activity of commensal organisms), and fecal microbiota transplantation (FMT)[7]. Though beneficial effects of probiotics have been reported in many studies, none show an alteration in fecal microbiota composition [8]. FMT on the other hand, causes significant changes in fecal microbiota composition [9]. FMT as a potential therapeutic has a long history. The successful practice of altering gut microbiota with FMT from a healthy to diseased individual was first recorded in the 4th century for the treatment of severe diarrhea [10]. Recently, randomized controlled clinical trials show astounding successes for recurrent, refractory *Clostridium dificile* infection (CDI). Multiple studies have reported greater than 90% efficacy, dramatically more successful than traditional therapy, in resolving recurrent CDI [11]. Recent evidence from animal and human models suggests FMT could also be used as a therapeutic intervention against obesity [12, 13]. In this review we will provide a status update on the role of FMT in treating obesity and its associated metabolic disorders.

1.1.2 Gut microbiota and host metabolism

Whereas inter-individual microbiota composition can vary dramatically, a conserved set of bacterial functional gene profiles are present in all healthy individuals, implying a role for the

microbiome in physiological gut functioning [1, 14, 15]. Alterations of this complex physiological bacterial population associated with negative functional outcomes or disease, known as dysbiosis, can cause low-level inflammation and altered intestinal homeostasis. Dysbiosis is linked to a variety of ailments, including obesity and its associated metabolic disturbances [16].

The mechanism by which dysbiosis leads to metabolic disturbances is not well understood. Leading theories include changes in the microbiome's digestive efficiency and perturbed intestinal signaling through alterations of luminal metabolites, low molecular weight signaling chemicals, released by bacteria in the intestinal lumen such as secondary bile acids (BAs) and short-chain fatty acids (SCFAs)[17]. The gut microbiome is essential for fermenting indigestible foodstuffs into products that can be used by, or modulate, the intestine (e.g. complex carbohydrates into SCFAs) [18]. In murine models, obesity-related microbes are able to harvest greater energy from ingested material [19]. In addition, the microbiome's metabolism of primary BAs to secondary BAs affects host metabolism by modulating activation of the farnesoid X receptor, a master regulator of hepatic triglyceride and glucose homeostasis [20], as well as G-protein coupled BA receptors, which can increase metabolic rate in brown adipose tissue [21-23]. Lastly, diet accounts for 57% of structural variation in the mouse gut microbiome [24], which shifts tremendously in response to the host's gender, diet, circadian rhythms, and feeding pattern [25-28], suggesting that it is a malleable system amenable to manipulation for therapeutic advantage.

1.1.3 Fecal matter transplant methodology

Currently, only recurrent CDI is approved by the FDA for FMT therapy without requiring an investigational new drug (IND) approval. Therefore, the majority of FMT recipients have been treated for severe CDI. These individuals failed repeated treatment with antibiotics and had few

therapeutic options left. In addition, FMT has been studied in inflammatory bowel disease (IBD) since the etiology of this disease, at least in part, results from dysbiosis. However, there have been few controlled, randomized trials for IBD patients and there is no evidence that FMT improves clinical outcomes [33]. In all, FMT has been performed in primarily ill individuals who are at high risk for complications. Hence, the potential risks and complications for relatively healthy patients with obesity or metabolic syndrome remain hypothetically lower compared to previous studies performed in patients with refractory, recurrent CDI or IBD.

Though FMT is relatively easy to perform, there is wide inter-institutional variability in methodology. For example, in preparation for FMT, some institutions give their patients multiple doses of doxycycline or vancomycin in an effort to reduce the native, dysbiotic population [29]. In many institutions, immediately prior to FMT, patients are typically given a polyethylene glycol colon preparation to increase the opportunity for the transplanted microbiome to successfully colonize the gut regardless of whether the FMT is introduced in the upper GI tract or through a colonoscopy. However, there is no published evidence suggesting that this preparation improves FMT clinical outcomes [22].

The processing of fecal matter for transplant is not standardized and needs to be experimentally validated for optimal efficacy. The general principal, however, is more or less universal. As outlined in **Figure 1**, the donated stool is first mixed with saline solution to homogenize it into a liquid sample, and is then filtered to remove any solid feces that may interfere with the transplant.

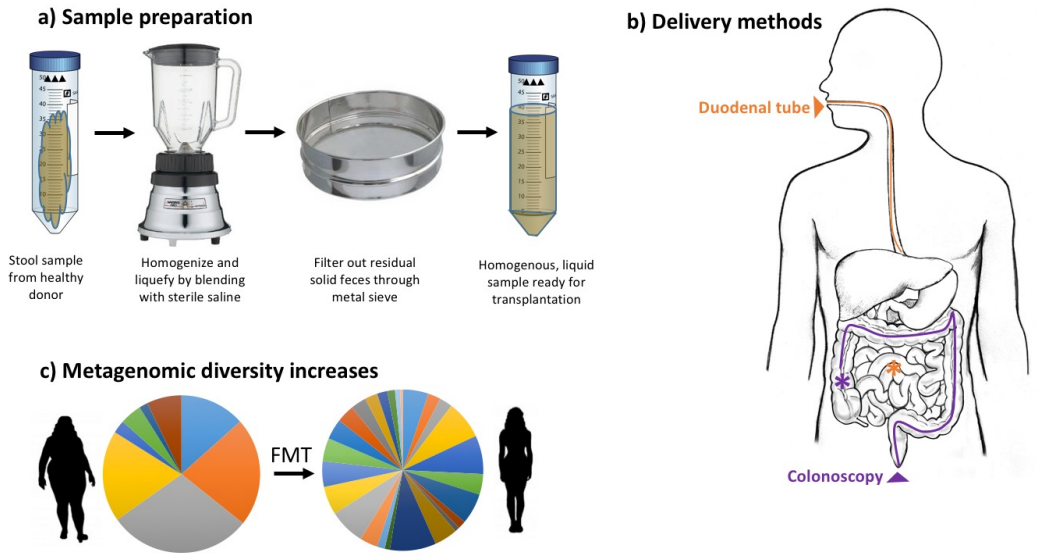

Figure 1.1.1. *Fecal Microbiota Transplantation schematic.* **A)** Donor fecal matter is blended with saline solution and pushed through a metal sieve to achieve a homogenous liquid solution. **B)** Processed fecal microbiota is either delivered via a duodenal tube or colonoscopy. **C)** Representative data showing metagenomic diversity increases following FMT from lean donor to obese recipient.

In order to standardize the processing of fecal matter, studies have compared the efficacy of frozen versus fresh stool samples prior to processing and transplantation. These studies have thus far shown no significant difference in primary outcomes [30, 31]. While studies have performed 16s rRNA sequencing before and after processing to evaluate sample loss, fecal matter contains 99% anaerobic species which may not survive vigorous aerobic blending [32, 33]. Furthermore, 16s rRNA sequencing does not discriminate viable from dead cells. Nevertheless, the overwhelming number of positive results obtained from FMT in treating CDI patients suggests that either the viability of the cells is relatively unimportant, or that a small proportion of survived cells is sufficient to induce a change in the recipient's microbiome and a therapeutic effect.

Processed fecal matter is typically delivered into the gastrointestinal tract of the patient by colonoscopy or duodenal tube/upper endoscopy (Figure 1B). While delivery route often varies from study to study, no statistically significant difference in outcome is reported between the

delivery methods for the treatment of CDI [11, 34]. This finding remains to be validated for the treatment of other diseases, such as IBD or obesity. Regardless, it is important to consider the potential risks associated with each potential delivery route.

The protocol for FMT is widely variable, as summarized in **Table 1**, and standardization of this technique should help elucidate FMT's efficacy.

Table 1.1.1. *Variability in fecal microbiota transplantation methodology.*

Points of variability	Potential methodology	Potential implications
Patient preparation	Type/length of antibiotic treatment, duration of colon preparation	State of patient's gut microbiome could impact susceptiblity to transplant
Donor	Patient relative, 'super donor', designer cultures?	The identification of 'super-donors' hints at the possibility of moving toward the creation of safer, more standardizable synthetic probiotic communities
Sample preparation	Aerobic vs anaerobic; fresh vs frozen vs lyophilized	A recent clincal trial reported no difference in clincal resolution between using fresh or frozen fecal sample for transplantation
Administration	Duodenal tube, colonoscopy, enema, pill	Maximizing practicality of this technique while maintaing efficacy could impacts its prescription and cost
Delivery site	Colon, small intestine	Spatial dynamics of the human microbiome remains poorly characterized, but could results in more targeted therapy

1.1.4 Insulin sensitivity transferred from donors to recipients

Recent studies in animal models show a functional relationship between the gut microbiome and obesity and its associated metabolic disturbances. For example, obesity and insulin resistance in adult rats on a high-fructose diet was reversed with orally administered antibiotics or oral FMT from control rats [13]. Transplanting fecal matter from twins discordant

for obesity into germ-free mice was recently examined [35]. Mice populated with the microbiome from the obese twin had increased adiposity and decreased bacterial diversity compared to mice populated with the microbiome from the lean twin. These results demonstrate the ability of the microbiome to alter the metabolic phenotype of the host.

To date there has only been one published study testing the efficacy of FMT specifically for treatment of metabolic disorders in humans. The hallmark characteristic of metabolic syndrome is insulin resistance, where cells are hypo-responsive to insulin and therefore cannot maintain glucose homeostasis. Fecal microbiota from healthy, lean donors transferred through a duodenal tube to obese individuals diagnosed with T2D affected host metabolism [12]. The study compared patients who received allogenic transplant (n = 9) (i.e. stool from a healthy donor) to autologous transplantation (n=9) (i.e., their own stool). Although there was no reported difference in body mass index 6 weeks after transplantation, there was a significant increase in insulin sensitivity (as measured by the median rate of glucose disappearance) and fecal microbiota diversity, and decrease in fecal SCFA in the allogenic versus autologous group. These promising results have been widely cited and inspired multiple clinical trials (discussed below). Although FMT can induce microbiome alteration towards the donor population for up to 24 weeks post-FMT [29], further studies are need to determine whether FMT can have long-term effects on insulin sensitivity or weight.

Additional clinical trials are necessary to validate the effects of FMT in those with obesity or metabolic syndrome. Importantly, these studies should be randomized, include autologous controls, contain meticulous metadata and track long-term microbiome and patient outcome data. ClinicalTrials.gov lists four ongoing clinical trials testing FMT for metabolic syndrome treatment. A phase 2 clinical trial at Massachusetts General Hospital is evaluating the impact of FMT capsules

on a primary outcome of body weight reduction over 18 weeks [ClinicalTrials.gov ID NCT02530385]. An Italian phase 3 clinical trial is tracking glucose homeostasis over a 6-month period following FMT in combination with diet and exercise [ClinicalTrials.gov ID NCT02050607]. Researchers from China's Nanjing Medical University are evaluating the results of a phase 3 clinical trial on a single, nasogastric-delivered FMT on T2D over a 2-year period [ClinicalTrials.gov ID NCT01790711]. A Canadian double-blind pilot study is testing FMT efficacy in both metabolic syndrome and non-alcoholic fatty liver disease, which is closely associated with obesity [ClinicalTrials.gov ID NCT02496390].

The results from these clinical trials should give us a better idea of the microbiome's functional role in human metabolic disorder. Future studies must be designed to identify which bacterial populations or functional microbe-host relationships underlie this phenomenon.

1.1.5 Super-donors

The selection of a donor for FMT is not standardized, although there is general consensus for the need to do so [36]. Initially, donors were typically family members identified by the patient. However, recent studies highlight the practical advantages of using standardized volunteer donors and creating screened biobanks [31, 34]. In general, donors are screened for healthy bowel movements according to the Bristol stool chart, communicable diseases, recent travel history and antibiotic history.

In subsequent publications and conferences, Vrieze et al. noted that the patients who had a more robust improvement of insulin sensitivity after FMT received transplantation from the same limited number of donors [37]. That is, a minority of donor samples elicited a robust response, whereas other samples had no effect on patient's metabolism. The success of the intervention,

hence, could be attributed to "super-donors." Studies on the effects of FMT in alleviating symptoms of IBD have similarly observed that fecal samples from certain donors have a much greater therapeutic effect on multiple recipients [38]. Currently there is no way to identify super-donors until after experiments have started. More recent FMT studies try to identify super-donors earlier in order to perform more rigorous analysis of their microbiome for the identification of therapeutic microbiota, which could allow for the design of a better alternative to FMT.

There is a strong social stigma with FMT [39]. Because fecal matter is difficult to standardize, the ethical and social complications in transplanting feces, and the difficulty in monetization, alternatives to direct FMT are being actively pursued [40]. Gel capsules of fecal microbiota is a promising new technique which excludes the need for any gastrointestinal procedure [34, 41] and is preferred by patients [42, 43]. In fact, private companies already deliver FMT through oral capsules, mainly for the treatment of CDI. However, it is unclear whether these capsules are as effective as FMT itself.

Another potential treatment is to design and produce probiotics in a donor-independent fashion. For example, the Vrieze et al., study identified increased butyrate-producing microbes in patients with increased insulin sensitivity following FMT [8]. If the increase of butyrate-producing bacteria is important for improvement of metabolic symptoms, then there is a possibility for more direct treatment of metabolic syndrome through pro/pre-biotics, which would be easier to control and administer.

1.1.6 Potential risks

One challenge with FMT is the difficulty in finding accurate measures of adverse reactions. Thus far, a vast majority of recipients are ill and it is difficult to differentiate between normal

disease progression and the effects of FMT. Nevertheless, although hundreds of individuals have undergone FMT, few negative outcomes have been reported, even in immunocompromised patients [44]. The majority of negative symptoms reported are mild, including diarrhea or fever [45-47]. Mortality has been observed in FMT trials, however it was attributed to unrelated causes in severely ill or elderly patients. Microbiota can predispose susceptibility to atherosclerosis using causative evidence in mice and correlative evidence in humans [48]. In addition, the spread of transmissible disease, while not reported, is still a viable threat, especially to the immunocompromised (e.g. IBD patient on immunomodulatory therapy, HIV patient with CDI). These reports underscore the importance of rigorous donor screening. Finally, these risks have to be tempered with the morbidity and mortality associated with obesity and its associated metabolic diseases, which as of yet have few effective treatments.

Surprisingly, obesogenic properties of the gut microbiome can be transmitted through FMT as well. A case report documented the transmission of an obese phenotype from an overweight donor to a lean patient following FMT for the CDI treatment [49]. The donor was a young, obese relative undergoing rapid weight gain at the time of donation. The recipient was an individual who had never been obese. After receiving FMT, the recipient had rapid unintentional weight gain that could not be explained by recovery from CDI alone. Interestingly, the recipient reported increased appetite. These observations remain controversial given that it's a case report. However, it is consistent with rodent studies where transfer of fecal matter from obese mice to germ-free mice transmits the metabolic phenotype [35]. Regardless, the results of this report have affected FMT protocol at many institutions that now exclude obese donors from donating.

1.1.7 Conclusion

FMT remains an exciting therapy with abundant potential. Nevertheless, there has been a lack of controlled, randomized trials for metabolic disease. Initially, the FDA considered FMT an IND, making it difficult for practitioners to use until all other therapeutic options had been exhausted. However, in 2014 the FDA stated that it would exercise enforcement discretion, allowing physicians to use FMT without IND applications for the treatment of CDI. For more investigational indications of FMT, an IND application with the FDA is still required.

Given the amount of controlled clinical studies currently testing FMT for metabolic syndrome we should have a clear indication in the next few years of whether or not microbiota changes are causative or correlative in this rising epidemic, and whether altering the gut microbiome through FMT or similar procedures will provide new therapeutic options for obesity and its associated metabolic disorders.

1.1.8 Acknowledgements

Chapter I, Introduction part 1, in full, is a reprint of previously published material: Marotz, Clarisse A., and Amir Zarrinpar. *Focus: Microbiome: Treating Obesity and Metabolic Syndrome with Fecal Microbiota Transplantation*. The Yale journal of biology and medicine 89.3 (2016): 383. I was the primary investigator and author of this paper. The co-author listed above supervised and provided support for the research and has given permission for the inclusion of the work in this dissertation.

1.1.9 References

1. Tremaroli V, Backhed F. Functional interactions between the gut microbiota and host metabolism. Nature. 2012;489(7415):242-9. doi: 10.1038/nature11552. PubMed PMID: 22972297.

2. Gerard P. Gut microbiota and obesity. Cell Mol Life Sci. 2016;73(1):147-62. doi: 10.1007/s00018-015-2061-5. PubMed PMID: 26459447.

3. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010;464(7285):59-65. doi: 10.1038/nature08821. PubMed PMID: 20203603; PubMed Central PMCID: PMCPMC3779803.

4. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, et al. Evolution of mammals and their gut microbes. Science. 2008;320(5883):1647-51. doi: 10.1126/science.1155725. PubMed PMID: 18497261; PubMed Central PMCID: PMC2649005.

5. Greenblum S, Turnbaugh PJ, Borenstein E. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(2):594-9. doi: 10.1073/pnas.1116053109. PubMed PMID: 22184244; PubMed Central PMCID: PMCPMC3258644.

6. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. Nature. 2006;444(7122):1022-3. doi: 10.1038/4441022a. PubMed PMID: 17183309.

7. Preidis GA, Versalovic J. Targeting the human microbiome with antibiotics, probiotics, and prebiotics: gastroenterology enters the metagenomics era. Gastroenterology. 2009;136(6):2015-31. PubMed PMID: 19462507; PubMed Central PMCID: PMCPMC4108289.

8. Kristensen NB, Bryrup T, Allin KH, Nielsen T, Hansen TH, Pedersen O. Alterations in fecal microbiota composition by probiotic supplementation in healthy adults: a systematic review of randomized controlled trials. Genome medicine. 2016;8(1):52. doi: 10.1186/s13073-016-0300-5. PubMed PMID: 27159972; PubMed Central PMCID: PMCPMC4862129.

9. van Nood E, Vrieze A, Nieuwdorp M, Fuentes S, Zoetendal EG, de Vos WM, et al. Duodenal infusion of donor feces for recurrent Clostridium difficile. The New England journal of medicine. 2013;368(5):407-15. doi: 10.1056/NEJMoa1205037. PubMed PMID: 23323867.

10. Zhang F, Luo W, Shi Y, Fan Z, Ji G. Should we standardize the 1,700-year-old fecal microbiota transplantation? Am J Gastroenterol. 2012;107(11):1755; author reply p -6. doi: 10.1038/ajg.2012.251. PubMed PMID: 23160295.

11. Kassam Z, Lee CH, Yuan Y, Hunt RH. Fecal microbiota transplantation for Clostridium difficile infection: systematic review and meta-analysis. Am J Gastroenterol. 2013;108(4):500-8. doi: 10.1038/ajg.2013.59. PubMed PMID: 23511459.

12. Vrieze A, Van Nood E, Holleman F, Salojarvi J, Kootte RS, Bartelsman JF, et al. Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. Gastroenterology. 2012;143(4):913-6 e7. doi: 10.1053/j.gastro.2012.06.031. PubMed PMID: 22728514.

13. Di Luccia B, Crescenzo R, Mazzoli A, Cigliano L, Venditti P, Walser JC, et al. Rescue of Fructose-Induced Metabolic Syndrome by Antibiotics or Faecal Transplantation in a Rat Model of Obesity. PloS one. 2015;10(8):e0134893. doi: 10.1371/journal.pone.0134893. PubMed PMID: 26244577; PubMed Central PMCID: PMCPMC4526532.

14. Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. Science. 2005;307(5717):1915-20. doi: 10.1126/science.1104816. PubMed PMID: 15790844.

15. Shen J, Obin MS, Zhao L. The gut microbiota, obesity and insulin resistance. Mol Aspects Med. 2013;34(1):39-58. doi: 10.1016/j.mam.2012.11.001. PubMed PMID: 23159341.

16. Everard A, Cani PD. Diabetes, obesity and gut microbiota. Best Pract Res Clin Gastroenterol. 2013;27(1):73-83. doi: 10.1016/j.bpg.2013.03.007. PubMed PMID: 23768554.

17. Matsumoto M, Kibe R, Ooga T, Aiba Y, Kurihara S, Sawaki E, et al. Impact of intestinal microbiota on intestinal luminal metabolome. Scientific reports. 2012;2:233. doi: 10.1038/srep00233. PubMed PMID: 22724057; PubMed Central PMCID: PMCPMC3380406.

18. Utzschneider KM, Kratz M, Damman CJ, Hullarg M. Mechanisms Linking the Gut Microbiome and Glucose Metabolism. J Clin Endocrinol Metab. 2016;101(4):1445-54. doi: 10.1210/jc.2015-4251. PubMed PMID: 26938201; PubMed Central PMCID: PMCPMC4880177.

19. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444(7122):1027-31. doi: 10.1038/nature05414. PubMed PMID: 17183312.

20. Calkin AC, Tontonoz P. Transcriptional integration of metabolism by the nuclear sterol-activated receptors LXR and FXR. Nature reviews Molecular cell biology. 2012;13(4):213-24. doi: 10.1038/nrm3312. PubMed PMID: 22414897; PubMed Central PMCID: PMC3597092.

21. Sayin SI, Wahlstrom A, Felin J, Jantti S, Marschall HU, Bamberg K, et al. Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. Cell metabolism. 2013;17(2):225-35. doi: 10.1016/j.cmet.2013.01.003. PubMed PMID: 23395169.

22. Zarrinpar A, Loomba R. Review article: the emerging interplay among the gastrointestinal tract, bile acids and incretins in the pathogenesis of diabetes and non-alcoholic fatty liver disease. Alimentary pharmacology & therapeutics. 2012;36(10):909-21. doi: 10.1111/apt.12084. PubMed PMID: 23057494; PubMed Central PMCID: PMC3535499.

23. Pols TW, Noriega LG, Nomura M, Auwerx J, Schoonjans K. The bile acid membrane receptor TGR5 as an emerging target in metabolism and inflammation. Journal of hepatology. 2011;54(6):1263-72. doi: 10.1016/j.jhep.2010.12.004. PubMed PMID: 21145931; PubMed Central PMCID: PMC3650458.

24. Zhang C, Zhang M, Wang S, Han R, Cao Y, Hua W, et al. Interactions between gut microbiota, host genetics and diet relevant to development of metabolic syndromes in mice. ISME J. 2010;4(2):232-41. doi: 10.1038/ismej.2009.112. PubMed PMID: 19865183.

25. Zarrinpar A, Chaix A, Yooseph S, Panda S. Diet and feeding pattern affect the diurnal dynamics of the gut microbiome. Cell metabolism. 2014;20(6):1006-17. doi: 10.1016/j.cmet.2014.11.008. PubMed PMID: 25470548; PubMed Central PMCID: PMC4255146.

26. Thaiss CA, Zeevi D, Levy M, Zilberman-Schapira G, Suez J, Tengeler AC, et al. Transkingdom control of microbiota diurnal oscillations promotes metabolic homeostasis. Cell. 2014;159(3):514-29. doi: 10.1016/j.cell.2014.09.048. PubMed PMID: 25417104.

27. Leone V, Gibbons SM, Martinez K, Hutchison AL, Huang EY, Cham CM, et al. Effects of diurnal variation of gut microbes and high-fat feeding on host circadian clock function and metabolism. Cell host & microbe. 2015;17(5):681-9. doi: 10.1016/j.chom.2015.03.006. PubMed PMID: 25891358; PubMed Central PMCID: PMC4433408.

28. Liang X, Bushman FD, FitzGerald GA. Rhythmicity of the intestinal microbiota is regulated by gender and the host circadian clock. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(33):10479-84. doi: 10.1073/pnas.1501305112. PubMed PMID: 26240359; PubMed Central PMCID: PMCPMC4547234.

29. Grehan MJ, Borody TJ, Leis SM, Campbell J, Mitchell H, Wettstein A. Durable alteration of the colonic microbiota by the administration of donor fecal flora. J Clin Gastroenterol. 2010;44(8):551-61. doi: 10.1097/MCG.0b013e3181e5d06b. PubMed PMID: 20716985.

30. Lee CH, Steiner T, Petrof EO, Smieja M, Roscoe D, Nematallah A, et al. Frozen vs Fresh Fecal Microbiota Transplantation and Clinical Resolution of Diarrhea in Patients With Recurrent Clostridium difficile Infection: A Randomized Clinical Trial. JAMA : the journal of the American Medical Association. 2016;315(2):142-9. doi: 10.1001/jama.2015.18098. PubMed PMID: 26757463.

31. Hamilton MJ, Weingarden AR, Sadowsky MJ, Khoruts A. Standardized frozen preparation for transplantation of fecal microbiota for recurrent Clostridium difficile infection. Am J Gastroenterol. 2012;107(5):761-7. doi: 10.1038/ajg.2011.482. PubMed PMID: 22290405.

32. Cui B, Xu F, Zhang F. Methodology, Not Concept of Fecal Microbiota Transplantation, Affects Clinical Findings. Gastroenterology. 2016;150(1):285-6. doi: 10.1053/j.gastro.2015.05.065. PubMed PMID: 26616573.

33. van der Waaij LA, Mesander G, Limburg PC, van der Waaij D. Direct flow cytometry of anaerobic bacteria in human feces. Cytometry. 1994;16(3):270-9. doi: 10.1002/cyto.990160312. PubMed PMID: 7924697.

34. Youngster I, Sauk J, Pindar C, Wilson RG, Kaplan JL, Smith MB, et al. Fecal microbiota transplant for relapsing Clostridium difficile infection using a frozen inoculum from unrelated donors: a randomized, open-label, controlled pilot study. Clin Infect Dis. 2014;58(11):1515-22. doi: 10.1093/cid/ciu135. PubMed PMID: 24762631; PubMed Central PMCID: PMCPMC4017893.

35. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, et al. Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science. 2013;341(6150):1241214. doi: 10.1126/science.1241214. PubMed PMID: 24009397.

36. Borody TJ, Paramsothy S, Agrawal G. Fecal microbiota transplantation: indications, methods, evidence, and future directions. Curr Gastroenterol Rep. 2013;15(8):337. doi: 10.1007/s11894-013-0337-1. PubMed PMID: 23852569; PubMed Central PMCID: PMCPMC3742951.

37. Nieuwdorp M, Gilijamse PW, Pai N, Kaplan LM. Role of the microbiome in energy regulation and metabolism. Gastroenterology. 2014;146(6):1525-33. doi: 10.1053/j.gastro.2014.02.008. PubMed PMID: 24560870.

38. Vermeire S, Joossens M, Verbeke K, Wang J, Machiels K, Sabino J, et al. Donor Species Richness Determines Faecal Microbiota Transplantation Success in Inflammatory Bowel Disease. J Crohns Colitis. 2016;10(4):387-94. doi: 10.1093/ecco-jcc/jjv203. PubMed PMID: 26519463.

39. Anderson JL, Edney RJ, Whelan K. Systematic review: faecal microbiota transplantation in the management of inflammatory bowel disease. Alimentary pharmacology & therapeutics. 2012;36(6):503-16. doi: 10.1111/j.1365-2036.2012.05220.x. PubMed PMID: 22827693.

40. Hawkins AK, O'Doherty KC. "Who owns your poop?": insights regarding the intersection of human microbiome research and the ELSI aspects of biobanking and related studies. BMC Med Genomics. 2011;4:72. doi: 10.1186/1755-8794-4-72. PubMed PMID: 21982589; PubMed Central PMCID: PMCPMC3199231.

41. Youngster I, Russell GH, Pindar C, Ziv-Baran T, Sauk J, Hohmann EL. Oral, capsulized, frozen fecal microbiota transplantation for relapsing Clostridium difficile infection. JAMA :

the journal of the American Medical Association. 2014;312(17):1772-8. doi: 10.1001/jama.2014.13875. PubMed PMID: 25322359.

42. Zipursky JS, Sidorsky TI, Freedman CA, Sidorsky MN, Kirkland KB. Patient attitudes toward the use of fecal microbiota transplantation in the treatment of recurrent Clostridium difficile infection. Clin Infect Dis. 2012;55(12):1652-8. doi: 10.1093/cid/cis809. PubMed PMID: 22990849.

43. Brandt LJ, Aroniadis OC, Mellow M, Kanatzar A, Kelly C, Park T, et al. Long-term follow-up of colonoscopic fecal microbiota transplant for recurrent Clostridium difficile infection. Am J Gastroenterol. 2012;107(7):1079-87. doi: 10.1038/ajg.2012.60. PubMed PMID: 22450732.

44. Kelly CR, Ihunnah C, Fischer M, Khoruts A, Surawicz C, Afzali A, et al. Fecal microbiota transplant for treatment of Clostridium difficile infection in immunocompromised patients. Am J Gastroenterol. 2014;109(7):1065-71. doi: 10.1038/ajg.2014.133. PubMed PMID: 24890442.

45. Baxter M, Colville A. Adverse events in faecal microbiota transplant: a review of the literature. J Hosp Infect. 2016;92(2):117-27. doi: 10.1016/j.jhin.2015.10.024. PubMed PMID: 26803556.

46. Kump PK, Grochenig HP, Lackner S, Trajanoski S, Reicht G, Hoffmann KM, et al. Alteration of intestinal dysbiosis by fecal microbiota transplantation does not induce remission in patients with chronic active ulcerative colitis. Inflamm Bowel Dis. 2013;19(10):2155-65. doi: 10.1097/MIB.0b013e31829ea325. PubMed PMID: 23899544.

47. Colman RJ, Rubin DT. Fecal microbiota transplantation as therapy for inflammatory bowel disease: a systematic review and meta-analysis. J Crohns Colitis. 2014;8(12):1569-81. doi: 10.1016/j.crohns.2014.08.006. PubMed PMID: 25223604; PubMed Central PMCID: PMCPMC4296742.

48. Gregory JC, Buffa JA, Org E, Wang Z, Levison BS, Zhu W, et al. Transmission of atherosclerosis susceptibility with gut microbial transplantation. The Journal of biological chemistry. 2015;290(9):5647-60. doi: 10.1074/jbc.M114.618249. PubMed PMID: 25550161; PubMed Central PMCID: PMCPMC4342477.

49. Alang N, Kelly CR. Weight Gain After Fecal Microbiota Transplantation. Open Forum Infectious Diseases. 2015;2(1):ofv004-ofv. doi: 10.1093/ofid/ofv004.

1.2

The Microbiome and Human Biology

Microbiome research has dramatically reshaped our understanding of human biology over the last few years. New insights range from an enhanced understanding of how microbes mediate digestion and disease processes e.g. inflammatory bowel disease, to surprising associations with Parkinson's Disease, autism and depression. In this review we describe how multiple new generations of sequencing technology, analytical advances coupled to new software capabilities, and the integration of animal model data have led to these new discoveries. We also discuss the prospects for integrating studies of the microbiome, metabolome and immune system towards the goal of elucidating mechanisms that govern their interactions. This systems-level understanding will change how we think about ourselves as organisms.

1.2.1 Introduction

When we think about what defines us as species, our thoughts naturally turn to the human genome. The Human Genome Project was a remarkable success in government-funded "Big Science": in 2013 in his State of the Union address, President Obama estimated the cost of the Human Genome Project at $3.8 billion, with a return on investment (ROI) of an incredible 140:1. Yet the human genome, which is essentially fixed at birth, represents a small fraction of genetic diversity associated with our bodies. Estimates of the gene content in microbiome, either from back-of-the-envelope calculations [1] to empirical observations [2, 3] place the number of microbial genes associated with our human body from 2-20 million, exceeding the ~20,000 human genes by at least a factor of 100. Microbial cells even outnumber our own cells; although the early and widely reproduced estimates of 10:1 microbial to human cells are overstated. The most

detailed report to date suggests that we are only about 47% human on average by cell count [4] (of course, because microbial cells are much smaller, this corresponds to only a couple of kilograms of microbial biomass in a typical adult). The impact of this enormous number of microbial genes and cells on human biology must be profound. Furthermore, unlike our fixed human genomes, our microbial gene repertoires are highly malleable, offering exciting prospects for novel therapies.

Our ability to read out these complex microbial communities and understand their impact on human biology has been transformed by advances in technology, especially DNA sequencing and computational methods. In just a decade, a typical study has advanced from collecting a few dozen sequences for each sample to a few hundred million. These advances open up a panoramic vista of how microbes change over space and time, and how they relate to processes ranging from the physiological and psychological.

1.2.2 Early development of the microbiome field

The term “microbiome” was often taken to refer to the collection of genes contained within a community of microbes, although today the term is also used to refer to the organisms themselves (often termed the “microbiota”, although we defer to common usage in this article). However, the first appearance of “microbiome” in the literature did not include a definition of relatedness according to genomic criteria. It represented the collection of all taxa that comprise microbial communities including bacteria, fungi and protists in the intestinal tract and their relationship to microbes in the oral cavity [5] or protozoan populations in sewage contaminated environmental waters [6]. At that time, microbiology systematics lacked a compelling phylogenetic context. Efforts by Stanier and van Niel [7] to transform the morphology-based microbial systematics of Bergey’s manual [8] into a phylogenetic framework had reached an impasse [9]. Morphology, cell

staining characteristics, physiological properties, capacity for biochemical reactions, and other anecdotal features could not identify membership or evolutionary relatedness between phylogenetic assemblages.

Resolution of this problem and the foundation for today's microbiome studies emerged from the physicist Carl Woese's quest for the origin of life. Influenced by the Zuckerkandl and Pauling publication "Molecules as Documents of Evolutionary History [10], Woese predicted that molecular analysis of RNA components of the ribosome would reveal primordial branching patterns in the tree of life. He reasoned that the earliest life forms must have invented protein synthesis machinery with RNA components that would then be locked inflexibility with their binding partners and therefore evolve slowly. Specifically, interactions with the multi-protein complexity of the ribosome and with all other cellular proteins during their synthesis would impose strong evolutionary constraints on the ribosomal RNAs.

Even before the development of any kind of DNA or RNA sequencing technology, RNA/DNA competition hybridization experiments confirmed Woese's hypothesis by demonstrating that ribosomal RNAs evolved 100-fold more slowly than protein-coding regions in bacterial genomes [11]. Comparisons of the fragmentary rRNA sequence information captured by the then "state-of-the-art" two dimensional oligonucleotide fingerprinting technology, initially for 5S rRNAs [11] but ultimately and more comprehensively for 16S rRNAs, provided a phylogenetic framework that still underpins contemporary microbial ecology and systematics [12, 13]. These early oligonucleotide cataloging efforts redefined microbial systematics through the discovery of 11 major microbial phyla, but more profoundly led to the discovery of a third domain of life, the Archaea [13, 14]. When coupled with the first report of a full-length 16S rRNA sequence [15], the comparisons of oligonucleotide catalogues from hundreds of cultured organisms revealed a series

of interspersed fast- and slow-evolving regions that would eventually provide important technical advantages for contemporary microbiome investigations. This interleaved conservation pattern of slow and fast evolving regions allows rRNAs to serve as multi-handed molecular chronometers over disparate evolutionary time scales [16]. The slow-evolving regions allow amplification primers to be constructed (see below), and differences between the highly-conserved regions identify phylogenetic relationships that span the longest evolutionary time scales. In contrast, the fast-evolving regions provide fine-scale resolution often suitable for human microbiome investigations. Because the rRNA evolves at different rates in different taxa, even the full-length rRNA sequence does not permit resolution of all taxa, but in general the faster-evolving regions provide better resolution that is often important for clinical or forensic questions.

The introduction of labor-intensive Maxam-Gilbert DNA sequencing [17] and Sanger dideoxy chain termination sequencing [18] led to a rapid expansion of cultivar rRNA databases that included sequences 100-300 nucleotides long. The conserved regions served as primer sites for reverse transcriptase-mediated direct sequencing [19] of purified 16S rRNAs or chain elongation sequencing of cloned rRNA genes [20]. The molecular biologist Norman Pace realized the revolution in microbial systematics enabled by molecular phylogenetics would have a profound impact on microbial ecology [21]. The initial molecular surveys of microbial diversity from environmental samples required the isolation and DNA sequencing of cloned rDNA inserts from recombinant libraries. Each sequence served as a proxy for the occurrence of a microbial genome in an environmental sample. Comparisons of environmental rDNA sequences to the rapidly growing rRNA database from cultivars revealed microbial diversity to be at least two orders of magnitude greater than previously appreciated. New technologies including the adaptation of polymerase chain reactions for amplifying and cloning rRNA gene sequences [22] and automated

DNA sequencing machines accelerated the rate of discovery [23]. Woese's eleven major bacterial phyla grew to more than 1000 [24]. Finally, the development of next-generation sequencing technologies capable of generating millions of reads for less than a penny/read once again pushed back the known limits of microbial diversity, much of which represents the very low abundance taxa that make up the rare biosphere [25].

1.2.3 The move to next-generation sequencing

Early studies of the microbiome were limited to relatively small numbers of sequences because of the costs and time required for cloning and Sanger sequencing (typically, several dollars per sequence). These studies traditionally focused on rRNA gene amplification by PCR, then relating the individual sequences to one another via multiple sequence alignment and subsequent phylogenetic reconstruction (e.g., [26]). One critical aspect in these phylogenetic analyses was the assessment of the taxonomic composition of a community or set of communities, often placing the new sequences into a characterized reference such as the Ribosomal Database Project [27] or into a phylogenetic tree with new sequences added to those with existing taxonomy and/or environment annotation in Genbank. These workflows, going from sequence to phylogeny, became commonplace and tools such as ARB [28], which allowed users to align sequences, insert them into a phylogeny, and visualize the tree, became widespread.

Because cost limited the number of sequences available per sample, obtaining abundance estimates of the organisms corresponding to the sequences was problematic, especially because many investigators used cheaper fingerprinting methods to choose only unique and diverse representative molecules for sequencing. Additionally, it was clear that the diversity was very high, with many new unknown sequences in each new dataset (including new sequences that were

artifacts of sequencing error), exceeding the capacity of phylogenetic reconstruction software and the computers of the time. Exacerbating the issue, there was considerable controversy about which phylogenetic methods to use, especially because theoretically more powerful methods such as maximum likelihood were limited to a few hundred sequences at most. To address these issues, researchers began to cluster sequences into operational taxonomic units (OTUs) often using a level of 97% sequence identity as a proxy for species; the number of sequences present in an OTU acted as a proxy for its abundance and allowed ecological diversity estimates to be computed [29]. Naively, the clustering could be performed using BLAST [30]. However, this approach was slow even with small numbers of sequences, because early methods relied on computing each pair of distances between sequences. Optimized methods for such calculations, notably DOTUR [31], which leveraged PHYLIP [32] to compute the distances, greatly improved the ability to generate OTU tables and pick representative sequences for phylogenetic reconstruction. DOTUR made the important contribution of refining the concept of an OTU beyond the idea of an identity threshold by introducing different methods of clustering (essentially, asking whether the threshold defined the maximum difference between any two sequences in an OTU, the average difference between any two sequences, or the maximum difference to the nearest sequence).

The adoption of the OTU concept paved the way for classically trained ecologists to get involved in microbiome research, bringing with them decades of ecological theory and methods developed for the qualitative and quantitative study of macroscopic environments. In particular, they brought approaches suited for assessing the relationships between environmental factors and the organisms in a community such as ordination techniques [33]. One extremely powerful concept was the notion of beta diversity, which represents the dissimilarity between a pair of samples. These distances can be computed pairwise across a large number of samples to produce a matrix

that describes how similar every sample is to every other sample. These distance matrices can then be assessed for the presence of systematic structure through techniques like PERMANOVA [34] and principal coordinates analysis.

There are many ways to compute beta diversity, but not all are equally informative for studying microbial communities [35]. In particular, methods that do not explicitly take phylogeny into account tend to underperform in expressing differences between samples, and can leading to inaccurate or implausible biological conclusions. In the mid 2000s, Cathy Lozupone asked whether including phylogeny could improve comparisons of microbial communities, and developed UniFrac [36] to perform such comparisons. Applying UniFrac to meta-analysis of over 100 existing 16S rRNA studies spanning every conceivable natural environment on the planet rapidly yielded a remarkable pattern: the structure of a microbial community is influenced more by whether it was sourced from a saline or non-saline environment than any other factor recorded [37]. However, once samples from vertebrate guts were included, the split between host-associated and non-host-associated samples was even more profound than the saline-non-saline split [38], suggesting that microbial communities residing within hosts are uniquely specialized relative to other communities.

In parallel with analytic improvements such as UniFrac, the price of DNA sequencing dropped precipitously due to the huge investment in sequencing technology from the Human Genome Project. In particular, the advent of pyrosequencing, which performs sequencing by synthesis and detects the incorporation of a nucleotide by monitoring the release of pyrophosphate (39). In addition to a cost reduction, a single sequencing run could produce hundreds of thousands of sequences without the need of laborious clone libraries used by Sanger sequencing, albeit with increased sequencing error. These errors are particularly problematic when assessing the presence

of rare members of a community, because an erroneous sequence may appear real. Relatively simple criteria could be applied to remove many of the low quality reads, yielding error rates that were better than Sanger sequencing [40, 41]. The first application of pyrosequencing to 16S rRNA data yielded an unprecedented view into the "rare biosphere" [25], revealing an extremely long tail of rare taxa in marine ecosystems that was rapidly extended to other environments.

Although pyrosequencing was cheaper than traditional capillary-based methods, it was not cost effective to perform a whole sequencing run per sample. For example, in 2007 the 454 platform typically yielded about 500,000 reads for $12,000; while this reduced a multi-year, multi-million dollar sequencing exercise to an 8-hour project, the cost per sample was still prohibitive for most applications. Simply combining libraries together would have made it impossible to track each sequence back to the sample from which it came. This challenge led to methods for multiplexing samples together such as by ligating nucleotide tags unique to each sample during PCR [42] or by including these tags on the PCR primers themselves, allowing readouts of multiple communities simultaneously [43]. Including formal "error-correcting" codes in the construction of these barcodes made it possible to tolerate errors within the barcode itself, and allowed confident multiplexing of hundreds of samples simultaneously on a run [44].

Through projects such as the Human Microbiome Project (HMP) [3] and MetaHIT [2], these advance in sequencing quickly moved the microbiome field from a data-poor science to a data-rich one. This wealth of data quickly resulted in computational bottlenecks. This analysis glut prompted the development of tools such as mothur [45] and QIIME [46]. The latter is an accessible and modular microbiome-focused analysis framework that can handle large volumes of data and operate in environments ranging from laptops to supercomputers.

The modularity of the QIIME platform greatly simplified subsequent and inevitable technology transitions. For example, at the same time as the HMP was being performed, Illumina sequencing technology began to offer tens of millions of sequences per run. We used QIIME to process the 454 amplicon data in the HMP. Within the same issue of Nature in 2012, we also used QIIME with data from the Illumina platform spanning over 1 billion amplicon reads, at a depth of coverage of over 1 million reads per sample [47].

A secondary benefit of the modularity of QIIME was the ability to add novel analyses methods easily. This approach allowed methods development to operate in concert with study design, such as the exploration of the biogeography of microbes on a person's hands [48]. Similarly, the flexibility of the platform made it possible to utilize the statistical and visualization components on different types of data beyond 16S rRNA (for example, shotgun metagenomic data and metabolomics) to study the distribution of small molecules across the body in relation to the microbial inhabitants [49]. Critically, this approach standardizes a large portion of the bioinformatics pipeline and reduces technical bias, which can often outweigh the biological differences among samples [50].

Most amplicon-based microbiome studies still perform OTU level analyses. However, in the last few years, there has been a strong motivation to go to the "sub" OTU level as methods have improved. One method, oligotyping [51], allowed researchers to partition OTUs into finer groups in a supervised fashion based on Shannon entropy of the variation within nucleotide positions, and to then assess whether these partitions were significantly correlated to study covariates. A related method, MED [52] operates in an unsupervised manner but is similar in concept. More recently, DADA2 [53] and Deblur [54] leverage the error profiles of the sequencing instruments themselves to determine the most probable molecules that were presented to the

sequencing instrument. Excitingly, although these approaches do not offer assured species or strain level resolution because the variation need to distinguish does not always exist in the target amplicon, they offer maximal precision and specificity from the data obtained.

Next-generation sequencing also enabled shotgun metagenomics, essentially isolating total DNA from a sample, fragmenting it, and sequencing all the fragments. Shotgun metagenomics has overall been adopted less than amplicon sequencing for human microbiome studies due to the far greater depth of coverage needed, and, correspondingly, cost, although it has been critically important for obtaining functional as well as taxonomic insight into the human microbiome (see examples below). A particularly exciting emerging area is metagenomic assembly, where complete genomes can be assembled from metagenomic data, allowing very detailed tracking of the colonization of individuals by specific strains of bacteria during development [55]. However, earlier gene-based approaches such as those used in MetaHIT [2] and the HMP [3], especially when correlations among reads across multiple samples are used to generate co-abundance groups corresponding to genomes or large genome fragments [56], have been very useful for functional investigations.

1.2.4 Placing the human microbiome in context

Next-generation sequencing techniques and the tools developed to effectively handle the data produced from these sequencing efforts have enabled deep exploration of the microbiome not possible even a decade ago. Several groups were studying the human microbiome, particularly the skin microbiome, in the early 2000s, and the first shotgun metagenomic analysis of the gut was performed using Sanger sequencing in 2006 [57]. Many patterns, such as the high level of diversity among individuals in the human gut [58] and the profound differences among human body sites

[59] were clear from the first amplicon studies to systematically explore these topics. However, microbiome research exploded after the release of data collected from the first large-scale effort to characterize the healthy human microbiome across the body, the Human Microbiome Project (HMP). The HMP was an NIH-funded multi-million dollar project which many components, but among the largest was a description of the microbiomes of up to 18 body sites in 242 healthy humans; the findings were published in two companion papers in Nature in 2012 [3, 60]. Unsurprisingly, drastically different microbial communities, both in terms of diversity and composition, were harbored by different body sites in the HMP cohort, consistent with earlier results from amplicon sequencing alone [59]. Stool and oral communities were the most diverse microbial communities in terms of number of different organisms present, and the microbial communities of vaginal samples proved least diverse, comprised mainly of *Lactobacillus spp.*[60]. Interestingly, sub-locations of body site classes (i.e., skin, mouth, vagina, or gut) harbor specific microbial communities; for example, while all locations in the oral cavity harbor *Streptococcus* as a major taxonomic group, the second most dominant group is different in the buccal mucosa (*Haemophilus*), supragingival plaque (*Actinomyces*), and subgingival plaque (*Prevotella*) [3]. Individual differences in the microbiome are sufficiently large and reproducible that they may be useful for forensic purposes [61, 62], and can even match up family members [63] and sexual partners [64], although courtroom applications of such technologies remain for the future.

Cross-individual differences were specific to body site; for example, while an individual can harbor a variety of species in their oral cavity, all individuals in the HMP cohort appear to carry the same or similar types of organisms (*Streptococcus, Neisseria, Haemophilus, Veillonella*), and while the skin microbiome in general is not highly diverse for a single individual (dominated by *Propionibacterium, Staphylococcus,* or *Corynebacterium*), individuals tend to harbor markedly

different communities from one another [3]. Stool samples collected from individuals in the cohort exhibited tremendous variability, ranging from complete dominance by the Firmicutes to complete dominance by Bacteroidetes (the two major bacterial phyla in the gut). Despite this wide variety in taxonomic composition, both within body sites and between body sites, metabolic and functional pathways in the metagenomes were much more constant and evenly diverse, with several "core" pathways including ribosome and translational machinery, ATP synthesis, and glycolysis ubiquitous across body sites and individuals [3], consistent with previous work identifying the same patterns in the gut alone [65]. This observation lends strong support to a multi-omics approach toward characterizing healthy (and diseased) microbiomes, as it will not be sufficient to determine the microbial composition of a community alone.

Although humans have explored a remarkably wide ecological niche relative to other vertebrate species, with an unprecedented geographic range and diversity of different diets, placing the human microbiome in the context of other mammals suggests that we are relatively unremarkable. Both at the level of taxonomy and function, we resemble other omnivorous primates such as chimpanzees closely [66, 67]. In general, mammals with the same diet and gut physiology harbor similar microbiomes, although there is substantial phylogenetic inertia, such that bears, which have diversified in a span of ~5 million years into obligate carnivores (polar bears), herbivores (pandas), and omnivores all harbor similar microbiomes [66]. Comparative studies of the microbiomes of great apes (including humans) show a considerable level of co-phylogeny, such that the gut microbiome tracks the overall pattern of host evolution [68, 69]. Unfortunately, very little data is available for other body sites such as the oral and skin communities, in part due to the difficulty of collection. Intriguingly, different species of apes have been observed to

converge in their gut microbiota when living in the same habitat, suggesting that there may be inter-species transfer [70].

The mouse microbiome is notably different from the human microbiome, and although the dominant phyla (Firmicutes and Bacteroidetes) are the same, the genus-level composition is notably different. In general, mouse samples are perfectly separable from human samples using techniques such as principal coordinates analysis [71]. This has important implications for the use of mouse models to make inferences about human biology: in general, because the background microbiota are so different, mouse studies are more effective for demonstrating possible mechanisms than identifying specific taxa associated with human biological processes. The use of gnotobiotic mice, i.e. mice colonized with known microbial communities such as those derived from human samples, is extremely useful in understanding the impact of human-associated bacteria on conserved mechanisms in the host, and for example individual human microbiomes transferred to mice can confer phenotypes ranging from adiposity [72] to features resembling Parkinson's disease [73]. Perhaps most excitingly, *Christensenella*, a microbe associated with low body mass indices in human twin studies, causes mice inoculated with an obesogenic microbiome from humans to remain lean [74], demonstrating the ability to move from population-level observations to mechanistic work in an individual microbe. However, human-derived microbes are displaced from gnotobiotic mice on exposure to mouse-derived microbes within a few days [75], underscoring the requirement to keep such mice completely isolated environmentally.

Placing the human microbiome in an environmental context, the microbiology of the built environment consists mainly of human-derived microbes with the dominant input being from the skin, and the individual-specific nature of this input can be tracked longitudinally to demonstrate transmission of microbes from individual people to surfaces they touch and spaces they inhabit

[76]. After death, the microbiome undergoes a specific suite of changes mixing endogenous community members with those derived from the soil, in a pattern so systematic that it can be used to estimate postmortem interval [77]. Intriguingly, the skin acts as a taxonomic bridge between other human body sites and environmental microbiomes, being dominated by Firmicutes (common in other human body sites), Actinobacteria (common in both human body sites and in the environment), and Proteobacteria (more common in environmental samples) [38]. However, the profound differences between host-associated and free-living microbial communities also apply to the human microbiome specifically, and this factor dominates large-scale comparisons of microbiomes.

1.2.5 Associations between the microbiome and human development

Although most studies of the human microbiome to date have focused on adults, there is immense interest in the microbiome during development both because of the profound changes that it undergoes during this process and because of the exciting prospects for early-life interventions that could promote health over a lifetime. Additionally, the microbiome appears to have far-reaching effects on reproductive and developmental biology that were unanticipated until recently.

The vaginal microbiome of pregnant women differs from that of the general population [78, 79], although the degree of stability within the community depends on its composition [79, 80] Overall, pregnancy is associated with a loss of microbial diversity within the vaginal community [78], and transition toward community structures dominated by *Lactobacillus* [78-80]. These shifts may be hormonally driven, with the bacterial community responding to increased estrogen during gestation [80]. The elevated *Lactobacillus* may also be protective for the

developing infant. *Lactobacillus*-dominated communities protect against bacterial vaginosis [81], a defect in the vaginal microbiome that is associated with elevated risk of preterm birth [82].

The mother's gut microbiome is radically remodeled between the first and third trimester [83], with the third trimester microbiome being markedly different from that of non-pregnant women. Late-term pregnancy has been associated with increases in the relative abundance of the phyla Proteobacteria and Actinobacteria and a loss of community diversity [83], an aberrant community relative to that of healthy adult women [83].

The changes in the maternal microbiome seeds the infant's first microbial communities. There is some controversy [84-86] whether this seeding occurs prenatally, or during birth. Recent papers [87, 88] have argued for a placental microbiome in the absence of amniotic infection. However, evidence suggests the observed placental microbiome may be a result of contamination in low-biomass samples [86]. One study was unable to detect differences between a placental microbiome and negative controls. Regardless of whether or not the microbial community is seeded prenatally, the mode of delivery has a strong role in shaping the neonatal microbiome. The microbial communities across multiple body sites of vaginally born infants more closely resemble an adult vaginal community, while infants born by C-section had communities that more closely resembled the skin microbiome [89, 90]. Additionally, birth method modulates the vertical transmission of gut microbes from mother to child [90], and the effect of vaginal birth could be seen in the microbiome through one year of age [90]. Studies in adults have not described a significant difference associated with delivery method, although it remains unclear whether the early-life changes in the microbiome can affect later phenotype by acting at a critical period in development. In mice, the time of colonization of initially germ-free mice with microbes

determines whether permanent alterations in behavior and in gene expression at distal tissues including the brain occur [91], and the same may be true in humans.

The microbiome undergoes rapid changes during the first three years of life, followed by a more gradual maturation [47, 90, 92]. The development of the infant microbiome correlates with changes in the breast milk composition and microbiome over the first year of development [90, 93–95]. Human breast milk contains unique oligosaccharides not found in any other mammalian milk [96]. Many of these sugars are recalcitrant to host digestion, but can be directly utilized as a primary carbon source by *Bifidobacterium*, and to a lesser extent, some *Bacteroides* species. This is directly reflected in the distinct infant gut microbiota which is dominated by *Bifidobacterium* [97]. Over the first four months of breast feeding, lactose increased while the concentration of both monosaccharides and oligosaccharides decreased [94]. The carbohydrate metabolism capacity of vaginally delivered, breast fed infants respond to this change in breast milk composition: at four months of age, infant metagenomes had an increased in abundance of lactose-specific transport genes [90].

Weaning forces a maturation of the microbiome [90, 92]. The introduction of solid food diversifies the microbiome and shifted carbohydrate metabolism to complex carbohydrates and starch [90]. Chronologically older infants who were exclusively breast feed appeared microbially younger than their peers who had been introduced to solid food. However, nutritional disruption can alter this maturation: children with persistent malnutrition were microbially delayed compared to their healthy peers [90, 92].

The widespread use of antibiotics, which began in the 1940s with large-scale production of penicillin (discovered in 1928; [98]) ushered in a new era of human medicine. Diseases and infections that had taken the lives of thousands were no longer considered dangerous, leading to

the designation of antibiotics as the “wonder” or “miracle” drugs. Inappropriate use of antibiotics, however, had led to a swath of serious health problems. Overuse of antibiotics in children in particular is a major public health problem. Most of the common illnesses experienced during childhood, such as diarrhea and upper respiratory tract infections (UTIs) are caused by viruses, not bacteria; therefore, the prescription of antibiotics in these cases is ineffective at best [99]. The detrimental effects of antibiotic usage on the gut microbiome have been described both in children and adults [100, 101]. Given that the first three years of life are a crucial time period for gut microbiome maturation, antibiotic treatments during this time period have the potential to inflict detrimental alterations to normal gut microbiome maturation, with potentially serious long-term effects. In mice, early life treatment with antibiotics increases weight gain and adiposity and delays microbiome maturation, while altering the metabolic activity of the fecal microbiome even into adulthood, long after antibiotic exposure [102]. The negative effects on microbiome maturation are also cumulative, becoming more pronounced with additional antibiotic courses, a significant observation given that the average child in the U.S. receives 10 courses of antibiotics by age 10 [102]. Similarly, the microbiomes of children exposed to multiple antibiotic treatments in the first three years of life have less diverse and more unstable gut microbiomes compared to their untreated counterparts [103]. Antibiotic resistance genes in the gut microbiome also rise sharply in this group after antibiotic administration [103]. Antibiotic treatment in children has also been associated with an increased risk for obesity [104], asthma and allergies [105, 106], diabetes [107], and inflammatory bowel disease [108], all diseases associated with dysbiosis and that, like antibiotic usage, have increased in prevalence over the recent decades.

Intriguingly, the rate of approach to the adult state is consistent in different populations, although the adult state that is reached in each case is markedly different [47]. Efforts to understand

the specific factors that lead to these cross-population differences in the adult microbiome are a major area of interest currently. Far less information is available about the maturation of the microbiome at other body sites, although characterizing these patterns of change will clearly be of great importance.

1.2.6 Large and small effects on the microbiome: what really matters?

As noted above, the human gut microbiome matures during the first three years of life, and remains relatively stable throughout adulthood [47] . Therefore, all factors influencing the microbiome likely have the greatest impact in this early developmental time window. Nevertheless, some factors have a dramatic impact on microbial composition at all life stages, and some factors have more subtle influences [109].

One of the most dramatic ways to influence the microbiome is with antibiotics. Even short term doses of antibiotics prescribed for acute infections can perturb gut microbial composition for years [110]. Resistance genes selected for during antibiotic treatment can persist in the microbial community long after the therapy has ended. While the degree and direction of dysbiosis in response to antibiotic treatment varies by individual, there are some overarching trends. For example, bacterial diversity decreases during the week immediately following antibiotic exposure and then begins to recover, although often the original state is not fully returned. In addition to changes in alpha diversity, antibiotic treatment can also have a significant microbial gene expression, protein activity, and overall microbial metabolic function [111].

Perhaps unsurprisingly, long-term diet is the primary determinant in the taxonomic and functional structure of the human gut microbiota. The nutritional composition of food affects which microbial species flourish, and, in some cases, selectively deplete certain taxa. While all

mammalian gut microbiomes share a core set of genes covering essential metabolic functions, the relative abundances of these genes and the specific taxa that carry them clearly distinguish carnivores, omnivores, and herbivores [67]. This leads to the observation that diet drives convergence of gut microbiomes across mammalian species.

The difference in the microbiome between western and nonwestern populations is profound, and likely driven at least in part by diet [47, 112], although controlled experiments to confirm this or detailed studies in immigrant populations have not yet been performed. Such studies comparing large numbers of populations, rather than individuals, will be essential for understanding the plasticity of function in the human microbiome in response to the many diverse diets represented in human populations worldwide.

Within the US population, long-term dietary patterns shape stable microbial communities, and small dietary changes are often not enough to disrupt the community factors that make an individual's microbiome unique [113, 114]. In two studies of short-term dietary interventions lasting less than a week, the microbiome reverted to its original state within days, and the magnitude of the change was smaller than the baseline differences among individuals [115]. In children with severe malnutrition, antibiotic treatment followed by a dietary intervention was not sufficient to alter microbial development long term, and the improvements in the microbial community structure associated with therapy required continuing dietary intervention [114, 116].

There are many examples emerging of the effects of individual dietary items. For example, dietary emulsifiers can disrupt the gut mucosal barrier inducing low-grade inflammation and changes in microbial composition [117]. Similar effects have been observed with artificial sweeteners, both in human and mouse studies, with individual differences in response largely explained by the microbiome [118]. The microbiome may explain the large individual-level

differences in response to specific dietary items that have long confounded weight loss studies. One recent groundbreaking study used continuous glucose monitors to explore post-prandial glucose response in a population of 800 people who were fed a controlled sequence of meals, allowing the effect of individual food items to be determined and related to various parameters including the microbiome [119]. Although the population average results recaptured the glycemic index for each food almost perfectly, individual variation in response was very high, and in a validation cohort of 100 individuals not involved in the initial study it was possible to design "good" and "bad" isocaloric, macronutrient-balanced diets for the same individual based on the microbiome, with markedly different impacts on glycemic response. These studies have tremendous potential to stratify patients effectively for dietary treatments of diabetes and obesity.

Host genetics also contribute to the structure of the microbiome, albeit with far smaller effect size. The best evidence for genetic influence on the microbiota in humans comes from twin studies comparing monozygotic and dizygotic twins. Initial 16S rRNA gene sequencing studies on dozens of people revealed no significant difference in the similarity between mono- and dizygotic twin gut microbiota, although both types of twins were more similar to their twin than to their mother or an unrelated adult [65]. However, a more targeted approach on a much larger cohort of hundreds of twin pairs revealed that community membership (alpha diversity) rather than community structure (relative abundances) are responsible for driving the similarities seen among monozygotic twins. While host genetics seem to play a relatively minor role in shaping the microbiome, certain taxa (for example *Christensenellaceae*) are indeed heritable.

The microbiome has been associated with many other processes, ranging from exercise [120] to infections [121] to stress [122] to sleep cycles [123], though typically with small effect size. Despite much interest in the possibility that probiotics modify the microbiome, the effects

tend to be very small, and may occur more at the level of transcription than of community change [124]. Identifying probiotics that have permanent, large effects on the microbiome remains an important topic for future research.

1.2.7 Microbiomes and disease

Bacteria in and on the human body have a significant impact on health and on the development of disease states. Microbiome alterations at different body sites have been associated with many diseases, which include perhaps obvious examples, such as dental caries and bacterial vaginosis; examples of chronic conditions, such as obesity, cardiovascular disease, inflammatory bowel disease and malnutrition; and even diseases not traditionally suspected to be linked to the microbiome, including as Parkinson's disease, autism and depression. A partial list of these diseases and their microbial links appears in Table 1; it is impossible to be comprehensive given the rapid discovery of new links between the microbiome and diseases.

Table 1.2.1. *Disease states and their microbial links.*

Disease	Description & microbiome link	Reference
C. difficile associated diarrhea	Typical example of change in gut microbiome leading to enduring disease state.	(149)
Obesity (metabolic disease)	Increased capacity of gut microbiome to harvest energy from the diet.	(150)
Inflammatory Bowel Disease	Gut inflammation disease driven by genetic, environmental and altered microbial factors. Adherent enterobacteria may promote initial ulceration events.	(151), (152)
Parkinson's Disease	The microbiome can promote disease progression in genetically susceptible individuals.	(73)
Atopic dermatitis	Skin inflammation driven by *Staphylococcus aureus* dominance (with genetic predisposition).	(126)
Acne vulgaris	Skin disorder mediated by certain *Propionibacterium acnes* strains, together with the vitamin B12 pathway.	(153)

Table 1.2.1. *Disease states and their microbial links, continued.*

Disease	Description & microbiome link	Reference
Chronic skin wounds	*Staphylococcus aureus*, *Pseudomonas aeruginosa* and other bacterial pathogenesis in chronic wounds.	(154)
Autism	Certain gut microbiota gained abundance in autism. Gut microbiota changes (in animal models) induce changes. Maternally produced microbially metabolites lead to an autism phenotype in mice.	(155, 156)
Asthma and allergies	Dust of traditional farms stimulates the immune response and protects against asthma and allergies.	(157)
Acute Anorexia	Lower gut alpha diversity in anorexia patients. Evidence that molecular mimicry of microbial metabolites may contribute to autoantibody production.	(158)
Rheumatoid Arthritis	Altered gut and oral microbiome in rheumatoid arthritis patients. RA patients have increased translocation of oral bacteria in the gut; treatment partially corrects this.	(145)
Atherosclerosis	Lack of intestinal symbiotic microbiota (in mice) induced plaque, while low levels of cholesterol were given in the diet.	(159)
Cystic Fibrosis	Characterized by chronic lung infections, commonly with hypermutable *Pseudomonas aeruginosa* strains.	(160, 161)
Bacterial vaginosis	Deviation from a low pH, *lactobacillus* dominated community to a higher pH, more diverse microbial community.	(162)
Dental caries	Increased phylogenetic diversity and overabundance of *Prevotella* taxa associated with dental caries.	(163, 164)
Depression	Transplantation of the microbiome into germ free mice induces depressions symptoms. These are associated with alterations in carbohydrate metabolism in the microbiome and hippocampus.	(131)
Type I diabetes	In mouse models, the microbiome is required for the development of diabetes, although low dose antibiotics increase susceptibility. Changes in microbial development mark the progression to disease but predate the clinical presentation.	(165, 166)
Type II diabetes	Lower levels of bacterial LPS in blood in type II diabetes patients.	(167)
Malnutrition	Altered gut microbiome strongly linked with childhood malnutrition.	(168)

Table 1.2.1. *Disease states and their microbial links, continued.*

Disease	Description & microbiome link	Reference
Cardiovascular disease	Diet & gut microbiome were linked with trimethylamine-N-oxide (TMAO) levels in plasma and cardiovascular disease risk (with genetic predisposition).	(169)
Multiple Sclerosis	Suggested evidence of gut microbiota changes related to autoimmunity and pathology of MS.	(170)
Alcoholic Liver Disease	Intestinal dysbiosis, bacterial overgrowth and increased gut permeability.	(171)
Osteoporosis	Direct and indirect impact of the gut microbiome on deregulated bone remodeling.	(128)
Colorectal cancer	Pathogenic microorganisms potentially initiate and facilitate the process of colorectal cancer	(129)
Addiction	Antibiotic treatment increased addictive behavior for animals receiving low dose opioids in a mouse model of addiction	(172)
Irritable Bowel Syndrome	Mucosal and luminal gut microbial changes, although causal effect is unproven.	(130)

In any study of the microbiome and disease, it is important to include the question of correlation versus causation. For example, *psoriasis vulgaris* patients have a different skin microbiome relative to healthy controls [125], but is this altered microbiome the cause of the disease state, or rather a consequence of the altered skin texture? For example, recent discoveries have suggested that atopic dermatitis may be caused by the altered microbiome [126], whereas the altered microbiome in psoriasis patients appears to be a side effect or an effect of the physiological changes of the skin [127]. Even if the microbiome is not causing a condition directly, the subsequent shift in the community may afford increased risk to other diseases, and furthermore, the microbiome may still be useful for assaying specific diseases or subtypes particularly in difficult to diagnose conditions like Crohn's Disease.

Interestingly, direct and indirect microbial links have been discovered for osteoporosis [128], colorectal cancer [129], irritable bowel syndrome [130] and mineral deficiency diseases, although it is still uncertain in these cases whether the altered microbiome is a cause or a consequence. An especially important method of determining causality has been demonstrating the ability to transfer human phenotypes to germ-free mice (wild type or other). This has now been done successfully for a number of diseases and conditions, including obesity [72], malnutrition [116], insulin resistance resulting from artificial sweeteners [118], depression [131], and even jetlag [132]. This technique has tremendous potential for establishing that microbes can cause a specific phenotype and for untangling molecular mechanisms, especially when combined with metabolomics, although negative results can be difficult to interpret in light of the many genetic and physiological differences between humans and mice. Gnotobiotic pigs, which provide a better physiological model, have been of intense interest in this respect recently, especially in studies of nutrition [133].

Like all other mammals, humans co-evolved with their microbiota and developed a wide range of innate immune responses to protect the body against infection while still sustaining bacterial presence. The gut microbiome is in constant and intimate interaction with the host immune system, and influences both the innate and adaptive immune function [134]. The innate immune response involves dendritic cells, neutrophils, natural killer cells. The adaptive immune response involves T and B cells activation. Specific microbiota are associated with the development of particular T cell subtypes [135]. A shift in the gut microbiome can cause beneficial as well as detrimental outcomes mediated by CD4+ T cell subtypes regulation [136]. This relationship between the gut microbiome and the immune system is strongly implicated in a range of inflammatory disorders. In particular, dysbiosis has been associated with increased oxidative

stress, which can drive a chronic inflammatory response [137] which is exacerbated by a decrease in community members known to produce short chain fatty acids (SCFAs) [138]. The lack of these SCFAs, such as butyrate, has the potential to promote an inflammatory response in the gut epithelium [139]. Given the association between inflammation and autoimmune diseases, it is no surprise that many of them have been linked with the gut microbiome.

However, the microbiome is only a piece of the puzzle in the development and progression of disease. The microbiome, along with genetic susceptibility, epigenetic regulation, and environmental factors create a complex interactome. Underlying genetic susceptibility is often required as an underlying etiology for conditions. While cystic fibrosis has long been considered a disease with classic Mendelian inheritance, the lung microbiome play an important role in long term prognosis [140]. In more complex genetic diseases such as inflammatory bowel disease [141] and Parkinson's disease [73, 142], the microbiome plays an important role in the etiology but is not causative alone. Crosstalk between the microbiome and epigenetic regulation may also modulate disease susceptibility, although the directionality of this interaction is unclear [143]. The role of diet and chemical exposure in disease management is also an important consideration. For instance, both diet and medication are used in the treatment of Type II diabetes. Conventional wisdom suggests type II diabetics avoid simple carbohydrates to control their blood sugar. However, there is a large degree of inter-individual glycemic response to foods. Some of this variation can be explained by the individual microbiome. Indeed, the microbiome is a far better predictor of a diet that limited the postprandial glucose spikes than a nutritionist [119]. The role of drug treatment in disease management, and the role of the microbiome in disease response, are also under active investigation. A large cross sectional study of patients with Type II diabetes found treatment with metformin was a better predictor of their microbiome than their diagnosis

[144]. Treatment played a role in microbiome remediation in Rheumatoid Arthritis: treatment disease-modifying antirheumatic drugs partially restore microbial balance [145]. However, microbiome can regulate the way in which drugs are metabolized: detrimental side effects of both metformin [144] and acetaminophen [146, 147] are related to microbial metabolism.

One challenge to overcome in therapeutic applications of the microbiome, however, is that our microbiomes are unique to each individual, and change over time. These properties complicate both diagnostic use of the microbiome and its viability as a therapeutic target. In diseases such as *C. difficile*, the shift in the microbial community is remarkable [148], but the picture is less clear cut for many of other associations of microbiome with disease. As in human genetics work, it is necessary to understand the scope of diversity associated with different human populations, as well as the context of lifestyle and diet, so that we can understand deviations from the norm and specific aspects of concern for a particular disease state relative to a particular background.

1.2.8 Conclusions

The microbiome is complex, dynamic, and spatially structured. It is clearly important for many physiological and disease processes in which its involvement was completely unsuspected until recently. With all this complexity, especially the emerging links to metabolism, host genetics, epigenetics, and immunology, one might easily give up hope of being able to untangle the complex relationships, let alone exploit them for therapeutic benefit. However, other complex fields of biology offer hope. For example, although mass spectrometry analysis of an orange yields many compounds that have not yet been identified, we know that the vitamin C it contains is essential for preventing scurvy. Similarly, we can expect that some effects on the microbiome will be so large that they can be identified and exploited in a systematic way for any individual (for example,

fecal transplant for recurrent *C. difficile*), whereas other dysbioses will be far more subtle and require individual-specific approaches, perhaps aided by model systems such as gnotobiotic mice, organoids, or in vitro models.

Although our capability to analyze the microbiome has expanded rapidly, it is clear that the ability to collect more sequences and timepoints will continue to revolutionize microbiome studies. Even depth of coverage of 1 million amplicon reads per sample merely scratches the surface of a 39 trillion-strong gut microbiota, and it is entirely possible that rare species that we cannot yet detect play important roles in many ecosystems and disease processes. Better methods of metagenome assembly and interpretation, especially from shallow-coverage samples, are urgently needed to provide statistical power for studies that need to cover hundreds to hundreds of thousands of individuals to reveal subtle effects. Computational methods, especially to integrate newly collected samples with large-scale resources such as the HMP, also need to be made far more accessible to a broad audience. However, we can imagine a day, perhaps soon, when easy tracking of the microbiome together with continuous analytics in the cloud, perhaps even at the consumer level (as it were, complementing one's Fitbit™ with a "Shitbit"), will place control of the microbiome for life-long health in the hands of each individual. Already, consumer products like the TweetPee™ allow a soiled diaper to alert the parents by Twitter or text message based on moisture alone – the potential for high-resolution tracking of additional microbiome variables is tremendous. The only question will be how to protect privacy, given the highly personal and personalized nature of the microbiome, and how to best place control of the microbiome in the empowered hands of each individual.

1.2.9 Acknowledgements

Chapter I, Introduction part 2, in full, is a reprint of previously published material: Knight R, Callewaert C, Marotz C, Hyde ER, Debelius JW, McDonald D, Sogin ML. *The microbiome and human biology*. Annual review of genomics and human genetics. 2017 Aug 31;18:65-86. I was one of the primary investigators and authors of this paper. The co-authors listed above supervised or provided support for the research and have given permission for the inclusion of the work in this dissertation.

1.2.10 References

1. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. 2007. The human microbiome project. Nature 449:804–810.

2. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Dore J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Meta HITC, Bork P, Ehrlich SD, Wang J. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464:59–65.

3. Human Microbiome Project C. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207–214.

4. Sender R, Fuchs S, Milo R. 2016. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. Cell 164:337–40.
5. Cambies. 1949. Parodontose et cure Thermale. Odontol Rev 447–452.

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

7. Stanier RY, Van Niel CB. 1941. The Main Outlines of Bacterial Classification. J Bacteriol 42:437–66.

8. Bergey DH, Brown CP, Etris S. 1939. Immunization Against Tetanus with Alum-Precipitated Tetanus Toxoid. Am J Public Health Nations Health 29:334–6.

9. Steele JA, Countway PD, Xia L, Vigil PD, Beman JM, Kim DY, Chow CE, Sachdeva R, Jones AC, Schwalbach MS, Rose JM, Hewson I, Patel A, Sun F, Caron DA, Fuhrman JA. 2011.

Marine bacterial, archaeal and protistan association networks reveal ecological linkages. ISME J 5:1414–1425.

10. Zuckerkandl E, Pauling L. 1965. Molecules as documents of evolutionary history. J Theor Biol 8:357–66.

11. Pace B, Campbell LL. 1971. Homology of ribosomal ribonucleic acid diverse bacterial species with Escherichia coli and Bacillus stearothermophilus. J Bacteriol 107:543–7.

12. Sogin SJ, Sogin ML, Woese CR. 1971. Phylogenetic measurement in procaryotes by primary structural characterization. J Mol Evol 1:173–84.

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

14. Woese CR, Kandler O, Wheelis ML. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 87:4576–9.

15. Brosius J, Palmer ML, Kennedy PJ, Noller HF. 1978. Complete nucleotide sequence of a 16S ribosomal RNA gene from Escherichia coli. Proc Natl Acad Sci U S A 75:4801–5.

16. Woese CR. 1987. Bacterial evolution. Microbiol Rev 51:221–71.

17. Maxam AM, Gilbert W. 1977. A new method for sequencing DNA. Proc Natl Acad Sci U S A 74:560–4.

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

19. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci U S A 82:6955–6959.

20. Elwood HJ, Olsen GJ, Sogin ML. 1985. The small-subunit ribosomal RNA gene sequences from the hypotrichous ciliates Oxytricha nova and Stylonychia pustulata. Mol Biol Evol 2:399–410.

21. Pace NR. 1997. A molecular view of microbial diversity and the biosphere. Science (80-) 276:734–740.

22. Medlin L, Elwood HJ, Stickel S, Sogin ML. 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Gene 71:491–9.

23. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE. Fluorescence detection in automated DNA sequence analysis. Nature 321:674–9.

24. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol 12:635–45.

25. Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere." Proc Natl Acad Sci U S A 103:12115–12120.

26. Bond PL, Hugenholtz P, Keller J, Blackall LL. 1995. Bacterial community structures of phosphate-removing and non-phosphate-removing activated sludges from sequencing batch reactors. Appl Environ Microbiol 61:1910–6.

27. Olsen GJ, Overbeek R, Larsen N, Marsh TL, McCaughey MJ, Maciukenas MA, Kuan WM, Macke TJ, Xing Y, Woese CR. 1992. The Ribosomal Database Project. Nucleic Acids Res 20 Suppl:2199–200.

28. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, Forster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, Konig A, Liss T, Lussmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, Schleifer KH. 2004. ARB: a software environment for sequence data. Nucleic Acids Res 32:1363–1371.

29. McCaig AE, Glover LA, Prosser JI. 1999. Molecular analysis of bacterial community structure and diversity in unimproved and improved upland grass pastures. Appl Environ Microbiol 65:1721–30.

30. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–10.

31. Schloss PD, Handelsman J. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Env Microbiol 71:1501–1506.

32. Baum BR. 1989. PHYLIP: Phylogeny Inference Package. Version 3.2 . Joel Felsenstein. Q Rev Biol 64:539–541.

33. Bray JR, Curtis JT. 1957. An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecol Monogr 27:325–349.

34. Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecol 26:32–46.

35. Kuczynski J, Liu Z, Lozupone C, McDonald D, Fierer N, Knight R. 2010. Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. Nat Methods 7:813–819.

36. Lozupone C, Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Env Microbiol 71:8228–8235.

37. Lozupone CA, Knight R. 2007. Global patterns in bacterial diversity. Proc Natl Acad Sci U S A 104:11436–11440.

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

39. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes X V, Godwin BC, He W, Helgesen S, Ho CH, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–80.

40. Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM. 2007. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol 8:R143.

41. Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, Head IM, Read LF, Sloan WT. 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. Nat Methods 6:639–641.

42. Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, Nielsen R, Willerslev E. 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS One 2:e197.

43. Huber JA, Mark Welch DB, Morrison HG, Huse SM, Neal PR, Butterfield DA, Sogin ML. 2007. Microbial population structures in the deep marine biosphere. Science 318:97–100.

44. Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Methods 5:235–237.

45. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–41.

46. 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. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336.

47. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI. 2012. Human gut microbiome viewed across age and geography. Nature 486:222–227.

48. Gonzalez A, Stombaugh J, Lauber CL, Fierer N, Knight R. 2012. SitePainter: a tool for exploring biogeographical patterns. Bioinformatics 28:436–8.

49. Bouslimani A, Porto C, Rath CM, Wang M, Guo Y, Gonzalez A, Berg-Lyon D, Ackermann G, Moeller Christensen GJ, Nakatsuji T, Zhang L, Borkowski AW, Meehan MJ, Dorrestein K, Gallo RL, Bandeira N, Knight R, Alexandrov T, Dorrestein PC. 2015. Molecular cartography of the human skin surface in 3D. Proc Natl Acad Sci U S A 112:E2120-9.

50. Liu Z, DeSantis TZ, Andersen GL, Knight R. 2008. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. Nucleic Acids Res 36:e120.

51. Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, Sogin ML. 2013. Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol Evol 4.

52. Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML. 2015. Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. Isme J 9:968–979.

53. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–3.

54. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Xu ZZ, Kightley EP, Thompson LR, Hyde ER, Peña AG, Knight R. Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems.

55. Olm MR, Brown CT, Brooks B, Firek B, Baker R, Burstein D, Soenjoyo K, Thomas BC, Morowitz M, Banfield J. 2017. Identical bacterial populations colonize premature infant gut, skin, and oral microbiomes and exhibit different in situ growth rates. Genome Res.

56. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, Arumugam M, Kultima JR, Prifti E, Nielsen T, Juncker AS, Manichanh C, Chen B, Zhang W, Levenez F, Wang J, Xu X, Xiao L, Liang S, Zhang D, Zhang Z, Chen W, Zhao H, Al-Aama JY, Edris S, Yang H, Wang J, Hansen T, Nielsen HB, Brunak S, Kristiansen K, Guarner F, Pedersen O, Doré J, Ehrlich SD, MetaHIT Consortium, Bork P, Wang J, MetaHIT Consortium. 2014. An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol 32:834–41.

57. No Title.

58. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. 2005. Diversity of the human intestinal microbial flora. Science (80-) 308:1635–1638.

59. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. 2009. Bacterial community variation in human body habitats across space and time. Science (80-) 326:1694–1697.

60. Human Microbiome Project C. 2012. A framework for human microbiome research. Nature 486:215–221.

61. Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R. 2010. Forensic identification using skin bacterial communities. Proc Natl Acad Sci U S A 107:6477–6481.

62. Franzosa EA, Huang K, Meadow JF, Gevers D, Lemon KP, Bohannan BJM, Huttenhower C. 2015. Identifying personal microbiomes using metagenomic codes. Proc Natl Acad Sci U S A 112:E2930-8.

63. Song SJ, Lauber C, Costello EK, Lozupone CA, Humphrey G, Berg-Lyons D, Caporaso JG, Knights D, Clemente JC, Nakielny S, Gordon JI, Fierer N, Knight R. 2013. Cohabiting family members share microbiota with one another and with their dogs. Elife 2:e00458.

64. Zozaya M, Ferris MJ, Siren JD, Lillis R, Myers L, Nsuami MJ, Eren AM, Brown J, Taylor CM, Martin DH. 2016. Bacterial communities in penile skin, male urethra, and vaginas of heterosexual couples with and without bacterial vaginosis. Microbiome 4:16.

65. 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. 2009. A core gut microbiome in obese and lean twins. Nature 457:480–484.

66. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R, Gordon JI. 2008. Evolution of mammals and their gut microbes. Science (80-) 320:1647–1651.

67. Muegge BD, Kuczynski J, Knights D, Clemente JC, Gonzalez A, Fontana L, Henrissat B, Knight R, Gordon JI. 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science (80-) 332:970–974.

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

69. Moeller AH, Caro-Quintero A, Mjungu D, Georgiev A V, Lonsdorf E V, Muller MN, Pusey AE, Peeters M, Hahn BH, Ochman H. 2016. Cospeciation of gut microbiota with hominids. Science 353:380–2.

70. Moeller AH, Peeters M, Ndjango J-B, Li Y, Hahn BH, Ochman H. 2013. Sympatric chimpanzees and gorillas harbor convergent gut microbial communities. Genome Res 23:1715–20.

71. Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. 2007. Short pyrosequencing reads suffice for accurate microbial community analysis. Nucleic Acids Res 35:e120.

72. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V, Henrissat B, Bain JR, Muehlbauer MJ, Ilkayeva O, Semenkovich CF, Funai K, Hayashi DK, Lyle BJ, Martini MC, Ursell LK, Clemente JC, Van Treuren W, Walters WA, Knight R, Newgard CB, Heath AC, Gordon JI. 2013. Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science (80-) 341:1241214.

73. Sampson TR, Debelius JW, Thron T, Janssen S, Shastri GG, Ilhan ZE, Challis C, Schretter CE, Rocha S, Gradinaru V, Chesselet M-F, Keshavarzian A, Shannon KM, Krajmalnik-Brown R, Wittung-Stafshede P, Knight R, Mazmanian SK. 2016. Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease. Cell 167:1469–1480.e12.

74. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van Treuren W, Knight R, Bell JT, Spector TD, Clark AG, Ley RE. 2014. Human genetics shape the gut microbiome. Cell 159:789–799.

75. Seedorf H, Griffin NW, Ridaura VK, Reyes A, Cheng J, Rey FE, Smith MI, Simon GM, Scheffrahn RH, Woebken D, Spormann AM, Van Treuren W, Ursell LK, Pirrung M, Robbins-Pianka A, Cantarel BL, Lombard V, Henrissat B, Knight R, Gordon JI. 2014. Bacteria from diverse habitats colonize and compete in the mouse gut. Cell 159:253–66.

76. Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM, Gibbons SM, Larsen P, Shogan BD, Weiss S, Metcalf JL, Ursell LK, Vazquez-Baeza Y, Van Treuren W, Hasan NA, Gibson MK, Colwell R, Dantas G, Knight R, Gilbert JA. 2014. Longitudinal analysis of microbial interaction between humans and the indoor environment. Science (80-) 345:1048–1052.

77. Metcalf JL, Xu ZZ, Weiss S, Lax S, Treuren W Van, Hyde ER, Song SJ, Amir A, Larsen P, Sangwan N, Haarmann D, Humphrey GC, Ackermann G, Thompson LR, Lauber C, Bibat A, Nicholas C, Gebert MJ, Petrosino JF, Reed SC, Gilbert JA, Lynne AM, Bucheli SR, Carter DO, Knight R. 2016. Microbial community assembly and metabolic function during mammalian corpse decomposition. Science (80-) 351:158–162.

78. Aagaard K, Riehle K, Ma J, Segata N, Mistretta T-A, Coarfa C, Raza S, Rosenbaum S, Van den Veyver I, Milosavljevic A, Gevers D, Huttenhower C, Petrosino J, Versalovic J. 2012. A metagenomic approach to characterization of the vaginal microbiome signature in pregnancy. PLoS One 7:e36466.

79. Romero R, Hassan SS, Gajer P, Tarca AL, Fadrosh DW, Nikita L, Galuppi M, Lamont RF, Chaemsaithong P, Miranda J, Chaiworapongsa T, Ravel J. 2014. The composition and stability

of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. Microbiome 2:4.

80. MacIntyre DA, Chandiramani M, Lee YS, Kindinger L, Smith A, Angelopoulos N, Lehne B, Arulkumaran S, Brown R, Teoh TG, Holmes E, Nicoholson JK, Marchesi JR, Bennett PR. 2015. The vaginal microbiome during pregnancy and the postpartum period in a European population. Sci Rep 5:8988.

81. 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. 2011. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A 108 Suppl:4680–4687.

82. Hillier SL, Nugent RP, Eschenbach DA, Krohn MA, Gibbs RS, Martin DH, Cotch MF, Edelman R, Pastorek JG, Rao A V. 1995. Association between bacterial vaginosis and preterm delivery of a low-birth-weight infant. The Vaginal Infections and Prematurity Study Group. N Engl J Med 333:1737–42.

83. Koren O, Goodrich JK, Cullender TC, Spor A, Laitinen K, Bäckhed HK, Gonzalez A, Werner JJ, Angenent LT, Knight R, Bäckhed F, Isolauri E, Salminen S, Ley RE. 2012. Host remodeling of the gut microbiome and metabolic changes during pregnancy. Cell 150:470–80.

84. Kliman HJ. 2014. Comment on “the placenta harbors a unique microbiome”. Sci Transl Med 6:254le4.

85. Aagaard KM. 2014. Author response to comment on “the placenta harbors a unique microbiome”. Sci Transl Med 6:254lr3.

86. Lauder AP, Roche AM, Sherrill-Mix S, Bailey A, Laughlin AL, Bittinger K, Leite R, Elovitz MA, Parry S, Bushman FD. 2016. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. Microbiome 4:29.

87. Zheng J, Xiao X, Zhang Q, Mao L, Yu M, Xu J. 2015. The Placental Microbiome Varies in Association with Low Birth Weight in Full-Term Neonates. Nutrients 7:6924–37.

88. Prince AL, Ma J, Kannan PS, Alvarez M, Gisslen T, Harris RA, Sweeney EL, Knox CL, Lambers DS, Jobe AH, Chougnet CA, Kallapur SG, Aagaard KM. 2016. The placental membrane microbiome is altered among subjects with spontaneous preterm birth with and without chorioamnionitis. Am J Obstet Gynecol 214:627.e1-627.e16.

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

90. Bäckhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, Li Y, Xia Y, Xie H, Zhong H, Khan MT, Zhang J, Li J, Xiao L, Al-Aama J, Zhang D, Lee YS, Kotowska D, Colding C, Tremaroli V, Yin Y, Bergman S, Xu X, Madsen L, Kristiansen K, Dahlgren J, Wang J, Jun

W. 2015. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe 17:690–703.

91. Diaz Heijtz R, Wang S, Anuar F, Qian Y, Björkholm B, Samuelsson A, Hibberd ML, Forssberg H, Pettersson S. 2011. Normal gut microbiota modulates brain development and behavior. Proc Natl Acad Sci U S A 108:3047–52.

92. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE. 2011. Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci U S A 108 Suppl:4578–4585.

93. Makrides M, Simmer K, Neumann M, Gibson R. 1995. Changes in the polyunsaturated fatty acids of breast milk from mothers of full-term infants over 30 wk of lactation. Am J Clin Nutr 61:1231–3.

94. Coppa G V, Gabrielli O, Pierani P, Catassi C, Carlucci A, Giorgi PL. 1993. Changes in carbohydrate composition in human milk over 4 months of lactation. Pediatrics 91:637–41.

95. Cabrera-Rubio R, Collado MC, Laitinen K, Salminen S, Isolauri E, Mira A. 2012. The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. Am J Clin Nutr 96:544–51.

96. Miller JB, Bull S, Miller J, McVeagh P. 1994. The oligosaccharide composition of human milk: temporal and individual variations in monosaccharide components. J Pediatr Gastroenterol Nutr 19:371–6.

97. Kwak M-J, Kwon S-K, Yoon J-K, Song JY, Seo J-G, Chung MJ, Kim JF. 2016. Evolutionary architecture of the infant-adapted group of Bifidobacterium species associated with the probiotic function. Syst Appl Microbiol 39:429–439.

98. Aminov RI. 2010. A brief history of the antibiotic era: lessons learned and challenges for the future. Front Microbiol 1:134.

99. Kutty N. 2011. Treating children without antibiotics in primary healthcare. Oman Med J 26:303–5.

100. Fouhy F, Guinane CM, Hussey S, Wall R, Ryan CA, Dempsey EM, Murphy B, Ross RP, Fitzgerald GF, Stanton C, Cotter PD. 2012. High-throughput sequencing reveals the incomplete, short-term recovery of infant gut microbiota following parenteral antibiotic treatment with ampicillin and gentamicin. Antimicrob Agents Chemother 56:5811–20.

101. Dethlefsen L, Relman DA. 2011. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci U S A 108 Suppl:4554–4561.

102. Nobel YR, Cox LM, Kirigin FF, Bokulich NA, Yamanishi S, Teitler I, Chung J, Sohn J, Barber CM, Goldfarb DS, Raju K, Abubucker S, Zhou Y, Ruiz VE, Li H, Mitreva M, Alekseyenko A V, Weinstock GM, Sodergren E, Blaser MJ. 2015. Metabolic and metagenomic outcomes from early-life pulsed antibiotic treatment. Nat Commun 6:7486.

103. Yassour M, Vatanen T, Siljander H, Hämäläinen A-M, Härkönen T, Ryhänen SJ, Franzosa EA, Vlamakis H, Huttenhower C, Gevers D, Lander ES, Knip M, DIABIMMUNE Study Group, Xavier RJ. 2016. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. Sci Transl Med 8:343ra81.

104. Azad MB, Bridgman SL, Becker AB, Kozyrskyj AL. 2014. Infant antibiotic exposure and the development of childhood overweight and central adiposity. Int J Obes (Lond) 38:1290–8.

105. Arrieta M-C, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S, Kuzeljevic B, Gold MJ, Britton HM, Lefebvre DL, Subbarao P, Mandhane P, Becker A, McNagny KM, Sears MR, Kollmann T, CHILD Study Investigators, Mohn WW, Turvey SE, Finlay BB. 2015. Early infancy microbial and metabolic alterations affect risk of childhood asthma. Sci Transl Med 7:307ra152.

106. Metsälä J, Lundqvist A, Virta LJ, Kaila M, Gissler M, Virtanen SM. 2013. Mother's and offspring's use of antibiotics and infant allergy to cow's milk. Epidemiology 24:303–9.

107. Kilkkinen A, Virtanen SM, Klaukka T, Kenward MG, Salkinoja-Salonen M, Gissler M, Kaila M, Reunanen A. 2006. Use of antimicrobials and risk of type 1 diabetes in a population-based mother-child cohort. Diabetologia 49:66–70.

108. Hviid A, Svanstrom H, Frisch M. 2011. Antibiotic use and inflammatory bowel diseases in childhood. Gut 60:49–54.

109. Debelius J, Song SJ, Vazquez-Baeza Y, Xu ZZ, Gonzalez A, Knight R. 2016. Tiny microbes, enormous impacts: what matters in gut microbiome studies? Genome Biol 17:217.

110. Jakobsson HE, Jernberg C, Andersson AF, Sjölund-Karlsson M, Jansson JK, Engstrand L. 2010. Short-term antibiotic treatment has differing long-term impacts on the human throat and gut microbiome. PLoS One 5:e9836.

111. Francino MP. 2015. Antibiotics and the Human Gut Microbiome: Dysbioses and Accumulation of Resistances. Front Microbiol 6:1543.

112. Girard C, Tromas N, Amyot M, Shapiro BJ. Gut Microbiome of the Canadian Arctic Inuit. mSphere 2.

113. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L, Li H, Bushman FD, Lewis JD. 2011. Linking long-term dietary patterns with gut microbial enterotypes. Science (80-) 334:105–108.

114. Subramanian S, Huq S, Yatsunenko T, Haque R, Mahfuz M, Alam MA, Benezra A, DeStefano J, Meier MF, Muegge BD, Barratt MJ, VanArendonk LG, Zhang Q, Province MA, Petri WA, Ahmed T, Gordon JI. 2014. Persistent gut microbiota immaturity in malnourished Bangladeshi children. Nature 510:417–21.

115. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling A V, Devlin AS, Varma Y, Fischbach MA, Biddinger SB, Dutton RJ, Turnbaugh PJ. 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559–563.

116. Smith MI, Yatsunenko T, Manary MJ, Trehan I, Mkakosya R, Cheng J, Kau AL, Rich SS, Concannon P, Mychaleckyj JC, Liu J, Houpt E, Li J V, Holmes E, Nicholson J, Knights D, Ursell LK, Knight R, Gordon JI. 2013. Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. Science (80-) 339:548–554.

117. Chassaing B, Koren O, Goodrich JK, Poole AC, Srinivasan S, Ley RE, Gewirtz AT. 2015. Dietary emulsifiers impact the mouse gut microbiota promoting colitis and metabolic syndrome. Nature 519:92–6.

118. Suez J, Korem T, Zeevi D, Zilberman-Schapira G, Thaiss CA, Maza O, Israeli D, Zmora N, Gilad S, Weinberger A, Kuperman Y, Harmelin A, Kolodkin-Gal I, Shapiro H, Halpern Z, Segal E, Elinav E. 2014. Artificial sweeteners induce glucose intolerance by altering the gut microbiota. Nature 514:181–6.

119. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, Ben-Yacov O, Lador D, Avnit-Sagi T, Lotan-Pompan M, Suez J, Mahdi JA, Matot E, Malka G, Kosower N, Rein M, Zilberman-Schapira G, Dohnalová L, Pevsner-Fischer M, Bikovsky R, Halpern Z, Elinav E, Segal E. 2015. Personalized Nutrition by Prediction of Glycemic Responses. Cell 163:1079–94.

120. Campbell SC, Wisniewski PJ. 2017. Exercise is a Novel Promoter of Intestinal Health and Microbial Diversity. Exerc Sport Sci Rev 45:41–47.

121. David LA, Materna AC, Friedman J, Campos-Baptista MI, Blackburn MC, Perrotta A, Erdman SE, Alm EJ. 2014. Host lifestyle affects human microbiota on daily timescales. Genome Biol 15:R89.

122. Lowry CA, Smith DG, Siebler PH, Schmidt D, Stamper CE, Hassell JE, Yamashita PS, Fox JH, Reber SO, Brenner LA, Hoisington AJ, Postolache TT, Kinney KA, Marciani D, Hernandez M, Hemmings SMJ, Malan-Muller S, Wright KP, Knight R, Raison CL, Rook GAW. 2016. The Microbiota, Immunoregulation, and Mental Health: Implications for Public Health. Curr Environ Heal reports 3:270–86.

123. Benedict C, Vogel H, Jonas W, Woting A, Blaut M, Schürmann A, Cedernaes J. 2016. Gut microbiota and glucometabolic alterations in response to recurrent partial sleep deprivation in normal-weight young individuals. Mol Metab 5:1175–1186.

124. McNulty NP, Yatsunenko T, Hsiao A, Faith JJ, Muegge BD, Goodman AL, Henrissat B, Oozeer R, Cools-Portier S, Gobert G, Chervaux C, Knights D, Lozupone CA, Knight R, Duncan AE, Bain JR, Muehlbauer MJ, Newgard CB, Heath AC, Gordon JI. 2011. The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. Sci Transl Med 3:106ra106.

125. Alekseyenko A V, Perez-Perez GI, De Souza A, Strober B, Gao Z, Bihan M, Li K, Methé BA, Blaser MJ. 2013. Community differentiation of the cutaneous microbiota in psoriasis. Microbiome 1:31.

126. Kobayashi T, Glatz M, Horiuchi K, Kawasaki H, Akiyama H, Kaplan DH, Kong HH, Amagai M, Nagao K. 2015. Dysbiosis and Staphylococcus aureus Colonization Drives Inflammation in Atopic Dermatitis. Immunity 42:756–66.

127. Niehues H, Schalkwijk J, van Vlijmen-Willems IMJJ, Rodijk-Olthuis D, van Rossum MM, Wladykowski E, Brandner JM, van den Bogaard EHJ, Zeeuwen PLJM. 2016. Epidermal equivalents of filaggrin null keratinocytes do not show impaired skin barrier function. J Allergy Clin Immunol.

128. Hernandez CJ, Guss JD, Luna M, Goldring SR. 2016. Links Between the Microbiome and Bone. J Bone Miner Res 31:1638–46.

129. Gao R, Gao Z, Huang L, Qin H. 2017. Gut microbiota and colorectal cancer. Eur J Clin Microbiol Infect Dis.

130. Simrén M, Barbara G, Flint HJ, Spiegel BMR, Spiller RC, Vanner S, Verdu EF, Whorwell PJ, Zoetendal EG, Rome Foundation Committee. 2013. Intestinal microbiota in functional bowel disorders: a Rome foundation report. Gut 62:159–76.

131. Zheng P, Zeng B, Zhou C, Liu M, Fang Z, Xu X, Zeng L, Chen J, Fan S, Du X, Zhang X, Yang D, Yang Y, Meng H, Li W, Melgiri ND, Licinio J, Wei H, Xie P. 2016. Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism. Mol Psychiatry 21:786–96.

132. Thaiss CA, Zeevi D, Levy M, Zilberman-Schapira G, Suez J, Tengeler AC, Abramson L, Katz MN, Korem T, Zmora N, Kuperman Y, Biton I, Gilad S, Harmelin A, Shapiro H, Halpern Z, Segal E, Elinav E. 2014. Transkingdom control of microbiota diurnal oscillations promotes metabolic homeostasis. Cell 159:514–29.

133. Charbonneau MR, O'Donnell D, Blanton L V, Totten SM, Davis JCC, Barratt MJ, Cheng J, Guruge J, Talcott M, Bain JR, Muehlbauer MJ, Ilkayeva O, Wu C, Struckmeyer T, Barile D, Mangani C, Jorgensen J, Fan Y, Maleta K, Dewey KG, Ashorn P, Newgard CB, Lebrilla C, Mills DA, Gordon JI. 2016. Sialylated Milk Oligosaccharides Promote Microbiota-Dependent Growth in Models of Infant Undernutrition. Cell 164:859–71.

134. Round JL, Mazmanian SK. 2009. The gut microbiota shapes intestinal immune responses during health and disease. Nat Rev Immunol 9:313–23.

135. Mazmanian SK, Liu CH, Tzianabos AO, Kasper DL. 2005. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. Cell 122:107–18.

136. Atarashi K, Tanoue T, Shima T, Imaoka A, Kuwahara T, Momose Y, Cheng G, Yamasaki S, Saito T, Ohba Y, Taniguchi T, Takeda K, Hori S, Ivanov II, Umesaki Y, Itoh K, Honda K. 2011. Induction of colonic regulatory T cells by indigenous Clostridium species. Science 331:337–41.

137. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward D V, Reyes JA, Shah SA, LeLeiko N, Snapper SB, Bousvaros A, Korzenik J, Sands BE, Xavier RJ, Huttenhower C. 2012. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol 13:R79.

138. Gevers D, Kugathasan S, Denson LA, Vazquez-Baeza Y, Van Treuren W, Ren B, Schwager E, Knights D, Song SJ, Yassour M, Morgan XC, Kostic AD, Luo C, Gonzalez A, McDonald D, Haberman Y, Walters T, Baker S, Rosh J, Stephens M, Heyman M, Markowitz J, Baldassano R, Griffiths A, Sylvester F, Mack D, Kim S, Crandall W, Hyams J, Huttenhower C, Knight R, Xavier RJ. 2014. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe 15:382–392.

139. Inan MS, Rasoulpour RJ, Yin L, Hubbard AK, Rosenberg DW, Giardina C. 2000. The luminal short-chain fatty acid butyrate modulates NF-kappaB activity in a human colonic epithelial cell line. Gastroenterology 118:724–34.

140. Quinn RA, Lim YW, Maughan H, Conrad D, Rohwer F, Whiteson KL. 2014. Biogeochemical forces shape the composition and physiology of polymicrobial communities in the cystic fibrosis lung. MBio 5:e00956-13.

141. Knights D, Lassen KG, Xavier RJ. 2013. Advances in inflammatory bowel disease pathogenesis: linking host genetics and the microbiome. Gut 62:1505–10.

142. Shulman JM, De Jager PL, Feany MB. 2011. Parkinson's disease: genetics and pathogenesis. Annu Rev Pathol 6:193–222.

143. Harris RA, Shah R, Hollister EB, Tronstad RR, Hovdenak N, Szigeti R, Versalovic J, Kellermayer R. 2016. Colonic Mucosal Epigenome and Microbiome Development in Children and Adolescents. J Immunol Res 2016:9170162.

144. Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S, Prifti E, Vieira-Silva S, Gudmundsdottir V, Krogh Pedersen H, Arumugam M, Kristiansen K, Voigt AY, Vestergaard H, Hercog R, Igor Costea P, Kultima JR, Li J, Jørgensen T, Levenez F, Dore J, MetaHIT consortium, Nielsen HB, Brunak S, Raes J, Hansen T, Wang J, Ehrlich SD, Bork P,

Pedersen O. 2015. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature 528:262–6.

145. Zhang X, Zhang D, Jia H, Feng Q, Wang D, Liang D, Wu X, Li J, Tang L, Li Y, Lan Z, Chen B, Li Y, Zhong H, Xie H, Jie Z, Chen W, Tang S, Xu X, Wang X, Cai X, Liu S, Xia Y, Li J, Qiao X, Al-Aama JY, Chen H, Wang L, Wu Q-J, Zhang F, Zheng W, Li Y, Zhang M, Luo G, Xue W, Xiao L, Li J, Chen W, Xu X, Yin Y, Yang H, Wang J, Kristiansen K, Liu L, Li T, Huang Q, Li Y, Wang J. 2015. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. Nat Med 21:895–905.

146. Maurice CF, Haiser HJ, Turnbaugh PJ. 2013. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. Cell 152:39–50.

147. Clayton TA, Baker D, Lindon JC, Everett JR, Nicholson JK. 2009. Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. Proc Natl Acad Sci U S A 106:14728–14733.

148. Weingarden A, González A, Vázquez-Baeza Y, Weiss S, Humphry G, Berg-Lyons D, Knights D, Unno T, Bobr A, Kang J, Khoruts A, Knight R, Sadowsky MJ. 2015. Dynamic changes in short- and long-term bacterial composition following fecal microbiota transplantation for recurrent Clostridium difficile infection. Microbiome 3:10.

149. Bartlett JG, Gerding DN. 2008. Clinical recognition and diagnosis of Clostridium difficile infection. Clin Infect Dis 46 Suppl 1:S12-8.

150. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444:1027–1031.

151. Kostic AD, Xavier RJ, Gevers D. 2014. The microbiome in inflammatory bowel disease: current status and the future ahead. Gastroenterology 146:1489–99.

152. Mimouna S, Goncalves D, Barnich N, Darfeuille-Michaud A, Hofman P, Vouret-Craviari V. 2011. Crohn disease-associated Escherichia coli promote gastrointestinal inflammatory disorders by activation of HIF-dependent responses. Gut Microbes 2:335–346.

153. Fitz-Gibbon S, Tomida S, Chiu B-H, Nguyen L, Du C, Liu M, Elashoff D, Erfe MC, Loncaric A, Kim J, Modlin RL, Miller JF, Sodergren E, Craft N, Weinstock GM, Li H. 2013. Propionibacterium acnes strain populations in the human skin microbiome associated with acne. J Invest Dermatol 133:2152–60.

154. Bessa LJ, Fazii P, Di Giulio M, Cellini L. 2015. Bacterial isolates from infected wounds and their antibiotic susceptibility pattern: some remarks about wound infection. Int Wound J 12:47–52.

155. Ding HT, Taur Y, Walkup JT. 2016. Gut Microbiota and Autism: Key Concepts and Findings. J Autism Dev Disord.

156. Hsiao EY, McBride SW, Hsien S, Sharon G, Hyde ER, McCue T, Codelli JA, Chow J, Reisman SE, Petrosino JF, Patterson PH, Mazmanian SK. 2013. Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. Cell 155:1451–1463.

157. Stein MM, Hrusch CL, Gozdz J, Igartua C, Pivniouk V, Murray SE, Ledford JG, Marques dos Santos M, Anderson RL, Metwali N, Neilson JW, Maier RM, Gilbert JA, Holbreich M, Thorne PS, Martinez FD, von Mutius E, Vercelli D, Ober C, Sperling AI. 2016. Innate Immunity and Asthma Risk in Amish and Hutterite Farm Children. N Engl J Med 375:411–21.

158. Kleiman SC, Watson HJ, Bulik-Sullivan EC, Huh EY, Tarantino LM, Bulik CM, Carroll IM. The Intestinal Microbiota in Acute Anorexia Nervosa and During Renourishment: Relationship to Depression, Anxiety, and Eating Disorder Psychopathology. Psychosom Med 77:969–81.

159. Jonsson AL, Bäckhed F. 2016. Role of gut microbiota in atherosclerosis. Nat Rev Cardiol.

160. Lynch S V, Bruce KD. 2013. The cystic fibrosis airway microbiome. Cold Spring Harb Perspect Med 3:a009738.

161. Oliver A, Cantón R, Campo P, Baquero F, Blázquez J. 2000. High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection. Science 288:1251–4.

162. Fredricks DN, Fiedler TL, Marrazzo JM. 2005. Molecular identification of bacteria associated with bacterial vaginosis. N Engl J Med 353:1899–1911.

163. Peterson SN, Snesrud E, Liu J, Ong AC, Kilian M, Schork NJ, Bretz W. 2013. The dental plaque microbiome in health and disease. PLoS One 8:e58487.

164. Yang F, Zeng X, Ning K, Liu KL, Lo CC, Wang W, Chen J, Wang D, Huang R, Chang X, Chain PS, Xie G, Ling J, Xu J. 2012. Saliva microbiomes distinguish caries-active from healthy human populations. ISME J 6:1–10.

165. Kostic AD, Gevers D, Siljander H, Vatanen T, Hyötyläinen T, Hämäläinen A-M, Peet A, Tillmann V, Pöhö P, Mattila I, Lähdesmäki H, Franzosa EA, Vaarala O, de Goffau M, Harmsen H, Ilonen J, Virtanen SM, Clish CB, Orešič M, Huttenhower C, Knip M, DIABIMMUNE Study Group, Xavier RJ. 2015. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. Cell Host Microbe 17:260–73.

166. Paun A, Yau C, Danska JS. 2017. The Influence of the Microbiome on Type 1 Diabetes. J Immunol 198:590–595.

167. Wang X, Xu X, Xia Y. 2016. Further analysis reveals new gut microbiome markers of type 2 diabetes mellitus. Antonie Van Leeuwenhoek.

168. Kane A V, Dinh DM, Ward HD. 2015. Childhood malnutrition and the intestinal microbiome. Pediatr Res 77:256–62.

169. Tang WHW, Hazen SL. 2014. The contributory role of gut microbiota in cardiovascular disease. J Clin Invest 124:4204–11.

170. Jangi S, Gandhi R, Cox LM, Li N, von Glehn F, Yan R, Patel B, Mazzola MA, Liu S, Glanz BL, Cook S, Tankou S, Stuart F, Melo K, Nejad P, Smith K, Topçuolu BD, Holden J, Kivisäkk P, Chitnis T, De Jager PL, Quintana FJ, Gerber GK, Bry L, Weiner HL. 2016. Alterations of the human gut microbiome in multiple sclerosis. Nat Commun 7:12015.

171. Hartmann P, Seebauer CT, Schnabl B. 2015. Alcoholic liver disease: the gut microbiome and liver cross talk. Alcohol Clin Exp Res 39:763–75.

172. Kiraly DD, Walker DM, Calipari ES, Labonte B, Issler O, Pena CJ, Ribeiro EA, Russo SJ, Nestler EJ. 2016. Alterations of the Host Microbiome Affect Behavioral Responses to Cocaine. Sci

1.3

Are microbiome studies ready for hypothesis-driven research?

Hypothesis-driven research has led to many scientific advances, but hypotheses cannot be tested in isolation: rather, they require a framework of aggregated scientific knowledge to allow questions to be posed meaningfully. This framework is largely still lacking in microbiome studies, and the only way to create it is by discovery- and tool-driven research projects. Here we describe the value of several such projects from our own laboratories, including the American Gut Project, the Earth Microbiome Project (which is an umbrella project integrating many smaller hypothesis-driven projects), and the knowledgebase-driven tools GNPS and Qiita. We argue that an investment of community resources in these infrastructure tasks, and in the controls and standards that underpin them, will greatly enhance the investment of hypothesis-driven research programs.

1.3.1 Introduction

Microbiome research is making dramatic progress, with thousands of papers now published each year linking specific microbes and/or host-microbe co-metabolites to specific diseases, physiological properties, or environmental parameters. Much of this research is performed in a traditional, hypothesis-driven way, or at least presented as a rational reconstruction that fits this model, much as Darwin re-wrote much of his discovery-driven work as hypothesis driven to increase its respectability under the influence of contemporary philosophers of science such as William Whewell [1]. However, it should be noted that hypothesis-driven science was not always so respectable -- Isaac Newton famously wrote "*Hypotheses non fingo*", or "I feign no hypotheses", in an essay appended to the second edition of the *Principia* [2] -- so the tradition of modifying how science is framed in order to meet respectability criteria dates back at least 300

years. In any case, what can be framed as a singular hypothesis suffers important limitations based on what we can measure, and what we already know.

Ten years ago Chris Anderson, editor of Wired magazine, set off an international debate with his article "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete" [3]. The idea was that with enough data, hypotheses will emerge from the data ("Let the data speak for itself") has become widely discussed in the rapidly growing data science profession. A thoughtful review of this topic was written in EMBO Reports in 2015-"Could Big Data be the end of theory in science? A few remarks on the epistemology of data-driven science" [4]. As the author points out:

"Francis Bacon, the "father of the scientific method" himself, in his Novum Organum (1620), argued that scientific knowledge should not be based on preconceived notions but on experimental data. Deductive reasoning, he argued, is eventually limited because setting a premise in advance of an experiment would constrain the reasoning so as to match that premise. Instead, he advocated a bottom-up approach: In contrast to deductive reasoning, which has dominated science since Aristotle, inductive reasoning should be based on facts to generalize their meaning, drawing inferences from observations and data."

One constant in microbiome research has been that most factors that we would intuitively suspect to drive differences in the microbiome are of minor importance. For example, although long-term dietary changes have a major effect on the microbiome, short-term changes don't[5,6]. Similarly, sex has a very limited impact on microbiomes across the human body [7,8] and has a much weaker effect than many other variables such as age (even within adults) and the time of year the sample was collected [9,10]. Perhaps more surprisingly, factors such as temperature and pH have a much smaller impact on environmental microbiomes than salinity [11,12], and even the

saline vs. non-saline difference is much smaller than the host-associated vs free-living difference [12,13]. Samples from different sites of the same person's body can be more different from one another in terms of their overall microbial communities than radically different free-living microbial communities, such as soils versus oceans [12]. Differences of this magnitude can also occur within the gut of a single person, with sufficiently large perturbation [DOI: 10.1101/277970].

As a consequence, it is easy to incorrectly frame hypotheses, especially when supervised ordination and classification techniques are used in experiments with many confounding variables. For example, suppose that for mouse experiments we don't know that cage effects are important in the microbiome [14], then we profile the microbiomes in each of two cages of each of two different genotypes of mice. Our results are likely to be driven by which pair of cages happens to resemble each other more closely. If the variable of cage is not measured, or not tested in an unsupervised model, we might never know that our results are driven by this important confounding variable! There may be many more important confounding variables that we are not yet aware of, so longitudinal studies with meticulous metadata annotation will be crucial for defining which environmental factors matter. This is especially important in the context of clinical samples, where single data points are often collected and obtaining contextual information in retrospect is exceedingly difficult [15].

Similarly, a frequent practice is to discard unannotated microbes or unannotated molecules, focusing on the subset of microbes or molecules that can be matched to an existing database. Because databases of both microbes and molecules are heavily biased (microbes, by studies of known pathogens which come from only a small number of taxonomic groups, and molecules, by commercially available compounds), the entities that actually best discriminate among classes of

samples may be lost in the analysis: often, only 60% of sequences and 2% of molecular features from an untargeted metabolomics experiment can be annotated by existing references [16,17]. However, a rational reconstruction of why the annotatable microbes or molecules are plausible can always be developed by creative scientists looking to respond to their reviewers' criticism that their manuscript is "too descriptive".

1.3.2 The Need for Maps

An important metaphor in science and information visualization is the idea of the map, whether of real spaces or of abstract spaces. Indeed, as data volumes increase, it is frequent that the field moves from tests of hypotheses among sites, to tests of these hypotheses with replicates at each site, to spatially or temporally explicit sampling, to detailed spatial maps. This progression has already occurred in 16S rRNA amplicon-based microbiome studies over the past decade [12,18], and has increasingly been taking place in mass spectrometry-based metabolome studies over the past four years [19-24].

The value of spatial maps is so self-evident that the results are often cursed by obviousness. For example, the finding that metabolomes cluster by individual, as revealed by principal coordinates analysis (PCoA), is interesting (Fig. 1A). However, the finding that a given molecule such as lauryl sulphate (*m/z* 355.219) covers one individual, but is absent from the other individual is obvious (Fig. 1B), especially when you know that individual subject A uses a stereotypically gendered product such as Nivea for Men, which is the source of the molecule [20]. How such personal lifestyle (often hygiene, health or beautification related) influences the microbiome is not known; it is also not known how even some basic parameters such as, skin temperature, skin pH, amount of sebum influences the microbial communities on the skin. Similarly, the finding that

samples from four individuals differ to a statistically significant extent in their levels of specific purines and that within an individual, such molecules are also non-randomly distributed, might well be an intriguing finding prompting more investigation. However, a spatial map with dense sampling of the same individuals (Fig. 1C) makes it obvious that the molecule is something that is touched and consumed, and sometimes spilled, allowing one to guess that it is probably caffeine and that one person likely spends time in the ocean based on the distribution of *Synechococcus* spp. (Fig. 1D) (both of which are in fact the case) [22].

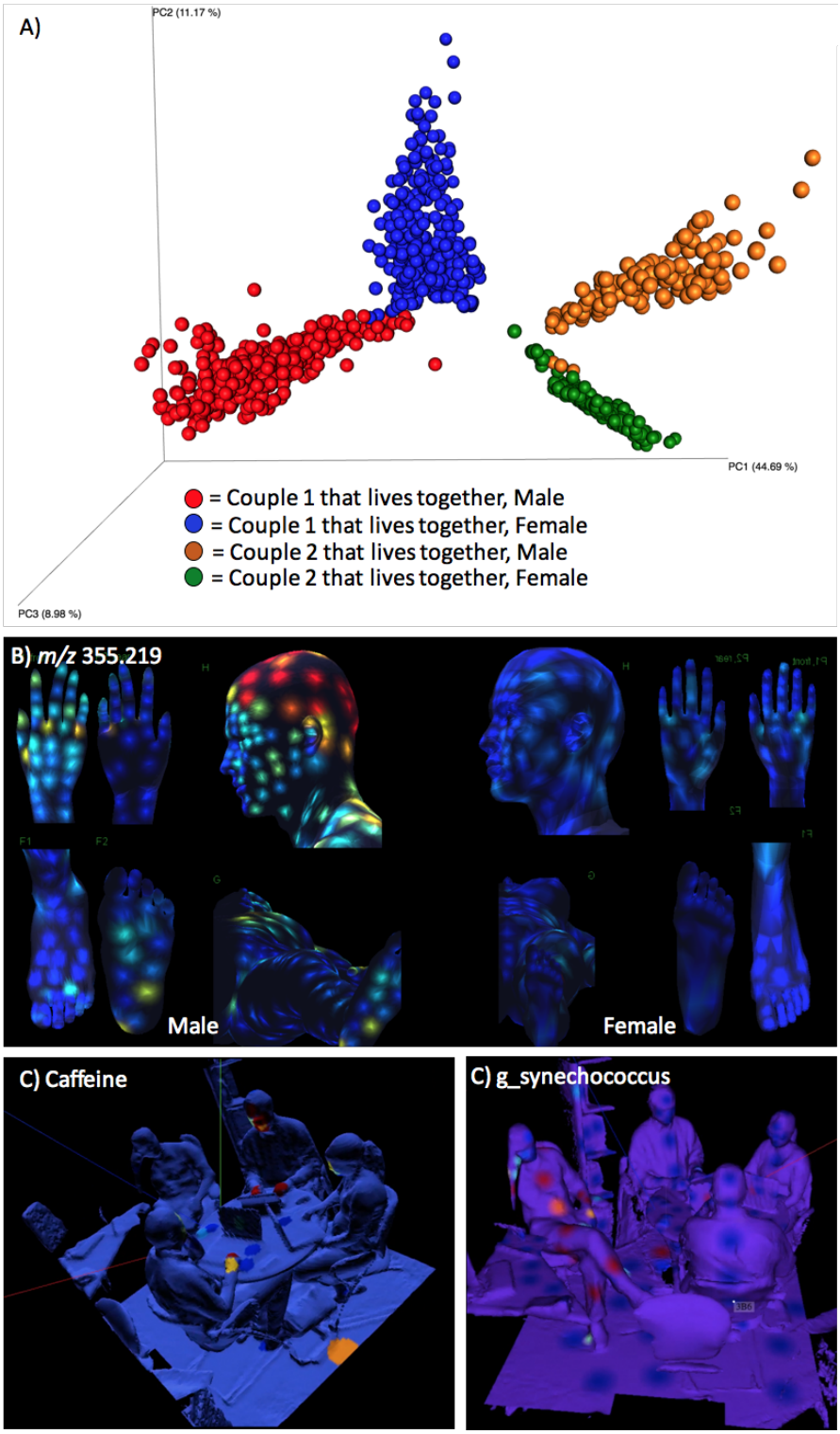

Figure 1.3.1. *Spatial analysis based on metabolomics of skin samples and a human habitat.* **A)** Principal coordinates analysis (Hellinger distance) of metabolomics data of skin swabs obtained from several hundreds locations on the human body of four volunteers. **B)** The detection of lauryl sulfate (*m/z* 355.219) from the shampoo Nivea for Men on a male volunteer. **C)** The distribution of caffeine (m/z 195.088) on four individuals and office environment. **D)** The distribution of *Synechococcus* spp. on within that same office environment.

However, the fact remains that for most microbes and for most molecules, we have no idea where they are in and on the human body, in natural environments, or in human-impacted environments including built environments. Just as John Snow's map of cholera instantly led to the hypothesis that this disease was water-borne and stemmed from the Broad Street pump, reinforced by the map's revelation that the block that drank alcohol had no incidence of disease [25]. The power of maps in shown by the history that this visual display of disease incidences by street became the foundation for the science and practice of epidemiology. In an analogous manner, systematically collected maps of microbes and of molecules across different spatial scales will dramatically improve our ability to make useful inferences from this data. Integration of these maps with other data layers ranging from air pollution to food deserts and neighborhood walkability, together with zoomable user interfaces (consider the utility of Google Maps versus earlier fixed-scale maps on DVD), will fundamentally transform the types of questions that can be asked of microbiome and metabolomics data.

The value of abstract maps, whether ordinations such as principal coordinates analysis (PCoA), non-metric multidimensional scaling (NMDS), t-distributed stochastic neighbor embedding (t-SNE), network diagrams obtained from object similarity (sequence or spectrum), or from co-occurrence across samples, is also considerable. In particular, when the right data frame and metrics are chosen, the key result is often immediately obvious. Consider, for example, the starting and ending time point of a fecal transplantation series [26] (Fig. 2A), where it's obvious that the clusters are statistically significantly different, but it is not obvious what direction this difference is in or what it means. However, when we perform a meta-analysis and put these samples in the context of the Human Microbiome Project data [8], one of the most important abstract maps in human microbiome science, we see immediately that the difference between start

and endpoint is much greater than the difference between healthy and diseased samples, and when we add the intermediate timepoints we see that the transition occurs very rapidly. These types of examples prompt similar data collection and visualization techniques in metabolomics, in order to understand how we can identify a desirable metabolomic state (for example, by comparing healthy and sick individuals), and guide an undesirable state into a desirable one by optimizing the trajectory towards the desired state in a series of perturbations. Only the existence of a map can allow rational hypotheses about what to try, especially in the context of n=1 studies or in cases where response heterogeneity among individuals is extreme.

1.3.3 The Need for Tools

We have seen, quite literally, the value of maps. But how do we build them? The key to acquiring high-resolution data, whether spatially or temporally resolved, or dense enough in an abstract space, is to make sampling fast, cheap, and sufficiently precise. Unfortunately, the trade-offs among these approaches are typically not well understood.

In DNA sequencing, a common question is whether, given a fixed sequencing budget, it is better to have more sequences per sample, or more samples. In general the answer to this question depends on the hypothesis to be tested. But, as noted above, all too frequently the "hypothesis" is retrofit to an arbitrarily collected dataset. What guidelines can be provided for aspiring microbial cartographers?

In our experience, for amplicon sequencing, the value of having more samples has always outweighed the value of having more sequences per sample, down to surprisingly low thresholds. For example, Fig. 3 shows the Earth Microbiome Project dataset [12] sampled at 500,000 sequences per sample, 1000 sequences per sample, and just 200 sequences per sample. The overall

pattern, e.g. the host/non-host split and the saline/non-saline split, are much clearer with more samples than with more precision about the location of each sample in PCoA space. Multinomial sampling considerations make it immediately clear why this is true: with 100 sequences per sample, the standard error in inferring the proportion of a taxon at 5% frequency is ~sqrt(100*.95*.05) or 2.18, or about 50% error in proportion; the standard error at a taxon at 1% frequency is about ±1, or about 100% error. Consequently, even low-abundance taxa are sampled with enough accuracy to place a sample in the context of an overall map with surprisingly few sequences. Logically, this must be true, or all ordination diagrams in microbial ecology before the advent of next-generation sequencing would have been useless, yet many revealed biologically interesting principles. The goal for better amplicon maps should therefore be to process vast numbers of additional samples inexpensively, exploiting the power of modern sequencers.

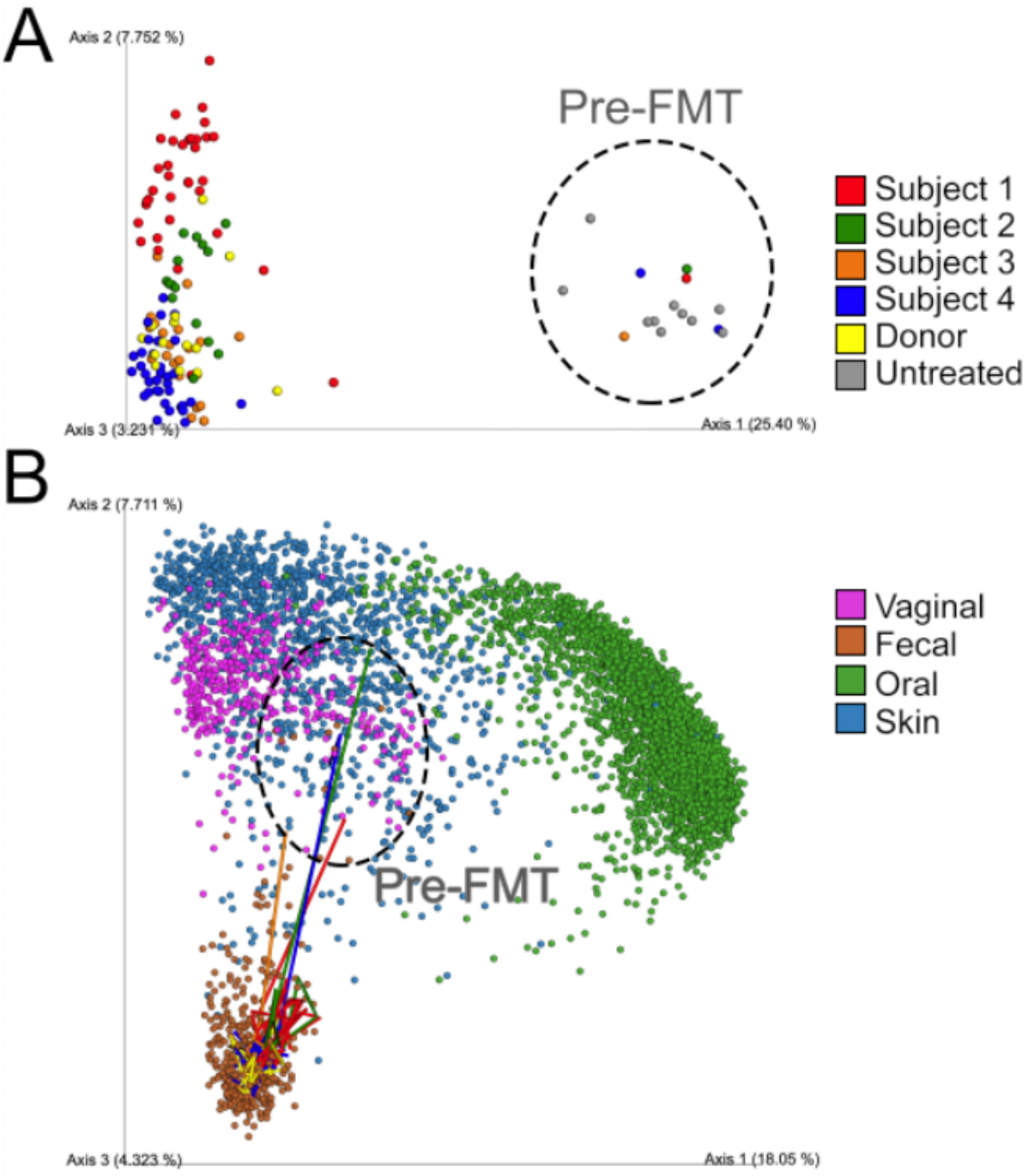

Figure 1.3.2. *Untangling the meaning of complex microbial interactions through meta-analyses.* (A) Principal coordinates analysis (unweighted UniFrac) of *Clostridium difficile* Infection subjects, before and after a fecal transplant, along with the fecal donor and 10 untreated subjects [26]. (B) Principal coordinates analysis (unweighted UniFrac) of the Human Microbiome Project (HMP) [8] combined with the data in panel A, the longitudinal samples for subjects 1-4 are connected as lines displaying the temporal variability and the shift from a disjointed untreated state of the patients vs. the healthy frame of the HMP.

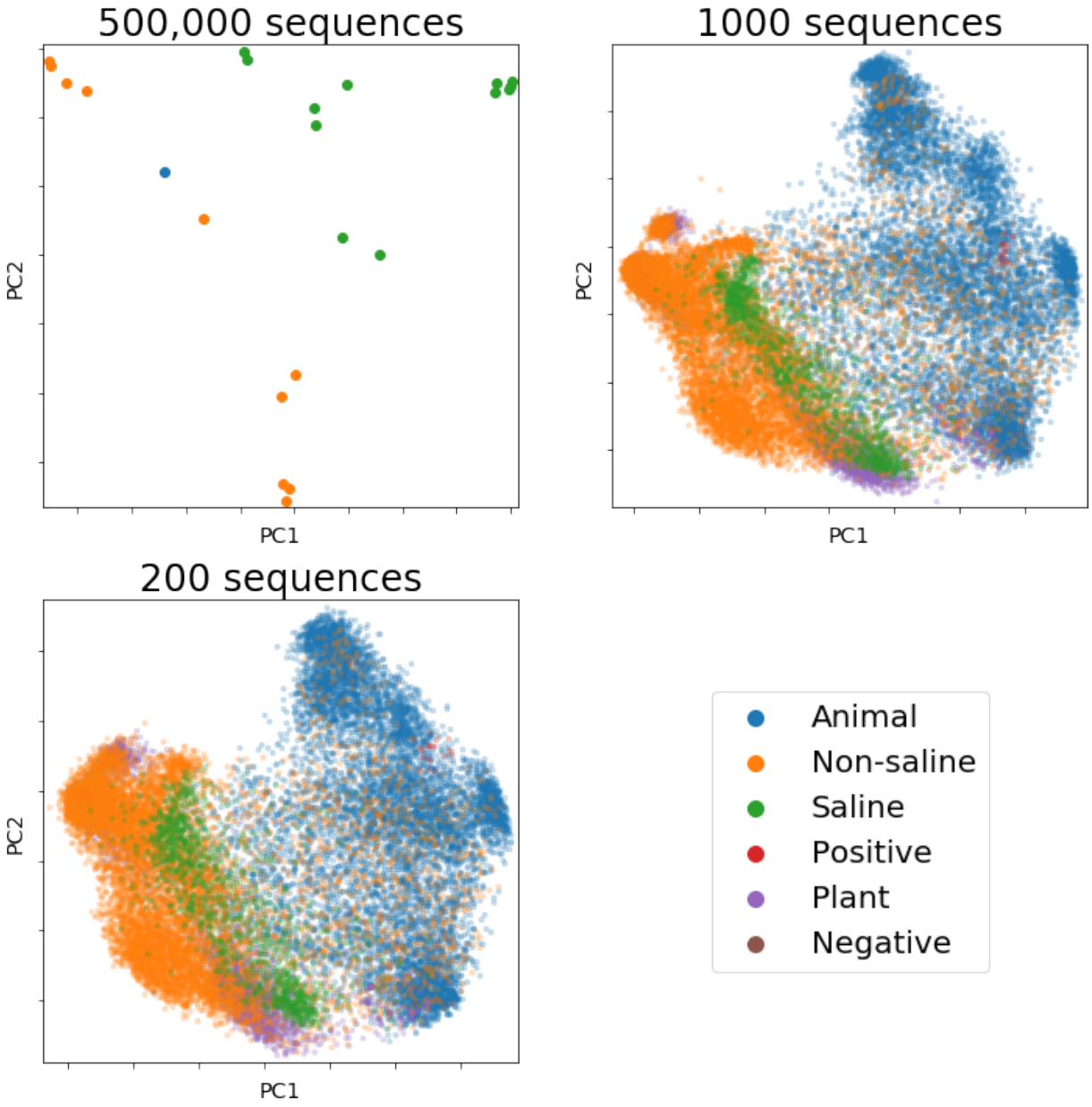

Figure 1.3.3. *Broader sampling improves maps of the microbial world, even with low resolution. All panels show principal coordinates analysis of unweighted UniFrac distances between samples.* (A) Samples rarefied to 500,000 sequences, showing only those exceeding this threshold sampling depth. (B) Samples rarefied to 1000 sequences. (C) Samples rarefied to 200 sequences. Even with few observations per sample, the overall relationships among sample types are preserved; in contrast, the overall pattern is lost with too few samples no matter how exquisitely characterized.

Shotgun metagenomics, however, poses a different challenge, because typically only a small fraction of the sequences can be confidently associated with known taxonomy or function. Further, the goals are often different because of the value of genome assembly in identifying biosynthetic pathways, allowing taxonomic resolution at the species or strain level, and generating high-resolution single nucleotide polymorphism (SNP) profiles to characterize novel strains and to confirm functional variants [27]. As a result, although the same sampling principles as for amplicon data apply if the goal is to provide a high-level taxonomic profile, far more sequences must be collected to have the same level of confidence in the result. Consequently, the most important areas for tool development in shotgun metagenomics are either several additional orders of magnitude drop in sequencing cost, reference databases that are more comprehensive and unbiased, and algorithms that are more efficient and accurate in read alignment, genome assembly and separation. In particular, methods that can identify genetic variation from lower-coverage data, and methods for estimating features of interest from less data or with efficient target capture, are of significant necessity. Another important consideration in shotgun metagenomics requires host DNA depletion, both experimentally and computationally, because total DNA extracts from biological specimens can be dominated by host DNA that is not picked up by standard PCR primers for bacterial/archaeal amplicon sequencing [28].

The challenges in metabolomics are somewhat different [29]. Sequencing has reduced in cost by nine orders of magnitude per data volume. In comparison, mass spectrometry, during the same time, has only reduced in cost of data volume collection by two orders of magnitude [29]. However the main limitation is the enormous diversity in chemistry. Unlike just four bases one has to “identify” to enable sequencing, there are hundreds to thousands of molecules that need to be identified from a list of millions, if the molecule is known to exist at all. The chemical diversity

also impacts the choice of extraction solvents during sample preparation, type of separation methods, type of instruments used and data analysis approaches. Further, because the multiplexing strategies that are successful in both amplicon- and shotgun-based sequencing approaches are not available in mass spectrometry, instrument time is directly proportional to the number of samples. Consequently, although it is easy to slip a few more samples into a mass spec run, instrument time is limiting for large-scale projects. As was the case with sequencing a decade ago, the vast majority of molecular features that are found in a sample are currently unidentified, and many are likely technical artifacts of various steps in the process, e.g. adducts formed in the gas phase, solvent artifacts [30] and multimers of the same compound [29]. Better methods and incentives for aggregating community knowledge [17] (e.g. retention of knowledge of the large number of manual annotations performed by the community) and for automatically assigning unknown mass peaks and fragmentation spectra to molecules and have an estimation of error rates [31], as opposed to heuristics subject to personal interpretation rules [32], are urgently needed. Global Natural Products Social molecular networking (GNPS) [17] offers alternative solutions for computational mass spectrometry infrastructure. Spectral datasets can be publicly deposited with a unique identifier and transformed to "living data" as they will be continuously searched against reference libraries to update users on new identifications. Furthermore, annotations can also be made by the scientific community within GNPS and propagated to all other data sets in the public domain with notifying subscribers on new annotations. This living data concept is crucial way to ensure that collected metabolomics data can still be useful over time. Other examples include automated species metabolome references [33] and the Molecular Explorer [17] for cross-searching annotated MS/MS spectra between datasets. Connections between several datasets, within the same knowledge base or between different spectral repositories such as Metabolights [34] and

Metabolomics Workbench [35], can be made to highlight annotated compounds found in several data sets Such analysis is a trivial task in sequencing but still novel in mass spectrometry.

Integration of taxonomic, genomic and metabolomic data remains an important unsolved challenge. Although genome mining is successful for identifying the sources of individual natural products, matching up the overall taxonomic or functional profile to a molecular profile remains challenging because of procedural and analytical differences in data acquisition. In particular, the likelihood of time lags in chemical production or in genomic response to environmental changes, which may appear on different timescales, make integrated analysis of snapshot data extremely challenging [36]. In cases where microbial and molecular composition is driven by a dominant effect (e.g. a dataset composed of soil and fecal samples), the molecular and metagenomic datasets will appear concordant by Procrustes analysis [37], which measures the fit of one ordination space to another. It is likely that an integrated systems biology approach that maps all data layers onto common pathways will be needed. This task is complicated at present not only because most genes, pathways, and molecules are unknown (especially those involving biotransformations of environmental or food inputs) but also because, even for the known components of the system, we still lack coherent ontological conventions across databases which may aid in connecting these data layers. Integrating this extended universe of possible molecules and their transformations across space, time, and species in complex ecologies will require fundamentally new approaches, and orders of magnitude more computing power, than are available today.

1.3.4 The Need for Standards

Another branch of non-hypothesis-driven research, but critically important to framing precise hypotheses, is the development of standards. In microbiome science these broadly take three tracks: procedural standards for sample collection and handling, analytical standards for determining the accuracy and fidelity of readouts, and annotation standards for integrating results across studies.

The lack of agreed-on standards stems from the origin of much of microbiome science in the discipline of ecology, where the fundamental questions revolved around finding new kinds of organisms to fill out the phylogenetic tree of life, and around finding statistically significant differences in microbial diversity or composition among sets of samples within the context of an individual study. Because the goal was to test whether any difference existed in the microbiome or metabolome as a function of disease, physiological, or environmental state, biases (including missing taxa, or missing classes of molecules) were not terribly important as long as a difference could be discovered.

However, this situation diverges radically from the present situation, where physicians and engineers expect to be able to measure the correct, absolute abundance of all microbes or molecules in a given sample simultaneously. The realities of nucleic acid or organic extraction, detection methods for sequences and molecules, and downstream data processing simply do not support this important goal. However, in general, we don't even know how far we are from it, or what the specific blind spots are. Consequently, without consistent and well-defined measurements underpinned by a mechanistic causal model of change, the state of microbiome-based predictions is much more like astrology than like astronomy.

In order to move from pre-science to science in predicting microbiome changes, we need known reference standards that can be spiked into samples at different stages, from original specimen to DNA or molecule, that are agreed on, widely used in the field, and have an inexhaustible supply. Previous efforts, such as the HMP standards, have been limited by insufficient availability of materials, taxonomic complexity, or both. KatharoSeq in particular [38] benefits from having different spike-in standards at the level of the primary sample and at the level of DNA, allowing different sources of contamination to be tracked down. Comparable development in mass spectrometry, perhaps with isotope-labeled molecules or molecules otherwise unlikely to occur in biological specimens and that can be introduced at different steps, would be of tremendous value.

Sample collection and storage can introduce biases of varying degrees in specimen readout [39-41], but for most sample types the precise implications of different forms of degradation are unknown. Consequently, the conservative recommendation is always to expensively collect pristine samples (e.g. flash-frozen in liquid nitrogen), even while more practical methods would often suffice. For a few sample types, such as amplicon processing of stool, considerable data is now available on a range of conditions [41-44], and researchers can make more informed decisions about which methods to use. However, we know much less about the implications of sample degradation for most other types of biospecimens, and for the implications for reading out different molecular fractions with mass spectrometry (although see [45]). Understanding these principles would greatly expand accessibility of these techniques to field, clinical, and self-collected specimens (by patients and citizen-scientists), as the American Gut Project is already doing for amplicon collection from stool.

Finally, integrating samples from different studies remains extremely challenging because of differences in annotation (often called "metadata"). For example, different studies may refer to "stool", "feces", "gut", or other synonyms or rely on different units of measurement (e.g., Celsius vs. Fahrenheit). Efforts such as the Genomic Standards Consortium MIxS family of standards [46], the Earth Microbiome Project Ontology (EMPO) [12], and other annotation schemes assist considerably in these tasks, but have been applied to relatively few datasets to date. The potential for natural language processing (NLP) and/or data-based methods for automatically applying annotations, perhaps semi-supervised by human guidance, is considerable. These types of strategies were successful in Qiita for inferring EMPO annotations for tens of thousands of samples in Qiita primarily based off the researcher reported "sample_type." Resources like Qiita, which allow researchers to deposit microbiome studies, provide mechanisms to help researchers use standard compliant metadata. However, further development is necessary to enable researchers to "discover" the types of variables and controlled vocabularies that are in common across the resource.

1.3.5 Conclusions

Although hypothesis-driven science has immense value, it depends to a considerable degree on a framework of maps, tools, and standards whose development often does not fit meaningfully into a hypothesis-driven framework and is therefore heavily criticized in settings such as grant review panels. However, without these types of development, hypotheses more explicit than "differences in the microbiome" or "elevation or depletion of specific taxa or molecules" cannot be tested, and completely new ideas about how to read out or control the microbiome will not be developed.

Extraordinary advances in data collection technologies leave us in a world where we regularly make millions of observations of organisms about which we know virtually nothing -- as exemplified by the recent 'discovery' of the most abundant phage in the human gut via metagenome mining [47]. The amount of information contained in these observations in principle is enough to allow us to fine-tune more labor-intensive experiments to test critical questions with great efficiency. In practice, though, much of this information remains inaccessible. In order to bring about a future of precision medicine and precision ecological remediation, where we can specify precise microbiome changes and bring them about through defined interventions, a vast amount of non-hypothesis-driven research, often dismissed as “technical work” or “fishing expeditions”, remains to be done.

1.3.6 Acknowledgements

This work was supported in part by National Institute of Justice Award 2015-DN-BX-K047, by the Alfred P. Sloan Foundation, and by the National Institutes of Health.

Chapter I, Introduction part 3, in full, is a reprint of previously published material: Tripathi A, Marotz C, Gonzalez A, Vázquez-Baeza Y, Song SJ, Bouslimani A, McDonald D, Zhu Q, Sanders JG, Smarr L, Dorrestein PC. *Are microbiome studies ready for hypothesis-driven research?* Current opinion in microbiology. 2018 Aug 1;44:61-9. I was one of the primary investigators and authors of this paper. The co-authors listed above supervised or provided support for the research and have given permission for the inclusion of the work in this dissertation.

1.3.7 References

1. Ruse M: The Darwinian Revolution: Science Red in Tooth and Claw. Chicago: University of Chicago Press; 1999.

2. Cohen IB: The First English Version of Newton's Hypotheses non fingo. Isis 1962, 53:379-388.

3. Anderson C: THE END OF THEORY: THE DATA DELUGE MAKES THE SCIENTIFIC METHOD OBSOLETE. In Wired. Edited by. San Francisco: Conde Nast; 2008.

4. Mazzocchi F: Could Big Data be the end of theory in science? A few remarks on the epistemology of data-driven science. EMBO Rep 2015, 16:1250-1255.

5. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, et al.: Diet rapidly and reproducibly alters the human gut microbiome. Nature 2014, 505:559-563.

6. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, et al.: Linking long-term dietary patterns with gut microbial enterotypes. Science 2011, 334:105-108.

7. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R: Bacterial community variation in human body habitats across space and time. Science 2009, 326:1694-1697.

8. Human Microbiome Project C: Structure, function and diversity of the healthy human microbiome. Nature 2012, 486:207-214.

9. Davenport ER, Mizrahi-Man O, Michelini K, Barreiro LB, Ober C, Gilad Y: Seasonal variation in human gut microbiome composition. PLoS One 2014, 9:e90731.

10. Smits SA, Leach J, Sonnenburg ED, Gonzalez CG, Lichtman JS, Reid G, Knight R, Manjurano A, Changalucha J, Elias JE, et al.: Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. Science 2017, 357:802-806.

11. Lozupone CA, Knight R: Global patterns in bacterial diversity. Proc Natl Acad Sci U S A 2007, 104:11436-11440.

12. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, et al.: A communal catalogue reveals Earth's multiscale microbial diversity. Nature 2017, 551:457-463.

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

14. McCafferty J, Muhlbauer M, Gharaibeh RZ, Arthur JC, Perez-Chanona E, Sha W, Jobin C, Fodor AA: Stochastic changes over time and not founder effects drive cage effects in microbial community assembly in a mouse model. ISME J 2013, 7:2116-2125.

15. Debelius J, Song SJ, Vazquez-Baeza Y, Xu ZZ, Gonzalez A, Knight R: Tiny microbes, enormous impacts: what matters in gut microbiome studies? Genome Biol 2016, 17:217.

16. da Silva RR, Dorrestein PC, Quinn RA: Illuminating the dark matter in metabolomics. Proc Natl Acad Sci U S A 2015, 112:12549-12550.

17. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, et al.: Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol 2016, 34:828-837.

18. Gonzalez A, Knight R: Advancing analytical algorithms and pipelines for billions of microbial sequences. Curr Opin Biotechnol 2012, 23:64-71.

19. Bouslimani A, Melnik AV, Xu Z, Amir A, da Silva RR, Wang M, Bandeira N, Alexandrov T, Knight R, Dorrestein PC: Lifestyle chemistries from phones for individual profiling. Proc Natl Acad Sci U S A 2016, 113:E7645-E7654.

20. Bouslimani A, Porto C, Rath CM, Wang M, Guo Y, Gonzalez A, Berg-Lyon D, Ackermann G, Moeller Christensen GJ, Nakatsuji T, et al.: Molecular cartography of the human skin surface in 3D. Proc Natl Acad Sci U S A 2015, 112:E2120-2129.

21. Garg N, Wang M, Hyde E, da Silva RR, Melnik AV, Protsyuk I, Bouslimani A, Lim YW, Wong R, Humphrey G, et al.: Three-Dimensional Microbiome and Metabolome Cartography of a Diseased Human Lung. Cell Host Microbe 2017, 22:705-716 e704.

22. Kapono CA, Morton JT, Bouslimani A, Melnik AV, Orlinsky K, Knaan TL, Garg N, Vazquez-Baeza Y, Protsyuk I, Janssen S, et al.: Creating a 3D microbial and chemical snapshot of a human habitat. Sci Rep 2018, 8:3669.

23. Petras D, Nothias LF, Quinn RA, Alexandrov T, Bandeira N, Bouslimani A, Castro-Falcon G, Chen L, Dang T, Floros DJ, et al.: Mass Spectrometry-Based Visualization of Molecules Associated with Human Habitats. Anal Chem 2016, 88:10775-10784.

24. Protsyuk I, Melnik AV, Nothias LF, Rappez L, Phapale P, Aksenov AA, Bouslimani A, Ryazanov S, Dorrestein PC, Alexandrov T: 3D molecular cartography using LC-MS facilitated by Optimus and 'ili software. Nat Protoc 2018, 13:134-154.

25. Paneth N, Vinten-Johansen P, Brody H, Rip M: A rivalry of foulness: official and unofficial investigations of the London cholera epidemic of 1854. Am J Public Health 1998, 88:1545-1553.

26. Weingarden A, Gonzalez A, Vazquez-Baeza Y, Weiss S, Humphry G, Berg-Lyons D, Knights D, Unno T, Bobr A, Kang J, et al.: Dynamic changes in short- and long-term bacterial composition following fecal microbiota transplantation for recurrent Clostridium difficile infection. Microbiome 2015, 3:10.
••This paper integrates timeseries microbiome data from Clostridium difficile patients with the HMP data frame to show their progress through the "microbiome map" provided by the HMP.

27. Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N: Microbial strain-level population structure and genetic diversity from metagenomes. Genome Res 2017, 27:626-638.

28. Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K: Improving saliva shotgun metagenomics by chemical host DNA depletion. Microbiome 2018, 6:42.

29. Aksenov AA, da Silva R, Knight R, Lopes NP, Dorrestein PC: Global chemical analysis of biology by mass spectrometry. Nat Rev Chem 2017, 1:0054.

30. Sauerschnig C, Doppler M, Bueschl C, Schuhmacher R: Methanol Generates Numerous Artifacts during Sample Extraction and Storage of Extracts in Metabolomics Research. Metabolites 2017, 8.

31. Scheubert K, Hufsky F, Petras D, Wang M, Nothias LF, Duhrkop K, Bandeira N, Dorrestein PC, Bocker S: Significance estimation for large scale metabolomics annotations by spectral matching. Nat Commun 2017, 8:1494.

32. Members MSIB, Sansone SA, Fan T, Goodacre R, Griffin JL, Hardy NW, Kaddurah-Daouk R, Kristal BS, Lindon J, Mendes P, et al.: The metabolomics standards initiative. Nat Biotechnol 2007, 25:846-848.

33. Salek RM, Conesa P, Cochrane K, Haug K, Williams M, Kale N, Moreno P, Jayaseelan KV, Macias JR, Nainala VC, et al.: Automated assembly of species metabolomes through data submission into a public repository. Gigascience 2017, 6:1-4.

34. Haug K, Salek RM, Conesa P, Hastings J, de Matos P, Rijnbeek M, Mahendraker T, Williams M, Neumann S, Rocca-Serra P, et al.: MetaboLights--an open-access general-purpose repository for metabolomics studies and associated meta-data. Nucleic Acids Res 2013, 41:D781-786.

35. Sud M, Fahy E, Cotter D, Azam K, Vadivelu I, Burant C, Edison A, Fiehn O, Higashi R, Nair KS, et al.: Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. Nucleic Acids Res 2016, 44:D463-470.

36. Nicholson JK, Lindon JC: Systems biology: Metabonomics. Nature 2008, 455:1054-1056.
37. Muegge BD, Kuczynski J, Knights D, Clemente JC, Gonzalez A, Fontana L, Henrissat B, Knight R, Gordon JI: Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science 2011, 332:970-974.

38. Minich JJ, Zhu Q, Janssen S, Hendrickson R, Amir A, Vetter R, Hyde J, Doty MM, Stillwell K, Benardini J, et al.: KatharoSeq Enables High-Throughput Microbiome Analysis from Low-Biomass Samples. mSystems 2018, 3.

39. Gika HG, Theodoridis GA, Wilson ID: Liquid chromatography and ultra-performance liquid chromatography-mass spectrometry fingerprinting of human urine: sample stability under different handling and storage conditions for metabonomics studies. J Chromatogr A 2008, 1189:314-322.

40. Lou JJ, Mirsadraei L, Sanchez DE, Wilson RW, Shabihkhani M, Lucey GM, Wei B, Singer EJ, Mareninov S, Yong WH: A review of room temperature storage of biospecimen tissue and nucleic acids for anatomic pathology laboratories and biorepositories. Clin Biochem 2014, 47:267-273.

41. Song SJ, Amir A, Metcalf JL, Amato KR, Xu ZZ, Humphrey G, Knight R: Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies. mSystems 2016, 1.

42. Choo JM, Leong LE, Rogers GB: Sample storage conditions significantly influence faecal microbiome profiles. Sci Rep 2015, 5:16350.

43. Hale VL, Tan CL, Niu K, Yang Y, Cui D, Zhao H, Knight R, Amato KR: Effects of field conditions on fecal microbiota. J Microbiol Methods 2016, 130:180-188.

44. Vogtmann E, Chen J, Amir A, Shi J, Abnet CC, Nelson H, Knight R, Chia N, Sinha R: Comparison of Collection Methods for Fecal Samples in Microbiome Studies. Am J Epidemiol 2017, 185:115-123.

45. Loftfield E, Vogtmann E, Sampson JN, Moore SC, Nelson H, Knight R, Chia N, Sinha R: Comparison of Collection Methods for Fecal Samples for Discovery Metabolomics in Epidemiologic Studies. Cancer Epidemiol Biomarkers Prev 2016, 25:1483-1490.

46. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, Gilbert JA, Karsch-Mizrachi I, Johnston A, Cochrane G, et al.: Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat Biotechnol 2011, 29:415-420.

47. Yutin N, Makarova KS, Gussow AB, Krupovic M, Segall A, Edwards RA, Koonin EV: Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. Nat Microbiol 2018, 3:38-46.

Chapter 2.

Technical advances for probing host-associated microbiomes

An incredible amount of knowledge has been gained through the advent of next generation sequencing and our ability to perform taxonomic profiling via amplicon sequencing. But there are still many limiting factors. The bottlenecks addressed in this chapter are 1) high throughput DNA extraction, 2) streamlined PCR amplification for next generation sequencing library preparation, and 3) host DNA depletion for shotgun sequencing. Together, these advances represent detailed protocols that allow for scaled host-associated microbiome analyses.

2.1 DNA extraction for streamlined metagenomics of diverse environmental samples

A major bottleneck for metagenomic sequencing is rapid and efficient DNA extraction. We compared the extraction efficiency of three magnetic bead-based platforms (KingFisher, epMotion, and Tecan) to a standardized column-based extraction platform across a variety of samples including feces, oral, skin, soil, and water. Replicate sample plates were extracted and prepared for 16S rRNA gene amplicon sequencing in parallel to assess extraction bias and DNA quality. The results found that any effect of extraction method on sequencing results was small compared to variability across samples and highlighted one platform for producing the largest number of high-quality reads in the shortest amount of time. This study thus identified an extraction pipeline that dramatically reduces sample processing time without sacrificing bacterial taxonomic or abundance information.

2.1.1 Introduction

As microbiome analyses become applicable to an increasing number of scientific areas, a streamlined process for efficiently extracting DNA to generate 16S rRNA gene amplicon (16S amplicons) or shotgun metagenomic sequencing data from a range of environmental sample types is increasingly important. DNA extraction is among the largest sources of experimental variability in 16S sequence analysis [1-3]. Complete bacterial lysis and DNA purification can be particularly time-consuming, and represent a significant bottleneck for high-throughput analyses. However, certain time-sensitive analyses with large sample sizes (e.g., clinical diagnostics or water quality

assessments) require expedient processing. We aimed to identify a DNA extraction method that yields faster results with no cost in taxonomic representation by comparing magnetic bead-based extraction robots with our benchmarked column-based extraction protocol.

2.1.2 Results

To compare the extraction efficiency of several automated methods, we collected environmental samples from a wide range of biological materials. Replicate plates were prepared from aliquots of identical samples for comparison of extraction efficiency among different platforms. In total, 48 fecal samples, 12 soil samples, 12 marine sediment samples, 6 seawater samples, 5 skin samples, 5 oral samples, and 6 mattress dust samples were included. Material for extraction was collected by swabbing each sample with Puritan wooden handle cotton swabs according to the Earth Microbiome Project standard protocol [http://www.earthmicrobiome.org/emp500/emp500-sample-submission-guide/]. Fecal samples were comprised of 2 mL homogenized human stool sample obtained from the Microbiome Quality Control Project (MBQC). Soil samples were selected from a published mouse decomposition study [4]. Marine sediment samples were donated by Dr. Paul Jensen and collected at a depth of either 300 or 700 m off the coast of San Diego. Seawater samples (4 L each) were collected off the Scripps Pier in San Diego, CA and filtered through a 0.22 μm Sterivex cartridge; the filters were then cut into segments and used for extraction [5]. Oral and saliva samples were collected using 5 swabs simultaneously (one for each extraction method) as previously described [http://americangut.org/how-it-works/, 6] from a total of six volunteers over three time points. Finally, three mattresses were vacuumed in duplicate and the resulting dust was swabbed for extraction. Approximately 10^6 *Vibrio fischerii* ES114 were used as a positive control, and one

blank well (no input material) was used as a negative control. Together, these samples represent a wide range of biological material, including some (soil and fecal samples) that had already been well characterized with widely accepted protocols. This diverse sample set permitted a broad assessment of the different extraction pipeline efficiencies.

We adapted the MoBio PowerMag Soil DNA isolation kit (Qiagen, Carlsbad, CA, USA) for use with a magnetic bead plate in place of the silica membrane column. This replacement significantly reduces sample preparation time by eliminating the need to sequentially load the silicon membrane three times for each sample. Importantly, it requires appreciably less hands-on time, although total processing time is also reduced. Replicate plates were extracted in parallel to compare the standard column-based method to three magnetic bead-based nucleic acid purification platforms; epMotion® 5075 TMX (TMX; Eppendorf, Hamburg, Germany), KingFisher™ Flex Purification System (KingFisher; ThermoFisher Scientific, Waltham, MA, USA), and Tecan Freedom EVO® Nucleic Acid Purification (Tecan; Tecan, Morrisville, NC, USA). An additional KingFisher run (KF2) was included with an overnight incubation at 4°C immediately prior to magnetic bead clean-up to test the processing flexibility of this system. All samples were eluted in 100 ul PCR grade H2O. DNA yield was broadly similar across extraction platforms except for the Tecan, which was lower (Fig. 1A). Notably, the Kingfisher extraction platform yielded the most gDNA from low biomass samples. DNA was prepared for 16S rRNA gene amplicon sequencing as previously described [7]. The demultiplexed fasta files were obtained from qiita.ucsd.edu (study ID 10178), and operational taxonomic units (OTUs) were detected using Deblur [8]. Raw reads were submitted to the European Bioinformatics Institute under accession number ERP021045.

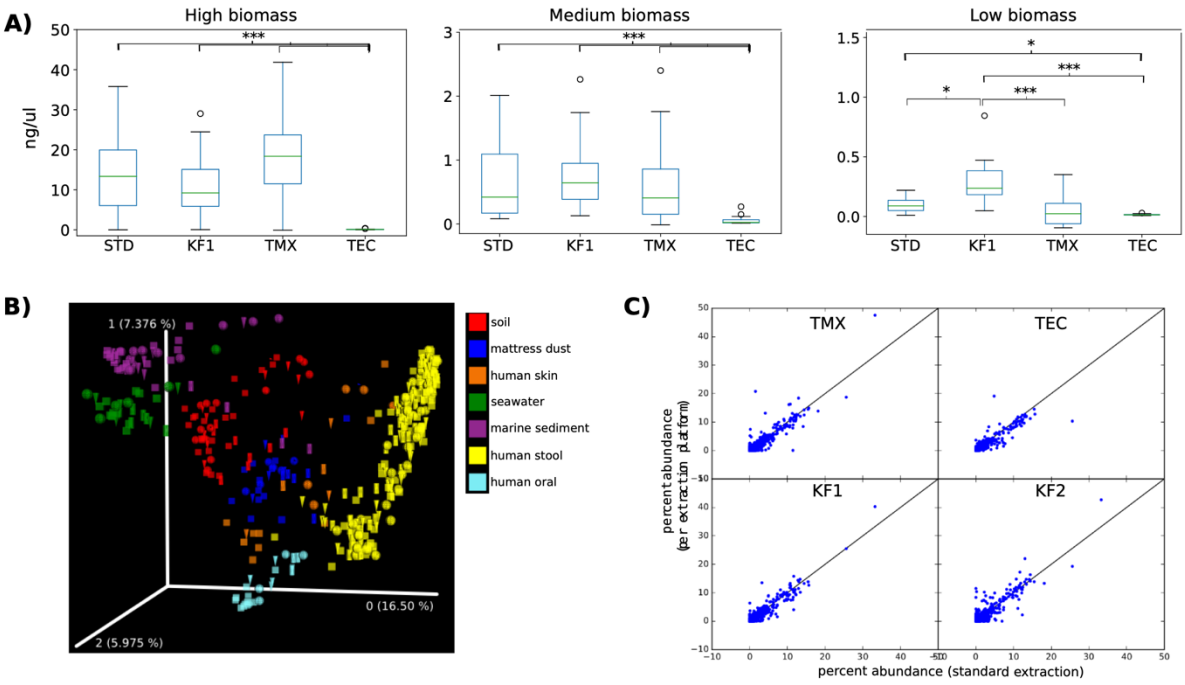

Figure 2.1.1. *Inter-sample variability outweighs extraction method bias*. **A)** Average concentration of DNA in ng/ul across extraction platforms. Statistics performed using non-parametric binomial two sided sign test; *<0.05, **<.01, ***<.001. **B)** Average number of quality filtered reads across extraction platforms. **C)** Unweighted principal coordinates analysis of all samples colored by sample type. Extraction method is denoted by shape: sphere = column-based, cube = KingFisher, cylinder = Tecan, cone = TMX **D**). Scatter plot showing taxonomic abundance differences between magnetic bead based extraction platform (y-axis) and standard column based extraction method (x-axis). TMX = epMotion TMX, TEC = Tecan, KF1 = KingFisher, KF2 = KingFisher with overnight pause.

We found that the bias introduced by extraction method is small compared to inter-sample variation (average distance across biological replicates, extraction method, and sample type, respectively; weighted UniFrac 0.11±0.001, 0.19±0.004, 0.29±0.001; unweighted UniFrac 0.44±0.01, 0.51±0.003, 0.63±0.001; Bray-Curtis 0.22±0.02, 0.38±0.007, 0.71±0.003). Confirming this, principal coordinates analysis (PCoA) of unweighted UniFrac distance matrices revealed that sequences clustered by sample type rather than extraction platform (Fig. 1B) [9]. OTU abundance across the magnetic extraction platforms revealed broadly similar results to the standard column-based protocol, even when the KingFisher extraction protocol was paused overnight before bead clean-up (Fig. 1C). Per sample alpha diversity levels were remarkably similar (within 2% of the

average across all extraction platforms) with the exception of samples extracted on the Tecan which produced on average 92%±2% of the average observed diversity. These results demonstrate that the taxonomic and community-level variation contributed by the different magnetic based extraction protocols was minor.

However, the number of quality sequencing reads across all sample types was highest in the KingFisher-extracted samples (Fig. 2A-B). Importantly, the KingFisher protocol requires only a fraction of the time required by the other extraction platforms. This streamlined extraction pipeline saves an average of 72 h per 192 samples compared to single-tube manual extraction and cuts the processing time in half compared to the second fastest method (Fig. 2C).

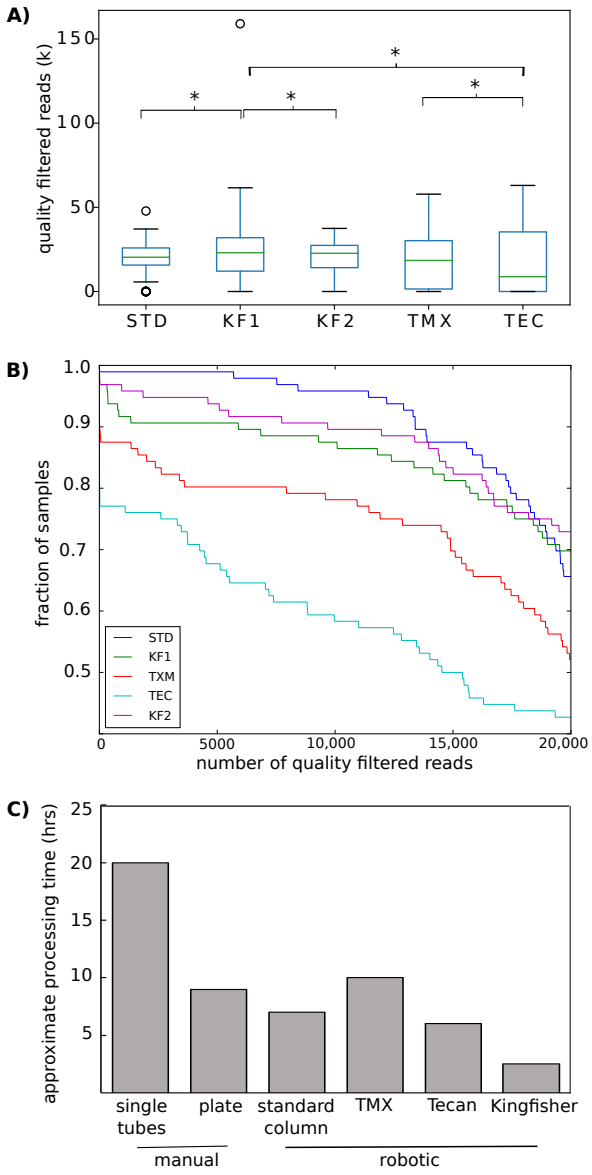

Figure 2.1.2. *DNA extracted on the KingFisher platform provides the highest quality reads and requires the least amount of processing time.* **A)** Average number of quality filtered reads produced by DNA extracted across different platforms. Error bars display standard error. Statistics performed using non-parametric binomial two sided sign test; *<0.05, **<.01, ***<.001. **B)** Kaplan–Meier estimator showing fraction of samples versus number of quality filtered reads for all samples on each extraction platform. **C)** Comparison of average time to process 96 samples across extraction protocols. Manual extractions were performed according to the manufacturer's protocol. The magnetic bead-based protocol on the KingFisher platform cuts the processing time in half from the second fastest platform.

2.1.3 Conclusion

We prepared replicate 96-well plates loaded with swabs from a variety of sources including feces, skin, oral, soil, and water. The MoBio PowerMag Soil DNA isolation kit was adapted for comparison of three magnetic bead isolation platforms for representative 16S rRNA sequencing results. Ultimately, we found the KingFisher Flex Purification system was the fastest, the most efficient for low biomass samples, and retained the largest number of high-quality amplicon sequencing reads. This streamlined protocol reduces the total time of DNA extraction to one-fourth of the original protocol while providing comparable sequencing results.

2.1.4 Acknowledgements

Chapter II, part 1, in full, is a reprint of previously published material: Marotz C, Amir A, Humphrey G, Gaffney J, Gogul G, Knight R. *DNA extraction for streamlined metagenomics of diverse environmental samples. Biotechniques*. 2017 Jun;62(6):290-3. I was the primary investigator and author of this paper. The co-authors listed above supervised or provided support for the research and have given permission for the inclusion of the work in this dissertation.

2.1.5 References

1. Wu, G.D., Lewis, J.D., Hoffmann, C., Chen, Y.Y., Knight, R., Bittinger, K., Hwang, J., Chen, J., Berkowsky, R., Nessel, L. and Li, H., 2010. Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. BMC microbiology, 10(1), p.206.

2. Yuan, S., Cohen, D.B., Ravel, J., Abdo, Z. and Forney, L.J., 2012. Evaluation of methods for the extraction and purification of DNA from the human microbiome. PloS one, 7(3), p.e33865.

3. Kennedy, N.A., Walker, A.W., Berry, S.H., Duncan, S.H., Farquarson, F.M., Louis, P. and Thomson, J.M., 2014. The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. PloS one, 9(2), p.e88982.

4. Metcalf, J.L., Parfrey, L.W., Gonzalez, A., Lauber, C.L., Knights, D., Ackermann, G., Humphrey, G.C., Gebert, M.J., Van Treuren, W., Berg-Lyons, D. and Keepers, K., 2013. A microbial clock provides an accurate estimate of the postmortem interval in a mouse model system. Elife, 2, p.e01104.

5. Gilbert, J.A., Field, D., Swift, P., Thomas, S., Cummings, D., Temperton, B., Weynberg, K., Huse, S., Hughes, M., Joint, I. and Somerfield, P.J., 2010. The taxonomic and functional diversity of microbes at a temperate coastal site: a 'multi-omic'study of seasonal and diel temporal variation. PloS one, 5(11), p.e15545.

6. Bouslimani, A., Porto, C., Rath, C.M., Wang, M., Guo, Y., Gonzalez, A., Berg-Lyon, D., Ackermann, G., Christensen, G.J.M., Nakatsuji, T., Zhang, L., Borkowski, A., Meehan, M., Dorrestein, K., Gallo, R., Bandeira, N., Knight, R., Alexandrov, T., and Dorrestein, P. 2015. Molecular cartography of the human skin surface in 3D. Proceedings of the National Academy of Sciences, 112(17), pp.E2120-E2129.

7. Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M. and Gormley, N., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME journal, 6(8), pp.1621-1624.

8. Amir, A., McDonald, D., Navas-Molina, J.A., Kopylova, E., Morton, J.T., Xu, Z.Z., Kightley, E.P., Thompson, L.R., Hyde, E.R., Gonzalez, A. and Knight, R., 2017. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems, 2(2), pp.e00191-16.

9. Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A. and Knight, R., 2013. EMPeror: a tool for visualizing high-throughput microbial community data. Gigascience, 2(1), p.16.

2.2

Triplicate PCR reactions for 16S rRNA gene amplicon sequencing are unnecessary

Conventional wisdom holds that PCR amplification for sequencing should employ pooled replicate reactions to reduce bias due to jackpot effects and chimera formation. However, modern amplicon data analysis employs methods that may be less sensitive to such artifacts. Here we directly compare results from single vs. triplicate reactions for 16S amplicon sequencing and find no significant impact of adopting a less labor-intensive single reaction protocol.

2.2.1 Introduction

For decades, 16S rRNA gene sequencing has been performed by pooling replicate PCR reactions, usually in triplicate. The primary benefit is to reduce "jackpotting": the stochastic nature of PCR means that some molecules are amplified earlier than others, and exponential amplification in subsequent rounds of PCR substantially distort the frequencies of different molecules in heterogeneous pools of target genes [1]. This phenomenon is particularly important in environmental DNA sequencing where the goal is an accurate, or at least consistent, readout of the different gene targets matching a primer set.

However, since the guideline that PCR should be performed in triplicate was introduced [1], there have been substantial improvements in the processivity and fidelity of DNA polymerases. Therefore, triplicate PCR may no longer provide the benefits it once did, although performing single PCR reactions instead of triplicate would provide significant time and cost savings. Several studies have tested single versus triplicate PCR for 16S rRNA sequencing in

limited settings with a small number of input samples (e.g. 18 soil samples [2], 2 soil and 2 stool samples [3], 3 soil samples [4]). However, it has never been tested across the wide range of samples and settings that would be needed to justify a general recommendation for change in protocol. We used the availability of standardized sample sets such as those from MBQC, the Microbiome Quality Control project[5], and from our previous technology testing to answer this question definitively across three different laboratories. In total, we tested the effects of replicate PCR pooling in 3 independent experiments containing nearly 373 samples from a diverse range of environments.

2.2.2 Results

First, we benchmarked single versus pooled-triplicate PCR across a broad range of sample types. In our previous study on comparison of DNA extraction methods [6] we assembled a set of 96 samples spanning a broad range of environments, including 48 fecal samples, 12 soil samples, 12 marine sediment samples, 6 seawater samples, 5 skin samples, 5 oral samples, and 6 mattress dust samples. We used the DNA from this previous study, extracted using the Earth Microbiome Project protocol [7] on the Kingfisher instrument, for this study. 16S rRNA gene amplification was performed according to the Earth Microbiome Project (EMP) protocol and is detailed in the supplemental file. We quantified amplicons by PicoGreen™ and pooled 240 ng of each for sequencing. We ran the entire sample set four times: twice with single PCR and twice with pooled-triplicate PCR. The pooled library was sequenced on the Illumina MiSeq sequencing platform with a MiSeq Reagent Kit v2 and paired-end 150 cycles. All data were processed and analyzed using the QIIME2 software suite [8] and Deblur [9]. Counterintuitively, single PCR reactions yielded significantly more reads than triplicate PCR reactions (mean ± SEM: 10,821 ± 298 versus 10,029

± 262, respectively, paired T-test p=0.0003), and fewer dropouts (Fig. 1A). We saw no significant difference in alpha diversity, regardless of environment (Fig. 1B). Beta diversity analysis with Unweighted UniFrac demonstrates that samples cluster by sample type and not number of PCR reactions (Fig. 1C). The Weighted UniFrac distances are significantly larger among samples from different environments than among biological replicates, and distances among biological replicates are significantly greater than technical replicates, with both single and triplicate PCR reactions (Fig. 1D). Negligible taxonomic changes between single and triplicate reactions were observed (97.8% shared taxonomy at the species level, genus 98.4%, and phylum 100%, Fig. S1A and Fig. S2).

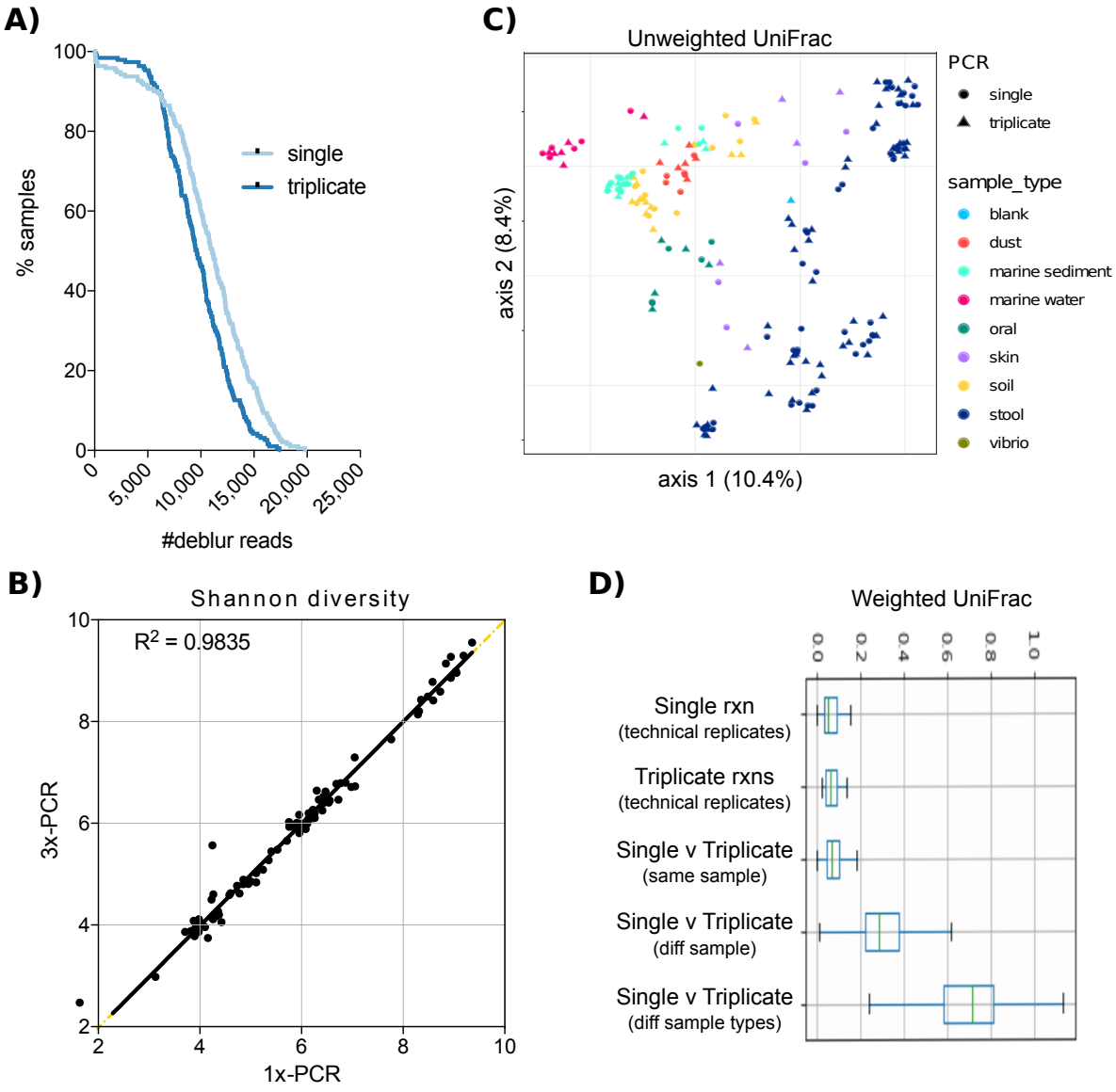

Figure 2.2.1. *Effect of 16S PCR reaction number across a broad range of sample types.* **A)** The sequencing dropout rate of all samples run with either single or triplicate PCR reactions**. B)** Shannon diversity index is nearly identical between single and triplicate PCR reactions of the same sample. **C)** Unweighted UniFrac PCoA plot shows that samples cluster by sample type (color) and not number of PCR reactions (shape). **D)** Weighted UniFrac distance among technical replicates (same sample) run with either single or triplicate PCR reactions are smaller than the distance between samples of the same type (diff samples) or among samples from different environments (types).

Second, because high-level conclusions crossing environment types might obscure relationships in particular sample types, we tested whether the conclusions held for a separate set of agricultural samples. We sampled root and rhizosphere samples from 3 different sites across 2 seasons. A variety of roots including crown, seminal, and primary roots were excavated and shaken for 1-2 min in 35 mL phosphate buffer and maintained on ice. In the laboratory, roots were surface sterilized by rinsing 30 seconds in 5.25% sodium hypochlorite + 0.01% Tween 20, followed by a 30 seconds rinse in 70% ethanol, followed by three rinses in sterile ultrapure water. Roots were blotted dry on a clean paper towel, placed in a 15 mL tube, frozen at -80° C and then ground in liquid nitrogen prior to DNA extraction. The rhizosphere samples were filtered through a sterile 100 μm mesh filter, pelleted at 3000 x g for 10 minutes, washed with 1.5 mL phosphate buffer, and re-pelleted by spinning for 5 minutes at full speed. The supernatant was drained off and the rhizosphere soil pellet was stored at -80° C until DNA extraction. DNA was extracted from soil, rhizosphere, and root samples using DNeasy PowerSoil HTP 96 Kit and quantified with the Quantifluor dsDNA reagent. Each sample was amplified both with a single PCR reaction and with pooled-triplicate reactions. The single PCR reactions yielded significantly more reads than triplicate PCR reactions (mean ± SEM: 3,631 ± 139 versus 3,000 ± 113, respectively, paired T-test $p<0.0001$), but had a similar dropout rate (Fig. 2A). Alpha diversity was not significantly different with single versus triplicate PCR (Fig. 2B), and as with the cross-environment comparison shown in Fig. 1, Weighted UniFrac analysis shows that the primary clustering is by sample type and the distances among samples does not differ in single versus triplicate PCRs (Fig. 2C,D). Negligible taxonomic changes between single and triplicate reactions were observed (99.3% shared taxonomy at the species level, genus 99.2%, and phylum 100%, Fig. S1B and Fig. S3).

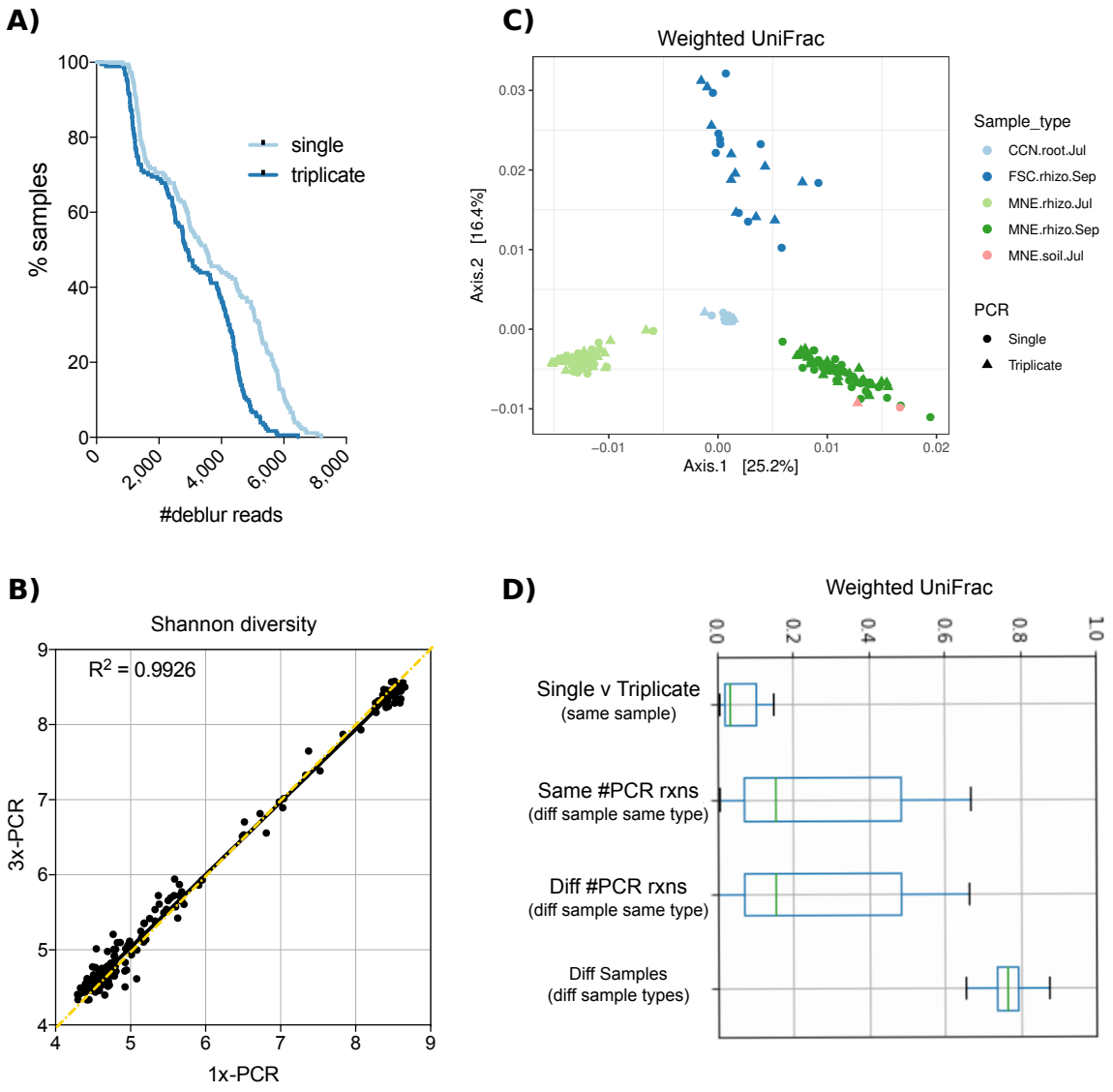

Figure 2.2.2. *Effect of 16S PCR reaction number across agricultural samples.*
A). The sequencing dropout number of all samples run with either single or triplicate PCR reactions. **B)** Shannon diversity index of each sample is similar between single and triplicate PCR reactions. **C)** Weighted UniFrac PCoA plot shows that samples cluster by sample type (color) and not number of PCR reactions (shape). **D)** Weighted Unifrac distances between single or triplicate PCR reactions of the same sample are smaller than the distance between different samples of the same type run with either single or triplicate PCR reactions, and both are smaller than the distance between samples from different environments.

Finally, the microbiology of the built environment has been a rapidly expanding topic of interest over the past decade, but poses unique challenges for molecular analysis. In particular, samples tend to be contaminated with high levels of human DNA and have low bacterial biomass [10]. We used samples from a previous study that collected 96 samples longitudinally from four commonly used building materials maintained at a high relative humidity (~94%) [11]. Genomic DNA was extracted from environmental samples using the PowerSoil DNA isolation kit as previously described[12], and genomic DNA was amplified using the EMP protocol as detailed in the supplemental file. Samples were processed both with single PCR and pooled-triplicate PCR reactions, and sequenced on an Illumina MiSeq sequencing platform with a MiSeq Reagent Kit v2 and paired-end 150 cycles. Once again, yields were higher with single PCR than triplicate PCR (Fig. 3A), Shannon diversity was not affected by single versus triplicate PCR (Fig. 3B), and beta diversity was driven by biological parameters of the sample rather than by single versus triplicate PCR (Fig. 3C,D). Negligible taxonomic changes between single and triplicate reactions were observed (96.5% shared taxonomy at the species level, genus 95.8%, and phylum 100%, Fig. S1C and Fig. S4). All data from each of the three experiments are publicly available from the EBI under accession number ERP113817.

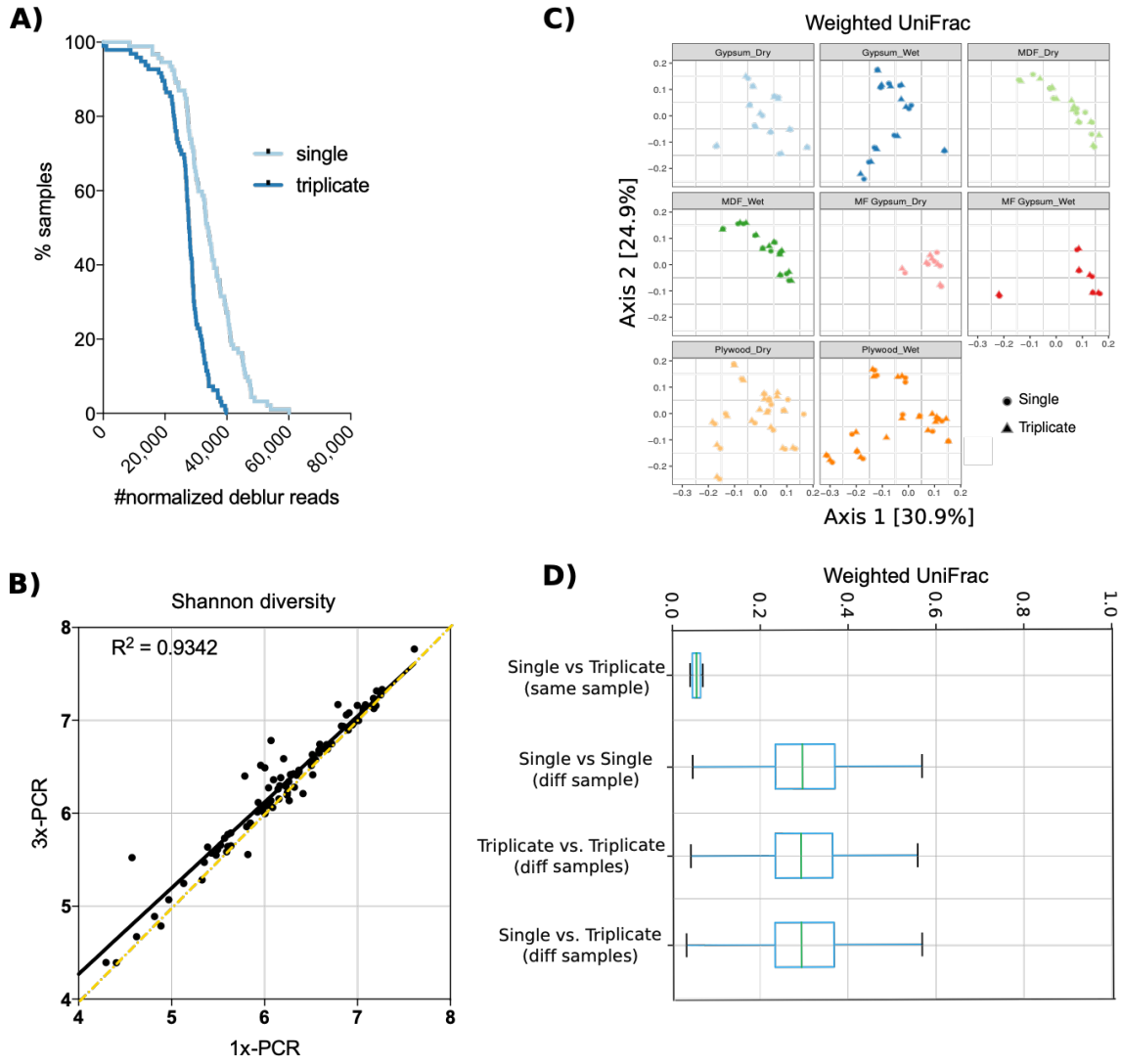

Figure 2.2.3. *Effect of 16S PCR reaction number across building materials.*
A) The sequencing dropout number of all samples run with either single or triplicate PCR reactions. **B)** Shannon diversity index is similar between single and triplicate PCR reactions. **C)** Weighted UniFrac PCoA plot shows that samples do not cluster by number of PCR reactions (shape). **D)** Weighted UniFrac distances between single or triplicate PCR reactions of the same sample are smaller than the distance between different samples with either single or triplicate PCR reactions.

Taken together, these results demonstrate that with modern methods pooling triplicate PCR reactions for 16S rRNA amplicon sequencing is more expensive and does not provide improvement over single PCR reactions. This result was confirmed in studies spanning three laboratories, hundreds of samples, and numerous distinct environment types. However, although these results hold true for the range of conditions tested here, there are so many variations in PCR techniques that this type of benchmarking effort should be validated for specific sample types and PCR protocols before a switch from established procedure is implemented for specialized protocols. For the general sample types tested here, we recommend using single PCRs rather than triplicate PCRs. Combined with other technical improvements in miniaturizing PCR reactions[13], this change in protocol will substantially reduce the cost and complexity of amplicon studies.

2.2.3 Acknolwedgements

Chapter II, part 2, in full, is a reprint of previously published material: Marotz C, Sharma A, Humphrey G, Gottel N, Daum C, Gilbert JA, Eloe-Fadrosh E, Knight R. *Triplicate PCR reactions for 16S rRNA gene amplicon sequencing are unnecessary*. BioTechniques. 2019 Jul;67(1):29-32. I was the primary investigator and author of this paper. The co-authors listed above supervised or provided support for the research and have given permission for the inclusion of the work in this dissertation.

2.2.4 Supplemental Figures

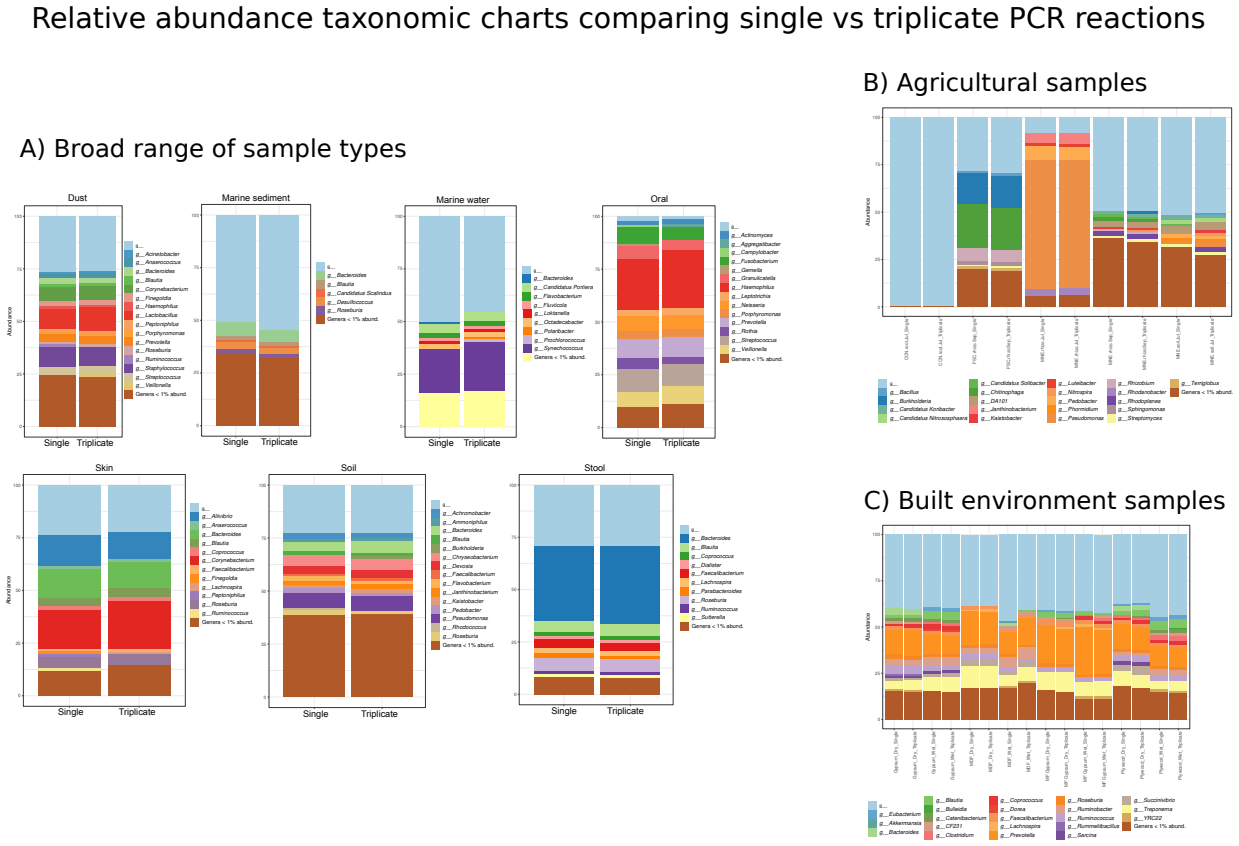

Figure 2.2.S1. *Taxonomic relative abundance profiles comparing single vs triplicate PCR reactions.* Bar charts comparing genera with greater than 1% relative abundance between single and triplicate reactions across a broad range of sample types **(A)**, agricultural samples **(B)**, and built environment samples **(C)**.

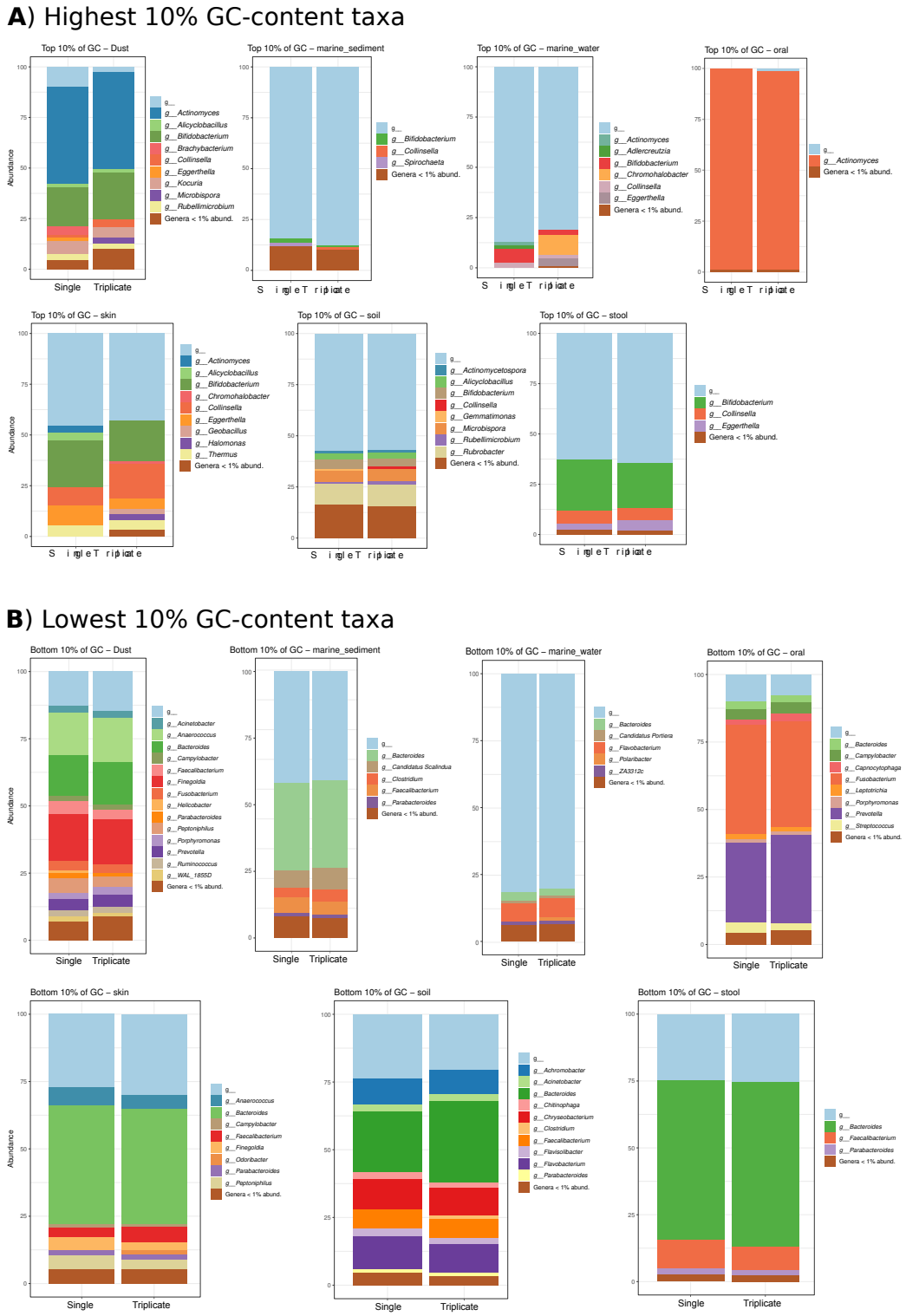

Figure 2.2.S2. *Relative abundance between single and triplicate PCR is not biased by GC content across a broad range of sample types.* Taxa were ranked according to their % GC content, and the highest 10% of genera (A) and lowest 10% of genera (B) are plotted to compare single vs triplicate PCR reactions.

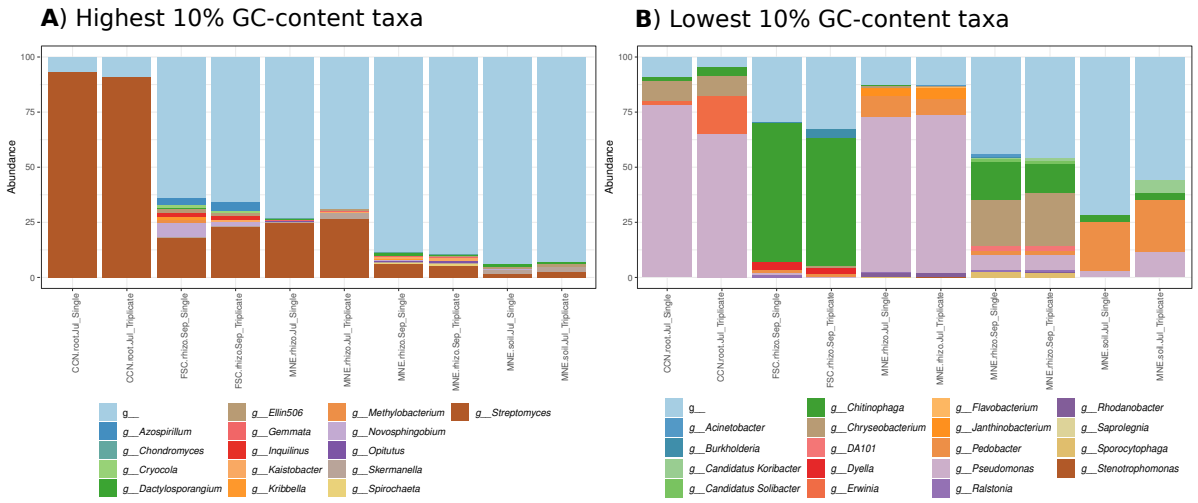

Figure 2.2.S3. *Relative abundance between single and triplicate PCR is not biased by GC content in agricultural samples.* Taxa were ranked according to their % GC content, and the highest 10% of genera **(A)** and lowest 10% of genera **(B)** are plotted to compare single vs triplicate PCR reactions.

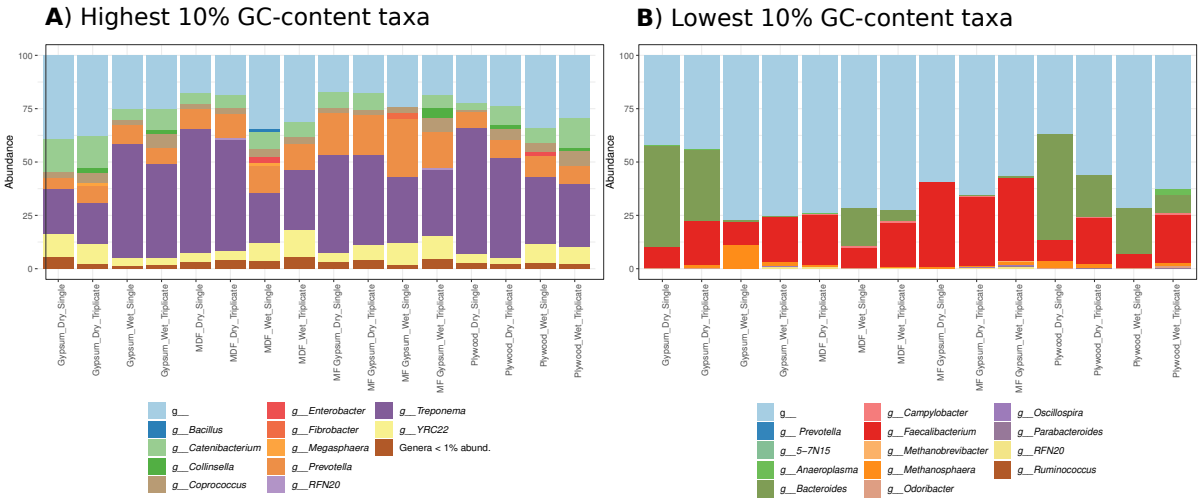

Figure 2.2.S4 *Relative abundance between single and triplicate PCR is not biased by GC content in built environment samples.* Taxa were ranked according to their % GC content, and the highest 10% of genera **(A)** and lowest 10% of genera **(B)** are plotted to compare single vs triplicate PCR reactions.

2.2.5 References

1. Polz MF, Cavanaugh CM. Bias in Template-to-Product Ratios in Multitemplate PCR [Internet]. Vol. 64, APPLIED AND ENVIRONMENTAL MICROBIOLOGY. 1998 [cited 2018 Dec 18].

2. Wen C, Wu L, Qin Y, Van Nostrand JD, Ning D, Sun B, Xue K, Liu F, Deng Y, Liang Y, Zhou J. Evaluation of the reproducibility of amplicon sequencing with Illumina MiSeq platform. Green SJ, editor. PLoS One [Internet]. 2017 Apr 28 [cited 2018 Dec 18];12(4):e0176716.

3. Kennedy K, Hall MW, Lynch MDJ, Moreno-Hagelsieb G, Neufeld JD. Evaluating Bias of Illumina-Based Bacterial 16S rRNA Gene Profiles. 2014

4. Smith DP, Peay KG. Sequence Depth, Not PCR Replication, Improves Ecological Inference from Next Generation DNA Sequencing. Kellogg CA, editor. PLoS One [Internet]. 2014 Feb 28 [cited 2018 Dec 18];9(2):e90234.

5. Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, Schwager E, Crabtree J, Ma S, Abnet CC, Knight R, White O, Huttenhower C, Huttenhower C. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat Biotechnol [Internet]. 2017 Oct 2;35(11):1077.

6. Marotz C, Amir A, Humphrey G, Gaffney J, Gogul G, Knight R. DNA extraction for streamlined metagenomics of diverse environmental samples. Biotechniques. 2017;62(6):290–3.

7. McDonald D, Hyde ER, Debelius JW, Morton JT, Gonzalez A, Ackermann G, Aksenov AA, Behsaz B, Brennan C, Chen Y, DeRight Goldasicha L, Dorrestein PC, Dunn RR, Fahimipourg AK, Gaffney J, Gilbert JA, Gogul G, Green JL, Hugenholtz P, Humphrey G, Knight R, et al. American Gut: an Open Platform for Citizen-Science Microbiome Research. bioRxiv. 2018 Jan 1

8. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Caporaso JG, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37:848–57.

9. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Xu ZZ, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems. 2017 Apr 21;2(2):e00191-16.

10. Kelley ST, Gilbert JA. Studying the microbiology of the indoor environment. Genome Biol [Internet]. 2013 Feb 28 [cited 2018 Dec 18];14(2):202.

11. Cardona C, Lax S, Larsen P, Stephens B, Hampton-Marcell J, Edwardson CF, Henry C, Bonn B Van, Gilbert JA. Environmental Sources of Bacteria Differentially Influence Host-Associated Microbial Dynamics. 2018 [cited 2018 Dec 18];

12. McPherson MR, Wang P, Marsh EL, Mitchell RB, Schachtman DP. Isolation and Analysis of Microbial Communities in Soil, Rhizosphere, and Roots in Perennial Grass Experiments. J Vis Exp [Internet]. 2018 Jul 24 [cited 2018 Dec 18];(137).

13. Minich JJ, Humphrey G, Benitez RAS, Sanders J, Swafford A, Allen EE, Knight R. High-Throughput Miniaturized 16S rRNA Amplicon Library Preparation Reduces Costs while Preserving Microbiome Integrity. Langille MGI, editor. mSystems [Internet]. 2018 Nov 6 [cited 2018 Dec 18];3(6):e00166-18.

2.3 Improving saliva shotgun metagenomics by chemical host DNA depletion

Shotgun sequencing of microbial communities provides in-depth knowledge of the microbiome by cataloging bacterial, fungal, and viral gene content within a sample, providing an advantage over amplicon sequencing approaches that assess taxonomy but not function, and are taxonomically limited. However, mammalian DNA can dominate host-derived samples, obscuring changes in microbial populations because few DNA sequence reads are from the microbial component. We developed and optimized a novel method for enriching microbial DNA from human oral samples and compared its efficiency and potential taxonomic bias with commercially available kits.

2.3.1 Background

In the past decade sequencing costs have plummeted, and 16S rRNA gene amplicon sequencing has become a nearly ubiquitous tool used to characterize bacterial populations from a wide range of environments and host systems [1,2]. This technique has revealed that bacteria inhabit a far greater range of human body sites than previously believed, including many long presumed to be sterile (e.g. urine [3], breast milk [4], blood [5], and atherosclerotic plaque [6]). However, 16S rRNA gene amplicon sequencing has several limitations.

Taxonomic resolution is intrinsically limited in amplicon analysis and can fail to distinguish species and strains with distinct biological functions. Primer choice can affect the representation of particular clades of bacteria [7]. Eukaryotic microbes are not captured by 16S rRNA gene amplicon sequencing and require 18S rRNA gene, internal transcribed spacer, or

mitochondrial sequencing approaches; viruses are not detected by any these methods and require custom clade-specific primers.

Shotgun metagenomic sequencing overcomes these hurdles because it analyzes total DNA extracted from a sample, and does not depend on target-specific primers. For the analysis of host-derived samples, this advantage of shotgun sequencing is also its vulnerability. Because the human genome is roughly one thousand times larger than an average bacterial genome (~$3x10^9$ versus ~$3x10^6$ bp), host DNA can quickly drown out microbial reads in samples containing even a relatively small number of human cells. The proportion of human cells to microbial cells varies widely by sampling site, and consequently the percentage of shotgun sequencing reads aligning to the human genome varies widely by sampling site. For example, fecal samples from healthy controls typically yield <10% human genome-aligned reads; but human saliva, nasal cavity, skin, and vaginal samples routinely contain >90% (Fig. 1). Therefore, identifying a method to reproducibly deplete host reads for shotgun sequencing is crucial for almost all host-derived microbiome studies.

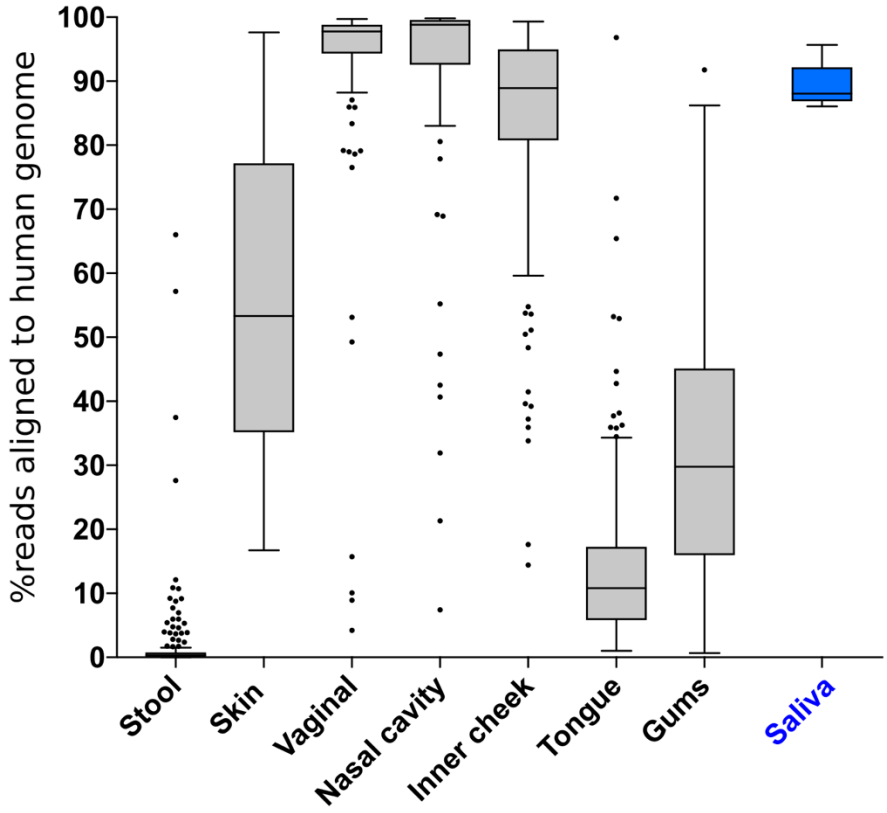

Figure 2.3.1. *Percent of shotgun metagenome sequencing reads aligning to human genome varies by sample type.* Data from the Human Microbiome Project (HMP; black) of healthy individuals demonstrates the percentage of human reads by sample type. Saliva data (blue) was collected from healthy individuals in this study. Stool n = 249, skin n= 29, vaginal n = 103, nasal cavity n = 112, inner cheek n = 175, tongue n = 208, gums n = 189, saliva n = 24 (this study).

Current approaches to deplete host reads can be divided into two major groups: those that act prior to DNA extraction (pre-extraction), and those that act on DNA after extraction (post-extraction). Pre-extraction approaches generally follow a two step procedure. The first step is to selectively lyse mammalian cells, which is easy because the mammalian cell membrane is more fragile than most microbial membranes/cell walls. The second step removes exposed DNA enzymatically, leaving only the intact microbial cells for downstream analysis. These kits have improved microbial sequencing yield in a variety of sample types, including bronchoalveolar lavage fluid [8], blood [9], and sonicate fluid from prosthetic joint infections [10]. However, the

multiple wash steps required limit the potential of low biomass samples to be successfully treated [11]. Furthermore, loss of bacterial DNA and a potential bias toward Gram-positive taxa has been reported [12].

Post-extraction separation approaches avoid some of these hurdles and thus pose an attractive alternative. One approach targets methylated nucleotides [13], which are typically more frequent in eukaryotic genomes. However, a bias against microbes with AT-rich genomes has been reported [14] and therefore this method is not suitable for eukaryotic microorganisms with methylation patterns similar to the host. Another approach is targeting host-specific sequences for hybridization-based depletion with CRISPR/Cas9. This method has been successfully employed for highly repetitive rRNA sequences [15], but is not easily adapted to depletion of sequences at the genome scale.

To overcome the disadvantages associated with each of the currently available host-depletion methods, we optimized a cost-effective technique with minimal sample processing and hands-on time. Similar to other pre-extraction methods, it starts with selective mammalian cell lysis, but instead of enzymatic digestion of exposed DNA we employed propidium monoazide (PMA). PMA has been used extensively over the past decade for detection of live/dead cells [16]. Similar to propidium iodide, PMA is a cell membrane impermeable DNA intercalator. Upon exposure to visible light, the azide group of the PMA molecule is photolytically cleaved and undergoes a C-H insertion reaction to form a covalent bond with DNA [17]. It is thought that this reaction fragments the DNA, effectively eliminating any exposed DNA from downstream analysis [17,18]. Any excess PMA in the sample reacts with water and becomes inert. We induced selective osmotic lysis of mammalian cells by resuspension in pure water followed by treatment with PMA

(lyPMA). This method requires less than five minutes hands on time and involves no special laboratory equipment.

To evaluate the efficiency and potential bias of the lyPMA treatment, we compared this protocol to raw samples and four alternative methods used for host depletion: 5 µm filtration (Fil), QIAamp DNA Microbiome Kit (QIA), MolYsis™ Basic (Mol), and NEBNext® Microbiome DNA Enrichment Kit (NEB). We chose saliva samples to compare methods of host depletion because it is easy to collect, has enough biomass to be divided into multiple groups per sample, and consistently has a high (~90%) percentage of human DNA in shotgun metagenomic sequencing (Fig. 1). The efficiency of host depletion and the effect on microbial community was assessed.

2.3.2 Results

Differential cell size-based approaches to host DNA depletion

One of the most obvious differences between mammalian and microbial cells is their size. Our preliminary attempts to reduce host DNA therefore focused on separating cells according to size. Because buccal epithelial cells are on average 50 μm wide, whereas a typical bacterial coccus is ~1 μm, we passed saliva samples across a 5 μm filter and analyzed the filtrate and residue compared to the raw sample. We designed a qPCR assay to evaluate the percentage of host DNA relative to untreated sample using a human-specific primer against the PTGER2 gene [19]. No significant difference across any of these three partitions was observed (Fig. S1A). To exclude the potential of host cell shearing during filtration, we next tried differential centrifugation to enrich for microbial DNA. First, a short, slow centrifugation (30 sec at 2,500 g) of human saliva was performed to pellet large cells [20], then the supernatant was washed with a longer, faster centrifugation (8 min at 10,000 g) to pellet all remaining cells. No significant difference in percentage of human DNA at any of these steps compared to the original raw sample was observed (Fig. S1B). Lastly, we attempted to take advantage of differences in forward and backward scatter (which canonically correlates to event size and density, respectively) using flow cytometry to separate microbial from human cells. Although three distinct groups of varying size were clearly observed (Fig. S1C), there was no significant difference in percentage of human DNA among the sorting gates compared to the raw sample (Fig. S1D).

These preliminary attempts to separate mammalian from microbial cells based on cell size were unsuccessful in reducing the amount of host DNA. We hypothesized that there must be a significant amount of extracellular host DNA that is not separated by size-based approaches. Indeed, DNAse treatment of saliva samples after a short, slow centrifugation significantly reduced

the percentage of human DNA (Fig. S1E). However, enzymatic treatment can be expensive, sensitive, and, because it must be processed on fresh samples, difficult to scale. As an alternative to enzymatic digestion of extracellular DNA, we next tested the ability of PMA to remove host DNA.

Optimization of lyPMA for host DNA removal

To optimize the lyPMA protocol, we compared different methods of selective mammalian cell lysis and multiple concentrations of PMA. Quantitative polymerase chain reaction (qPCR) analysis of the human-specific PTGER2 gene revealed that, compared to untreated controls, PMA treatment reduced the percentage of human DNA following selective mammalian cell lysis by sonication (25.6%) and osmotic lysis with H2O (1.7%), and following mammalian cell removal by differential centrifugation (1.4%) (Fig. S2A). Interestingly, PMA treatment of raw saliva sample (without an initial selective lysis step) also reduced the percentage of human DNA (16.8%), suggesting that the majority of human DNA in saliva is already exposed. We also evaluated the effect of PMA concentration (1 μM, 10 μM, and 50 μM) on the reduction of extracellular host DNA following differential centrifugation. Treatment with 10 μM PMA was the optimal concentration to achieve host DNA reduction without compromising microbial DNA recovery, although the results were not highly sensitive to PMA concentration (Fig. S2B). The relative percentage of bacterial DNA was also evaluated by qPCR to ensure specific removal of human DNA. Compared to raw saliva samples, osmotic lysis (7.82 fold) and 10 μm PMA concentration (13.4 fold) had the greatest increase in the proportion of bacterial DNA (Fig. S2C-D). We thus used osmotic lysis followed by 10 μM PMA treatment for comparison of the lyPMA approach to commercially available alternatives.

Efficiency of host depletion across microbial enrichment methods

To compare lyPMA to other methods of host DNA depletion, saliva samples were collected from 8 healthy participants (4 mL each). Each sample was homogenized and divided into 18 separate 200 µl aliquots. Triplicate aliquots were processed in parallel for each of the six methods (i.e. untreated (raw) samples, Fil, NEB, Mol, QIA, and lyPMA, see Methods). DNA was extracted from all samples and shotgun DNA sequencing libraries were prepared in parallel. The concentration of DNA following host depletion was significantly lower in all five methods compared to raw samples (Fig. S3A). Samples for sequencing were pooled such that the raw samples had twice as many reads compared to the host-depleted samples, with the assumption that we could thereby increase the number of microbial reads in the raw samples to better assess potential taxonomic biases (Fig. S3B). Samples with less than 50,000 quality filtered microbial reads (n=7) were excluded from downstream analysis, leaving 137 samples to evaluate efficiency of host depletion and microbial community effect.

The percentage of shotgun metagenomic sequencing reads mapping to the human genome in each sample was evaluated using Bowtie 2. The average percentage of human reads in the raw samples (89.29 ± 0.61%) was no different than in samples filtered across a 5 µm filter (89.69 ± 0.84%), or samples treated with the NEB kit (90.83 ± 0.77%) (Fig. 2). Treatment with Mol (62.88 ± 3.46%), QIA (29.17 ± 5.04), and lyPMA (8.53 ± 2.08%) all significantly depleted host reads compared to the raw samples (One-way ANOVA with Tukey's multiple comparison $p<0.0001$). These three methods were all significantly different from each other, with lyPMA performing the best followed by QIA and then Mol ($p<0.0001$).

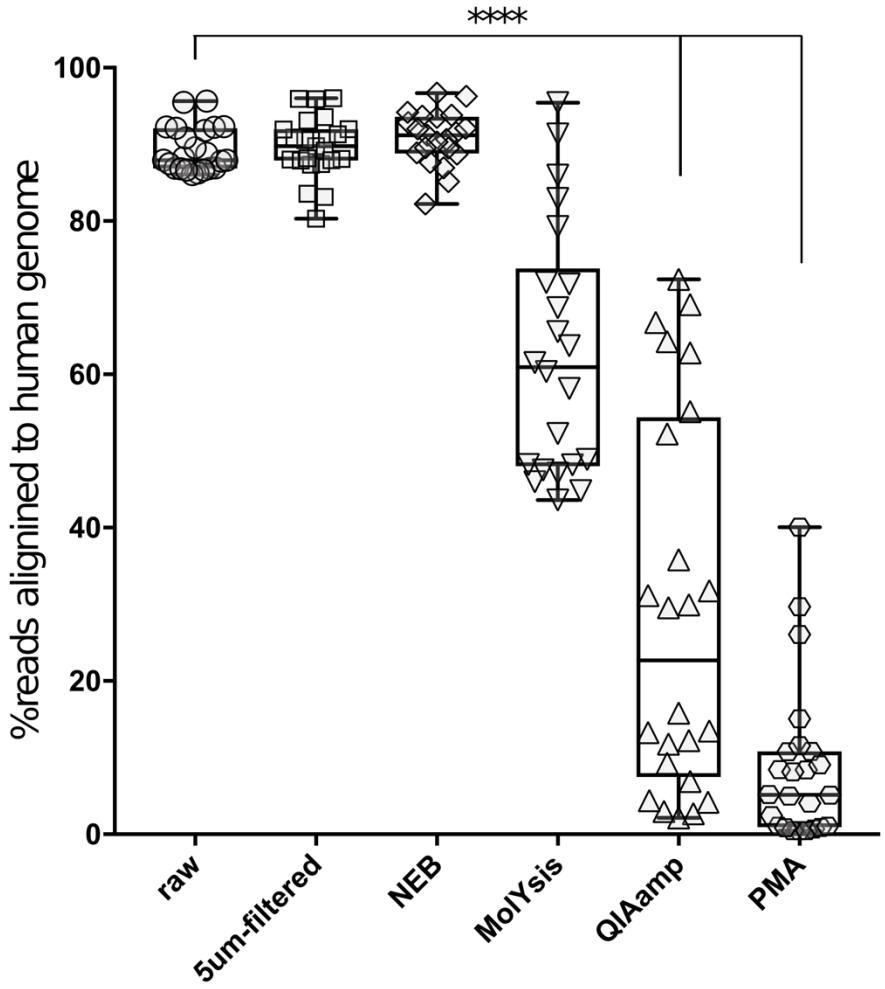

Figure 2.3.2. *Host DNA depletion in saliva reduces the percentage of sequencing reads aligning to the human genome.* Saliva was collected from eight individuals and divided into triplicate aliquots for each of the processing methods. The fraction of quality filtered shotgun sequencing reads mapping to the human genome was assessed with Bowtie 2. One-way ANOVA with Tukey's multiple comparison correction, significance $p<0.0001$.

Effect on microbial community composition caused by host depletion methods

Each participant had a distinct pattern of microbial genera that altered slightly upon host depletion (Fig. S4). In principal coordinates analyses (PCoA) [21], samples cluster by participant and not method of host depletion (Fig. 3A-B), suggesting that differences in the microbial community were driven more by biological differences among participants rather than technical effects from any of the host depletion methods. Indeed, Bray-Curtis (BC) dissimilarities between

samples from different participants processed with the same host depletion method were significantly greater than dissimilarities from the same participant processed with different host depletion methods (Fig. 3C). This held true across other beta-diversity metrics including phylogeny-informed weighted and unweighted UniFrac (Table 1). Furthermore, relative abundances of microbial taxa within participants among host depletion methods were tightly correlated with few obvious outliers and Spearman's rank correlation coefficient ≥ 0.75 (Fig. S5).

Table 2.3.1. *Adonis statistical assessment of beta diversity driven by participant or host DNA depletion method*

Beta-diversity metric	variable	Degrees of Freedom	R2	F.model	*p* value
unweighted unifrac	method	5	0.092	6.345	0.001
	participant	7	0.548	26.932	0.001
weighted unifrac	method	5	0.118	12.331	0.001
	participant	7	0.644	47.947	0.001
Bray-curtis	method	5	0.149	14.444	0.001
	participant	7	0.594	41.019	0.001
Binary Jaccard	method	5	0.168	9.537	0.001
	participant	7	0.396	16.059	0.001

To evaluate whether differences in read depth affected these results, we subsampled 50,000 non-human reads from each sample and found similar results; namely that the participant explained more of the variability in microbial taxa than method of host DNA depletion (Adonis of Bray-Curtis distance R2 = 0.169 by method, 0.556 by participant, F.model = 15.152 by method, 36.614 by participant; *p* value <0.001 for all).

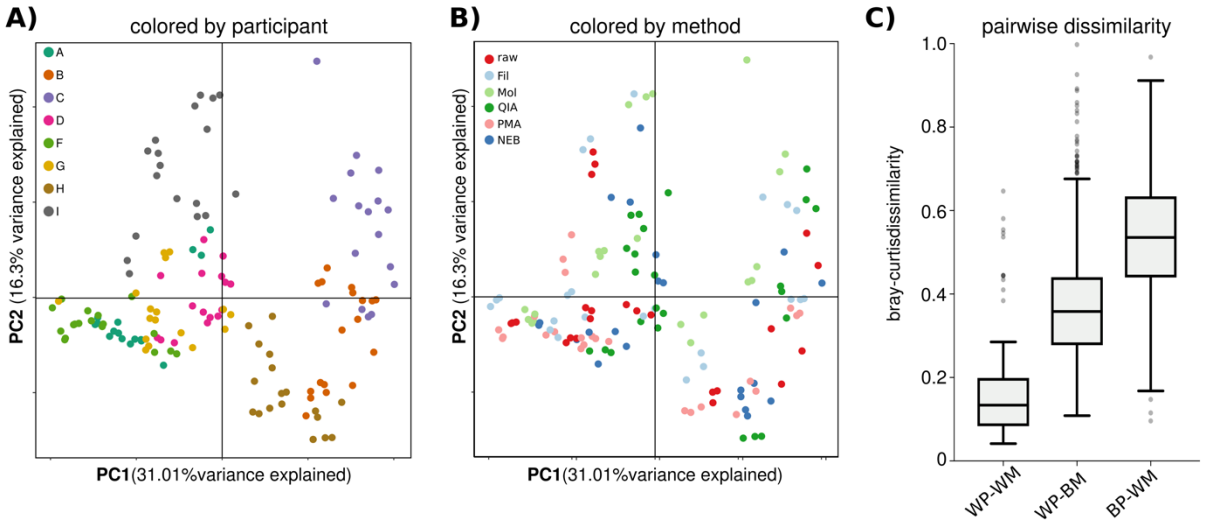

Figure 2.3.3. *Differences in saliva microbiome driven by participant and not method of host depletion.* Microbial reads cluster by participant **A)** and not method of host depletion **B)** in PCoA space using Bray-Curtis distance. **C)** Pairwise Bray-Curtis dissimilarities: within participant, within method (WP-WM); within participant, between methods (WP-BM); between participants, within methods (BP-WM). Each category is statistically significantly different from each other group (Kruskal-Wallis with Benjamini and Yekutieli FDR correction $p<0.0001$).

However, BC dissimilarities among host-depleted samples from the same participant were significantly higher than noise from technical replication (within raw triplicate samples), indicating that there is a significant effect of host depletion on microbial community composition (Fig. 3C). We then compared the BC dissimilarities between each method of host depletion and the corresponding raw samples (Fig. 4). Each of the five treatments had significantly greater BC dissimilarity than technical variation among raw replicates (0.115 ± 0.009; Kruskal-Wallis with Benjamini and Yekutieli FDR correction $p<0.05$). However, lyPMA (0.273 ± 0.011) and Fil (0.226 ± 0.009) were significantly more similar to raw samples than NEB (0.333 ± 0.010), Mol (0.321 ± 0.015), and QIA (0.342 ± 0.008); Kruskal-Wallis with Benjamini and Yekutieli FDR correction $p<0.05$). There was no statistical difference observed among NEB, Mol, and QIA distance from corresponding raw samples. To look for bacteria affected by host DNA depletion, we performed a

pairwise comparison of the relative abundance of each taxon in raw versus host DNA depletion method for each individual using t-tests with Benjamini, Krieger and Yekutieli false discovery rate correction at 1%. No taxa were identified to be consistently differentially abundant across the host DNA depletion methods.

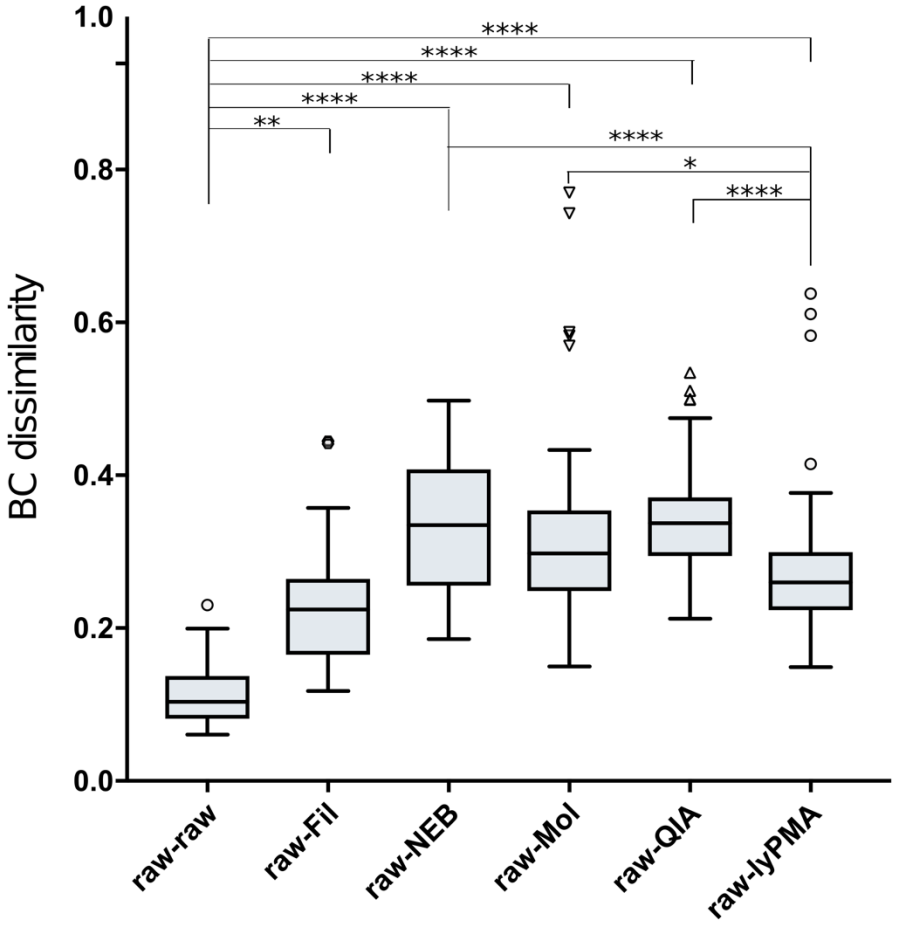

Figure 2.3.4. *Bray-Curtis dissimilarity between host depleted and raw sample from same participant.* The pairwise Bray-Curtis dissimilarity value was calculated between each sample with every other sample in this study. The dissimilarity values between each sample and the matched participant raw sample is presented here. Statistical significance calculated with Kruskal-Wallis with Benjamini and Yekutieli FDR correction $p<0.05$. raw-raw n=22, raw-Fil n=66, raw-NEB n=63, raw-Mol n=63, raw-QIA n=69, raw-lyPMA n=69.

Evaluation of lyPMA treatment on frozen saliva samples

Microbiome sampling often requires samples to be frozen and preserved for downstream processing. We therefore tested the effectiveness of lyPMA on previously frozen samples. Saliva samples from three participants were divided into 200 µl aliquots and either immediately stored at -20°C or cryopreserved by mixing with 20% glycerol prior to freezing (Fig. S6). After three days samples were thawed and replicate frozen and cryopreserved samples were treated with lyPMA. Similar to the freshly processed samples, the majority of reads from the untreated samples aligned to the human genome (84.73 ± 2.56%). lyPMA samples that were cryopreserved with glycerol had a similar reduction in host-aligned reads to the freshly processed samples (7.18 ± 3.09%). Without cryopreservation, lyPMA was less efficient and more variable (53.78 ± 27.43%). The BC dissimilarity value was similar for technical replicates of raw samples (0.146 ± 0.004) and raw versus matched cryopreserved lyPMA (0.276 ± 0.071) but was higher for raw versus matched non-cryopreserved lyPMA (0.348 ± 0.002).

2.3.3 Discussion

We compared the efficiency of five methods of host DNA depletion on human saliva for shotgun metagenomic sequencing as outlined in Figure 5. Although filtering saliva across a 5 µm filter excludes intact host cells, no difference was observed in the percentage of host-aligned reads in DNA extracted from the filtrate. This is likely due to the high amount of extracellular DNA in saliva and explains why preliminary experiments based on separating microbial from host cells based off size (i.e. 5µm-filtration and flow cytometry) were unsuccessful.

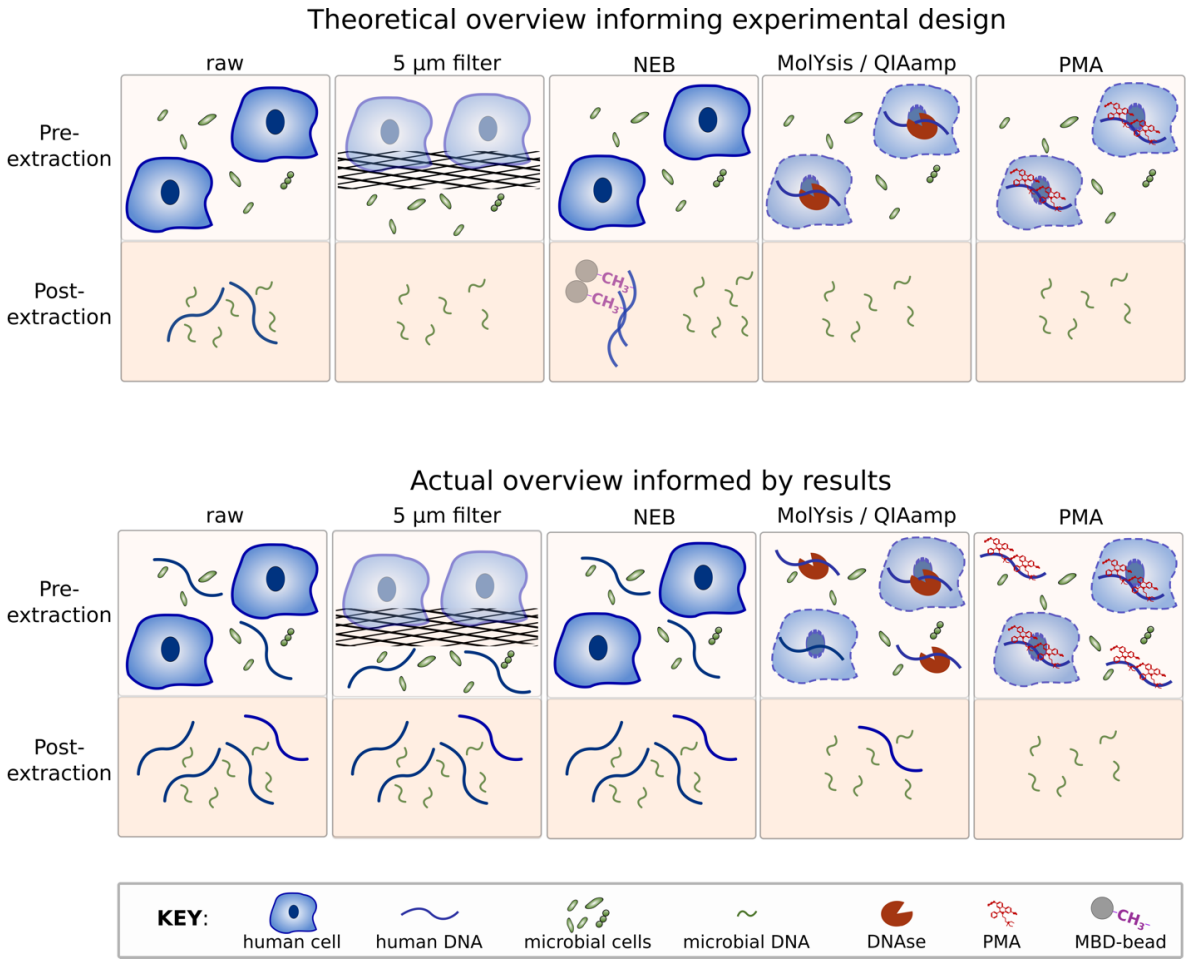

Figure 2.3.5. *Experimental overview.* A graphical summary of the experimental design and results.

In our hands, immunoprecipitation of methylated eukaryotic DNA was unsuccessful at reducing the percentage of host-aligned sequences, as evidenced by our post-extraction processing of DNA with the NEBNext enrichment kit. It is important to note that this protocol recommends an input of high molecular weight gDNA (>15kb fragments) and our input gDNA peaked at ~10kb. To achieve maximal efficiency, samples should be extracted with phenol chloroform followed by isolation of appropriately sized DNA from a low melt agarose gel [13]. However, we extracted all samples in parallel using a high-throughput DNA extraction pipeline [22] to reduce any confounding variables from differing DNA extraction methods.

Selective lysis of mammalian cells followed by removal of exposed DNA was a consistently effective method of reducing host-aligned sequencing reads. MolYsis, QIAamp, and lyPMA treatments all significantly improved microbial yield, with lyPMA outperforming alternative treatments. It is possible that increasing the enzyme concentration of the MolYsis or QIAamp kits would further reduce the percentage of host-aligned reads. Regardless, lyPMA treatment has an advantage over enzymatic degradation in that it requires fewer washing steps, less hands-on time, and has a fraction of the reagents costs compared to commercial alternatives (PMA ~0.15$/sample; QIAmp ~10$/sample; MolYsis ~7$sample; NEB ~30$/sample).

Any method to enrich microbial sequences will invariably have some effect on the microbial community. However, the Bray-Curtis distance between lyPMA and matched raw samples was significantly smaller than for every other host depletion method. This suggests lyPMA treatment can be used to reduce the percentage of reads in saliva samples while minimizing biases in representation in the microbial community. Importantly, the distortions induced were less than the differences among individuals, strongly suggesting that the ability to read out clinically significant microbiome states would be preserved. Furthermore, the differences observed may

actually be biologically relevant. PMA treatment only detects intact, or live, microbial cells, which can have a statistically significant impact on biological interpretation, as has been shown in recent studies on a broad range of topics including the soil microbiome [23], spaceship clean rooms [24], cystic fibrosis patient samples [25], food safety [26,27], and saliva [28].

The lyPMA method has been optimized for saliva; however, saliva is only one of many sample types where microbial analysis is hampered by a large amount of host DNA. Expanding this technique to different sample types will require tailoring the method to account for selective lysis, ideal PMA concentration, and optimal temperature and duration of light exposure.

2.3.4 Methods

qPCR evaluation of human and microbial gDNA

A total of 1 ng purified DNA from human saliva was used as a template to amplify the human-specific primer PTGER2: hPTGER2f (5'- GCTGCTTCTCATTGTCTCGG -3'), PTGER2r (5'- GCCAGGAGAATGAGGTGGTC -3') [19], and the 16S rRNA gene: Bakt-805R (5'- CCTACGGGNGGCWGCAG -3'), Bakt_341f (5'- GACTACHVGGGTATCTAATCC -3') [7]. All reactions were performed in triplicate. The final qPCR reaction volume totaled 10 µl containing 5 µl KAPA HiFi HotStart ReadyMix (2X), 1 ng DNA, 0.5 µM forward and reverse primer, 1x SYBR green (Life Technologies), and the remainder water. The qPCR amplification was carried out over 35 cycles (20 s at 98°C, 15 s at 60°C, 35 s at 72°C) with an initial 3 min hot start at 95°C and a final extension step (1 min at 72 °C). In each experiment a standard curve was included comprising known ratios (100:0, 25:75, 50:50, 75:25, and 0:100) of human gDNA (extracted HEK293T cells) and bacterial gDNA (extracted from *E. coli*) in order to extrapolate the

percentage of human versus microbial DNA. All samples were run in triplicate reactions and the error bars represent standard deviation among these technical replicates.

Cell size-based attempts at host DNA removal

Flow cytometry: approximately 20 mg of frozen fecal sample was homogenized with 1 mL sterile phosphate buffered saline (PBS) by vortexing at maximum speed for 10 minutes. The sample was centrifuged for 3 min at 2,000 g, diluted with an additional 2 mL PBS and filtered across a 35-µm filter. Triplicate 50 µl aliquots were stored for analysis of the unsorted sample, and the remaining sample was stained with a final concentration of 2x SYBR green I in the dark for 15 min. The sample was diluted 1:10 in sterile PBS and run on a Sony SH800 FACS using a 100-µm nozzle with threshold set on the forward scatter detector at 1%. Events with SYBR-specific fluorescence emission (520 nm) stronger than vehicle control were selected for analysis. Of these SYBR positive events, 3 distinct populations were gated in the forward and backward scatter axes (representing event size and density, respectively). 100,000 events per gate were sorted and centrifuged at 10,000 g for 8 min to pellet cells. DNA was extracted from the cell pellets as detailed below.

Sonication: 200 µl saliva samples were sonicated in an ice bath for 15 min at 40 hz (Branson 2510, Marshall Scientific), which was previously shown to separate microbial biofilms without lysing bacteria cells [29] and then treated with PMA as detailed below.

DNAse treatment: raw saliva samples were centrifuged for 5 min at 5,500 g and the pellet was resuspended in 100 µl 1x TURBO™ DNase buffer. Next, 3 units of TURBO™ DNAse I was added and the samples were incubated at 37°C for 20 min. The sample was washed with 500 µl sterile 1x PBS containing 0.1 µM EDTA to inhibit the DNAse and DNA was extracted from the pellet as detailed below.

Comparative Study Details

Sample collection and host depletion

Volunteers were asked to refrain from eating or drinking for one hour prior to sample collection. A total of 4 ml of unstimulated saliva was collected from 8 volunteers into sterile 15-ml conical tubes. The sample was vortexed for 30 seconds and 200-µl aliquots were made for each method in triplicate (18 replicate samples in total per individual). The samples were immediately processed in parallel as described below.

Raw (untreated): samples were kept on ice while the other samples were processed, then stored at -20°C.

5 µm filtration (Fil): 200 µl of sterile, 1x PBS was added to each sample and vortexed for 15 seconds. The diluted sample was run across a pluriStrainer® 5-µm filter (PluriSelect) by inducing low pressure with a 10-ml syringe on the Connector Ring. The effluent was retained and frozen at -20°C.

MolYsis™ Basic kit (Mol)*:* samples were processed according to the manufacturer's instructions. After removal of MolDNase A, samples were frozen at -20°C.

Qiagen QIAamp DNA Microbiome enrichment kit (QIA): samples were processed according to the manufacturer's instructions. After proteinase K treatment samples were frozen at -20°C.

PMA treatment (lyPMA): 200-µl unstimulated saliva aliquots were centrifuged at 10,000 g for 8 min. The supernatant was discarded and the cell pellet was resuspended in 200 µl sterile H_2O by pipetting and a brief vortex then left at room temperature for 5 min to osmotically lyse mammalian cells. A final concentration of 10 µm PMA (Biotium) was added (10 µl of 0.2 mM PMA solution added to 200 µl sample) and the sample was briefly vortexed, then incubated in the

dark at room temperature for 5 min. Samples were then laid horizontally on ice <20 cm [23] from a standard, bench top fluorescent light bulb (Philips F28T5/835 ALTO 40PK) for 25 min, with brief centrifugation and rotation every ~5 min. After exposure, samples were frozen at -20°C.

NEBNext Microbiome DNA Enrichment kit (NEB): samples were treated as raw throughout sample collection and DNA extraction, then processed according to the manufacturer's instructions.

DNA extraction

Frozen samples were thawed and transferred into 96-well plates containing garnet beads and extracted using Qiagen PowerSoil DNA kit adapted for magnetic bead purification as previously described [22]. DNA was eluted in 100 µl Qiagen elution buffer.

Library generation and sequencing

All data presented combines two independent experiments performed identically, with each experiment containing replicate saliva samples processed as described above for four individuals each. Extracted DNA was quantified via Qubit™ dsDNA HS Assay (ThermoFisher Scientific) and 1 ng of input DNA was used in a 1:10 miniaturized Kapa HyperPlus protocol. For samples with less than 1 ng DNA, a maximum volume of 3.5 µl input was used. Library concentration was determined with triplicate readings of the Kapa Illumina Library Quantification Kit. 20 fmol of raw sample libraries and 10 fmol of host-depleted libraries were pooled and size selected for fragments between 300 and 800 bp on the Sage Science PippinHT to exclude primer dimers. The pooled library was sequenced as a paired-end 150-cycle run on an Illumina HiSeq2500 v2 in Rapid Run mode at the UCSD IGM Genomics Center.

Sequencing data analysis

Demultiplexed sequences were processed using an in-house modular workflow employing Snakemake [30] (https://github.com/tanaes/snakemake_assemble/commits/mash, commit 1c393f4). First, reads were trimmed and quality filtered using Atropos v 1.1.5, a fork of Cutadapt [31]. Reads aligning to the host genome (GRCh38.p7) were identified using Bowtie 2 v2.3.0 [32] with parameters set by the flag --very-sensitive-local. A total of seven samples with fewer than 50,000 quality filtered non-human reads were excluded from downstream analysis. The host-filtered microbial reads from the remaining 137 samples were profiled using MetaPhlAn v2.0 [33] with standard parameters. The MetaPhlAn taxonomic output matrix was filtered to represent only the relative abundance of the most specific taxonomic level. Taxa only identified in one out of the 137 samples were excluded from analysis, resulting in 175 taxa. This filtered matrix was used for Bray-Curtis and Binary Jaccard beta diversity analysis using QIIME. For phylogenetic analyses including UniFrac [34], a tree was created using the MetaPhlAn2 taxonomy, with internal branches assigned a length of 1. Because some taxa could not be assigned to the tips of the tree, internal nodes were added as tips assigned a length of 0, allowing these taxa to contribute to the analysis.

2.3.5 Conclusion

Osmotic lysis in distilled water followed by treatment with PMA (lyPMA) is a novel method to significantly reduce the percentage of human DNA in shotgun metagenomic sequencing. The method requires only standard laboratory equipment and is suitable for any DNA extraction technique. lyPMA increases microbial reads in human saliva samples by an order of magnitude. Given a low consumables cost of around 15 cents per sample, lyPMA can therefore reduce the sequencing cost by an order of magnitude.

2.3.6 Declarations

Ethics approval and consent to participate

All human subjects who participated in this project were consented under the UCSD HRPP approved protocol #150275 Explaining Variability in Human-associated microbial Communities PI Rob Knight (Federalwide Assurance number, FWA00004495).

Availability of data and material

The dataset supporting the conclusions of this article (all human-filtered microbial sequencing data) is publicly available at qiita.ucsd.edu under Qiita ID 11541 and the European Nucleotide Archive (ENA) under Study ID PRJEB24090, accession ERP105893.

Competing interests

The authors declare that they have no competing interests.

Funding

Research reported in this publication was supported in part by the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health under Award Number AR071731. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This research was also supported in part by the Center for Microbiome Innovation, UC San Diego, and a fellowship award from BASF (Germany) awarded to CM.

Authors' contributions

CM, LZ, RK, and KZ designed the experiment. CM performed the experiment and wrote the manuscript with input from all authors. CM, JS, CZ and LZ analyzed, interpreted, and visualized the data. All authors read and approved the final manuscript.

We would like to acknowledge Dr. Paul Carini for his advice in optimizing the PMA protocol used in this manuscript, Dr. Anup Mahurkar for providing HMP metagenomic data, and Serene Sun for statistical analysis advice.

Chapter II, part 3, in full, is a reprint of previously published material: Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K. *Improving saliva shotgun metagenomics by chemical host DNA depletion*. Microbiome. 2018 Dec;6(1):42. I was the primary investigator and author of this paper. The co-authors listed above supervised or provided support for the research and have given permission for the inclusion of the work in this dissertation.

2.3.7 Supplemental Figures

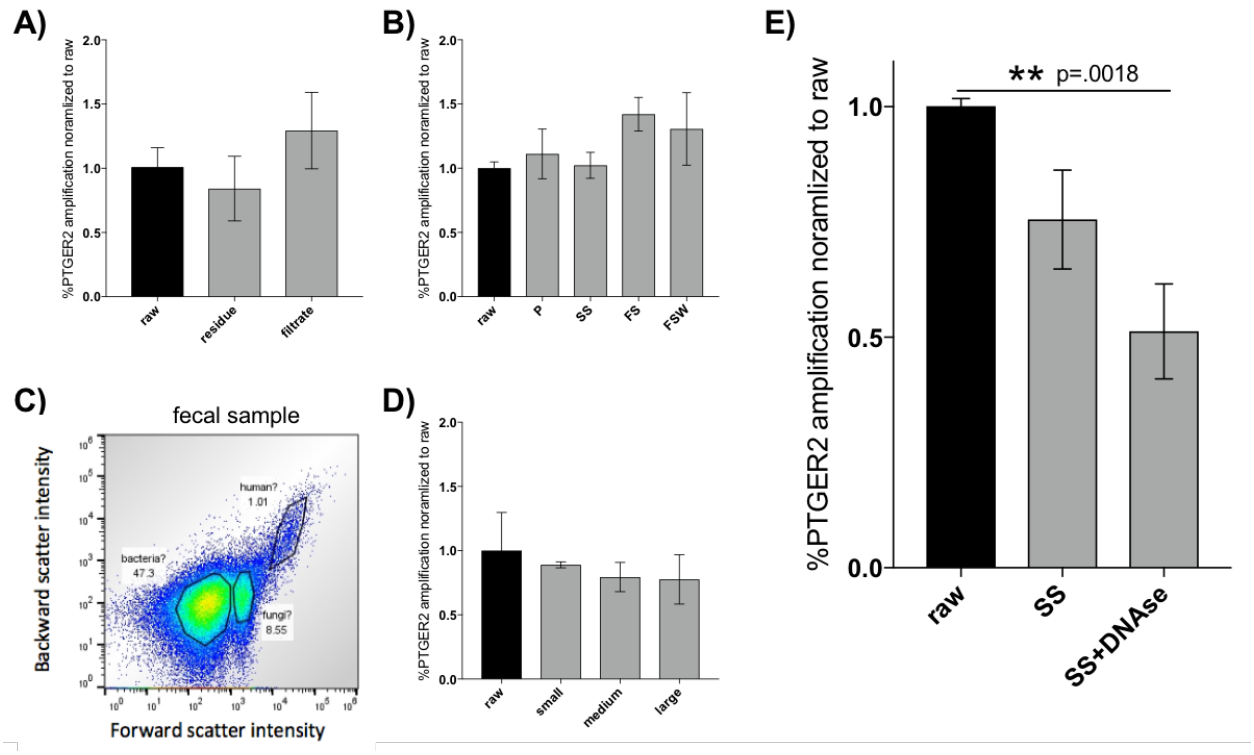

Figure 2.3.S1. *Physical approaches to separate human from microbial cells does not reduce percentage human DNA*. Unless otherwise stated, evaluation of size-driven host DNA depletion methods was performed by qPCR analysis of the human-specific PTGER2 gene normalized to raw sample. **A)** Raw saliva was passed across a 5-µm filter and the original sample (raw), residue left on top of the filter (res) and filtrate (fil) were compared. **B)** The pellet of a raw saliva sample after a 30 sec centrifugation at 2,500 g (P), its supernatant (SS), the SS after pelleting all cells at 10,000 g for 8 min (FS), and the FS pellet washed with 1x PBS (FSW) were compared. **C)** Distinct populations of small, medium (med), and large events by flow cytometry of a human fecal sample. **D)** Percentage of human DNA by shallow shotgun sequencing normalized to raw sample of distinct FACS populations from **C**. **E)** The SS of a raw saliva sample after treatment with DNAse. Significance test ordinary one-way ANOVA with Dunnett's multiple comparisons test $p < 0.01$.

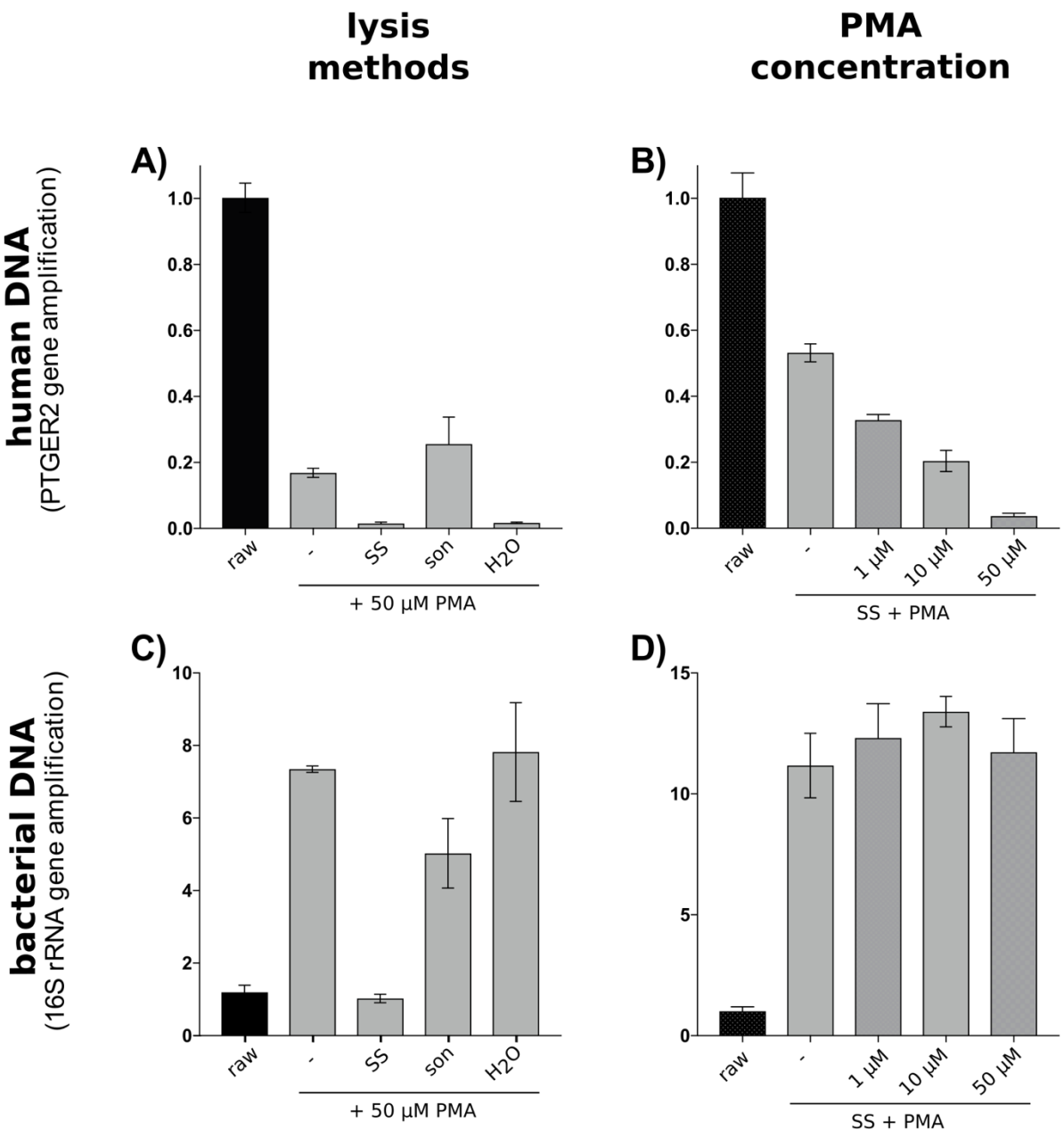

Figure 2.3.S2. *Optimization of lyPMA conditions for human DNA depletion. qPCR* analysis of the relative abundance of the human-specific PTGER2 gene normalized to raw saliva across methods of selective mammalian cell lysis **A**) and PMA concentration **B**). qPCR analysis of the fold change of the bacteria-specific 16S rRNA gene normalized to raw saliva across methods of selective mammalian cell lysis **C**) and PMA concentration **D**). SS = slow centrifugation (30 sec at 2,500g), son = sonication (15 min at 60 Hz), H_2O = osmotic lysis with pure water.

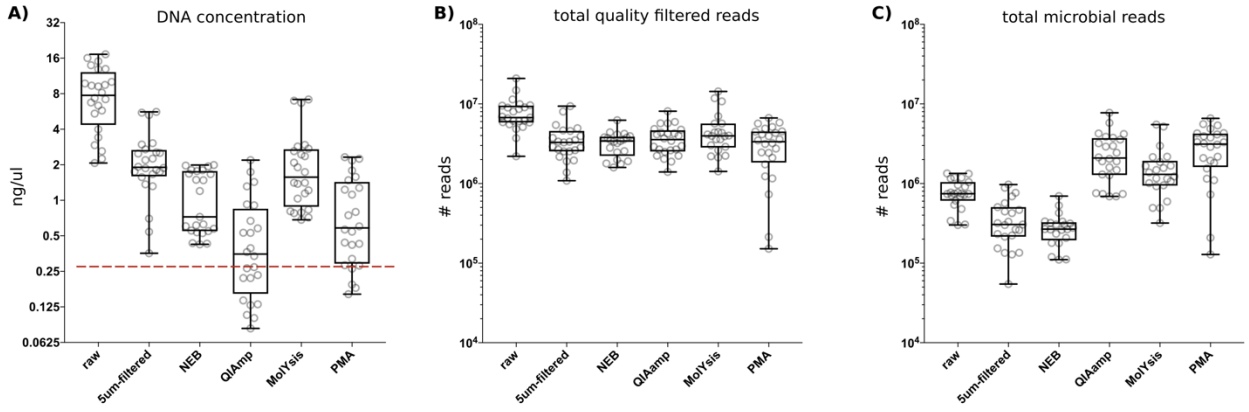

Figure 2.3.S3. *Quality control information.* **A)** DNA quantification pre-library-prep, but post-host-DNA-depletion. The red line indicates the concentration necessary to obtain 1 ng DNA input for library preparation given the volume limitations. **B)** Total number of quality filtered reads by processing method. Libraries were normalized to obtain twice as many reads for the raw samples compared to host depleted samples. **C)** Total number non-human reads after filtering using Bowtie 2.

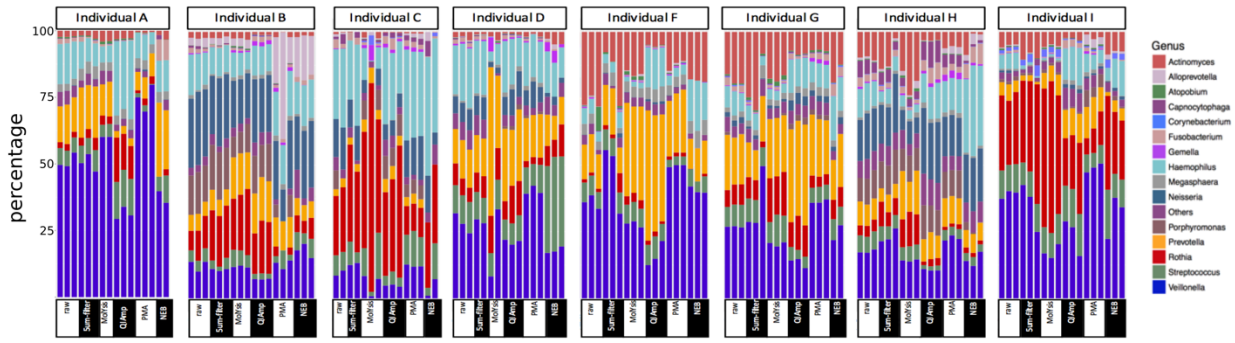

Figure 2.3.S4. *Relative abundance of the top 15 most abundant genera assigned by MetaPhlAn2 across individual and host depletion method.*

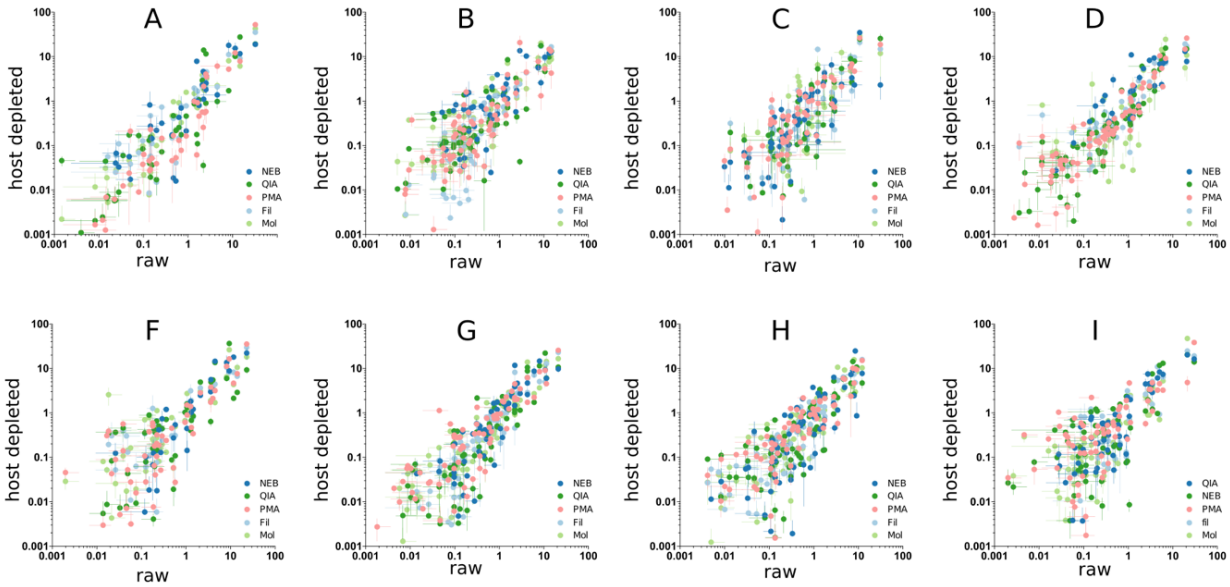

Figure 2.3.S5. *Relative taxon abundance correlation between raw and host-depleted samples.* Each plot represents data from a single participant. The x-axis represents relative abundance in the raw sample and the y-axis represents relative abundance in the corresponding host depleted sample where each dot represents a distinct taxon. Error bars represent SEM across triplicate samples. The correlation values averaged across individuals for each method were not statistically different from each other (average Spearman's rank correlation coefficient ± standard deviation: Fil = 0.789 ± 0.09, NEB = 0.75 ± 0.13, Mol = 0.82 ± 0.08, QIA = 0.83 ± 0.05, PMA = 0.82 ± 0.08)

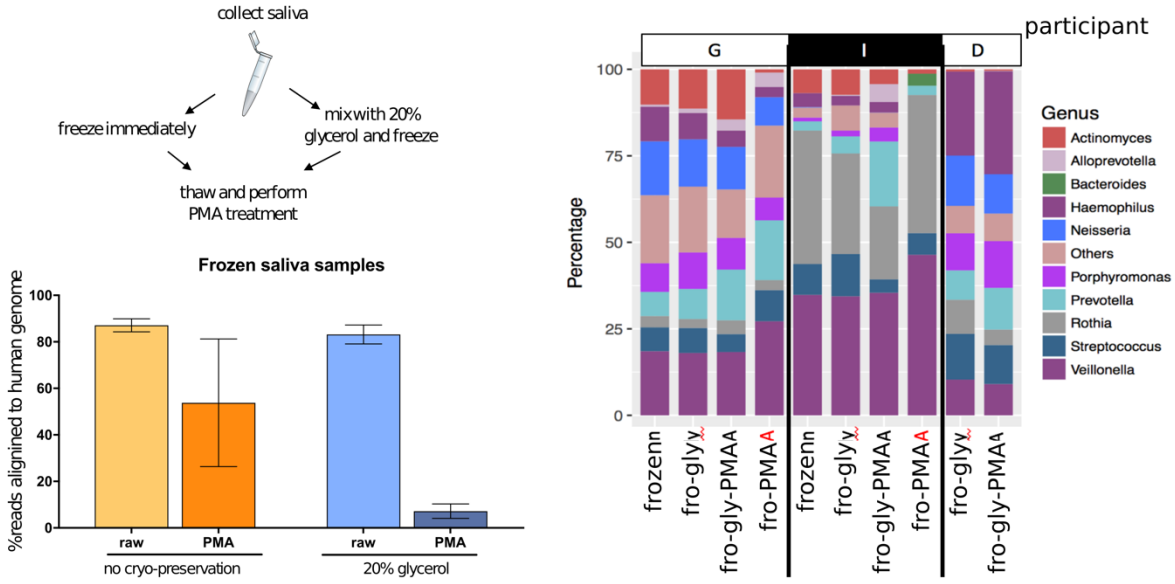

Figure 2.3.S6. *Host depletion via PMA treatment is possible for cryo-preserved samples.* Raw saliva samples were aliquoted and either frozen immediately at -20°C or mixed with a final concentration of 20% glycerol for cryopreservation. The percentage of human reads was assessed by Bowtie2, and the top 15 most abundant genera was assessed by MetaPhlAn2.

2.3.8 Acknowledgements

Chapter II, part 3, in full, is a reprint of previously published material: Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K. *Improving saliva shotgun metagenomics by chemical host DNA depletion*. Microbiome. 2018 Dec;6(1):42. I was the primary investigator and author of this paper. The co-authors listed above supervised or provided support for the research and have given permission for the inclusion of the work in this dissertation.

2.3.9 References

1. Zuñiga C, Zaramela L, Zengler K. Elucidation of complexity and prediction of interactions in microbial communities. Microb. Biotechnol. 2017;10:1500–22.

2. Knight R, Callewaert C, Marotz C, Hyde ER, Debelius JW, Mcdonald D, et al. The Microbiome and Human Biology. Annu. Rev. Genomics Hum. Genet. 2017;183:65–86.

3. Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ, Mueller ER, et al. Urine is not sterile: use of enhanced urine culture techniques to detect resident bacterial flora in the adult female bladder. J. Clin. Microbiol. 2014;52:871–76.

4. Martín R, Heilig HGHJ, Zoetendal EG, Jiménez E, Fernández L, Smidt H, et al. Cultivation-independent assessment of the bacterial diversity of breast milk among healthy women. Res. Microbiol. 2007;158:31–7.

5. Potgieter M, Bester J, Kell DB, Pretorius E. The dormant blood microbiome in chronic, inflammatory diseases. FEMS Microbiol. Rev. 2015;39:567–91.

6. Koren O, Spor A, Felin J, Fak F, Stombaugh J, Tremaroli V, et al. Human oral, gut, and plaque micro-biota in patients with atherosclerosis. Proc. Natl. Acad. Sci. 2011;108:4592–8.

7. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity stud-ies. Nucleic Acids Res. 2013;41:e1-e1.

8. Wen Y, Xiao F, Wang C, Wang Z. The impact of different methods of DNA extraction on microbial community measures of BALF samples based on metagenomic data. Am. J. Transl. Res. 2016;8:1412–25.

9. Hansen WLJ, Bruggeman CA, Wolffs PFG. Pre-analytical sample treatment and DNA extraction pro-tocols for the detection of bacterial pathogens from whole blood. Methods Mol. Biol. 2013;943:81–90.

10. Thoendel M, Jeraldo P, Greenwood-Quaintance K, Yao J, Abdel M, Chia N, et al. Identification of prosthetic joint pathogens directly in clinical specimens by metagenomic shotgun s¬¬¬equencing. Open Forum Infect. Dis. 2016;3:1.

11. Ferretti P, Farina S, Cristofolini M, Girolomoni G, Tett A, Segata N. Experimental metagenomics and ribosomal profiling of the human skin microbiome. Exp. Dermatol. 2017;26:211–9.

12. Horz H-P, Scheer S, Huenger F, Vianna ME, Conrads G. Selective isolation of bacterial DNA from human clinical specimens. J. Microbiol. Methods. 2008;72:98–102.

13. Feehery GR, Yigit E, Oyola SO, Langhorst BW, Schmidt VT, Stewart FJ, et al. A Method for selec-tively enriching microbial DNA from contaminating vertebrate host DNA. PLoS One. 2013;8:e76096.

14. Liu G, Weston CQ, Pham LK, Waltz S, Barnes H, King P, et al. Epigenetic segregation of microbial genomes from complex samples using restriction endonucleases hpaii and McrB. PLoS One. 2016;11: e0146064.

15. Gu W, Crawford ED, O'Donovan BD, Wilson MR, Chow ED, Retallack H, et al. Depletion of Abun-dant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. Genome Biol. 2016;17:41.

16. Nocker A, Cheung CY, Camper AK. Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of DNA from dead cells. J. Microbiol. Methods. 2006;67:310–20.

17. Fittipaldi M, Nocker A, Codony F. Progress in understanding preferential detection of live cells using viability dyes in combination with DNA amplification. J. Microbiol. Methods. 2012;91:276–89.

18. Soejima T, Iida K, Qin T, Taniai H, Seki M, Takade A, et al. Photoactivated ethidium monoazide di-rectly cleaves bacterial DNA and is applied to PCR for discrimination of live and dead bacteria. Microbi-ol. Immunol. 2007;51:763–75.

19. Alcoser SY, Kimmel DJ, Borgel SD, Carter JP, Dougherty KM, Hollingshead MG. Real-time PCR-based assay to quantify the relative amount of human and mouse tissue present in tumor xenografts. BMC Biotechnol. 2011;11:124.

20. Ji P, Zhang Y, Wang J, Zhao F. MetaSort untangles metagenome assembly by reducing microbial community complexity. Nat. Commun. 2017;8:14306.

21. Kuczynski J, Stombaugh J, Walters WA, González A, Caporaso JG, Knight R. Using QIIME to ana-lyze 16S rRNA gene sequences from microbial communities. Curr. Protoc. Microbiol. 2012;1-5.

22. Marotz C, Amir A, Humphrey G, Gaffney J, Gogul G, Knight R. DNA extraction for streamlined met-agenomics of diverse environmental samples. Biotechniques. 2017;62:290–3.

23. Carini P, Marsden PJ, Leff JW, Morgan EE, Strickland MS, Fierer N. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. Nat. Microbiol. 2016;2:16242.

24. Mahnert A, Vaishampayan P, Probst AJ, Auerbach A, Moissl-Eichinger C, Venkateswaran K, et al. Cleanroom maintenance significantly reduces abundance but not diversity of indoor microbiomes. PLoS One. 2015;10:8: e0134848.

25. Nguyen LDN, Deschaght P, Merlin S, Loywick A, Audebert C, Van Daele S, et al. Effects of propidi-um monoazide (PMA) treatment on mycobiome and bacteriome analysis of cystic fibrosis airways dur-ing exacerbation. PLoS One. 2016;11: e0168860.

26. Weber M, Geißert J, Kruse M, Lipski A. Comparative analysis of bacterial community composition in bulk tank raw milk by culture-dependent and culture-independent methods using the viability dye propidium monoazide. J. Dairy Sci. 2014;97:6761–76.

27. Zhao F, Liu H, Zhang Z, Xiao L, Sun X, Xie J, et al. Reducing bias in complex microbial community analysis in shrimp based on propidium monoazide combined with PCR-DGGE. Food Control. 2016;68:139–44.

28. Exterkate RAM, Zaura E, Brandt BW, Buijs MJ, Koopman JE, Crielaard W, et al. The effect of pro-pidium monoazide treatment on the measured bacterial composition of clinical samples after the use of a mouthwash. Clin. Oral Investig. 2015;19:813–22.

29. Kerstens M, Boulet G, Van kerckhoven M, Clais S, Lanckacker E, Delputte P, et al. A flow cytomet-ric approach to quantify biofilms. Folia Microbiol. 2015;60:335–42.

30. Köster, J, Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics. 2012:28:2520-22.

31. Didion JP, Martin M, Collins FS. Atropos: specific, sensitive, and speedy trimming of sequencing reads. PeerJ. 2017;5:e3720.

32. Langmead B, Trapnell C, Pop M, Salzberg S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25.

33. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, et al. MetaPhlAn2 for en-hanced metagenomic taxonomic profiling. Nat. Methods. 2015; 12:902–3.

34. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: An effective distance metric for microbial community comparison. ISME J. 2011;5:69–72

Chapter 3.

Computational advances for analyzing microbiome sequencing data

There are multiple properties inherent to data resulting from next generation sequencing experiments of micobial communities that complicate downstream analyses. First, this data is compositional, meaning that it only represents relative abundances or proportions of each detected microbe, and is not necessarily reflective of the number of microbes present in a given sample. The first section of this chapter describes this problem in detail and offers alternative ways to view this data to avoid making incorrect conclusions.

The second section summarizes three interrelated research advances that I have had the opportunity to contribute to. This includes a novel way to handle the inherent sparsity of microbiome sequencing datasets, interative visualizations for identifying differentially abundant microbes in compositional datasets, and an incredible, community-wide effort to consolidate emerging microbiome sequencing analysis tools into a single language. These tools complement the wet-lab advances described in Chapter 2, and together form a comprehensive view of microbiome sequencing dataset generation and analysis.

3.1

Establishing microbial composition measurement standards with reference frames

Differential abundance analysis is controversial throughout microbiome research. Current gold standard approaches require laborious measurements of total microbial load, or absolute number of microorganisms, to accurately determine taxonomic shifts among samples. Therefore, most studies rely on making conclusions based off changes in relative abundance. We highlight commonly made pitfalls in comparing relative abundance across samples and identify two solutions that reveal microbial changes without the need to estimate total microbial load. We define the notion of "reference frames", which provide deep intuition about the compositional nature of microbiome data. In an oral time series experiment, reference frames alleviate false positives and produce consistent results on both raw and cell count normalized data. Furthermore, reference frames identify consistent, differentially abundant microbes previously undetected in two independent published datasets from subjects with atopic dermatitis. These methods allow re-assessment of published relative abundance data to reveal reproducible microbial changes from standard sequencing output without the need for new molecular assays.

3.1.1 Introduction

Next-generation sequencing data used to study the microbiome is inherently compositional and provides information in the form of relative abundances, independent of the total microbial load of the original sample. Numerous analytical approaches including rarefaction [1], median [2], and quan- tile normalization [2,3] have been proposed for comparing compositional samples.

However, these analytical solutions cannot control false discovery rates [4,5], and their application contributes to lack of reproducibility among microbiome studies [6-8]. Here we illustrate mathematical challenges in analyzing compositional microbiome data from DNA sequence reads, and define the concept of "reference frames" for inferring changes in abundance.

To illustrate the pitfalls of inferring changes in abundance among samples using relative abundance data, consider the following example (Fig. 1). Samples from a population containing only two taxa (orange and blue) are collected pre- and post-treatment. Before treatment, the two taxa occur in equal proportions. After treatment, the orange taxon is twice as abundant as the blue taxon. It is tempting to conclude that orange increased and blue decreased.

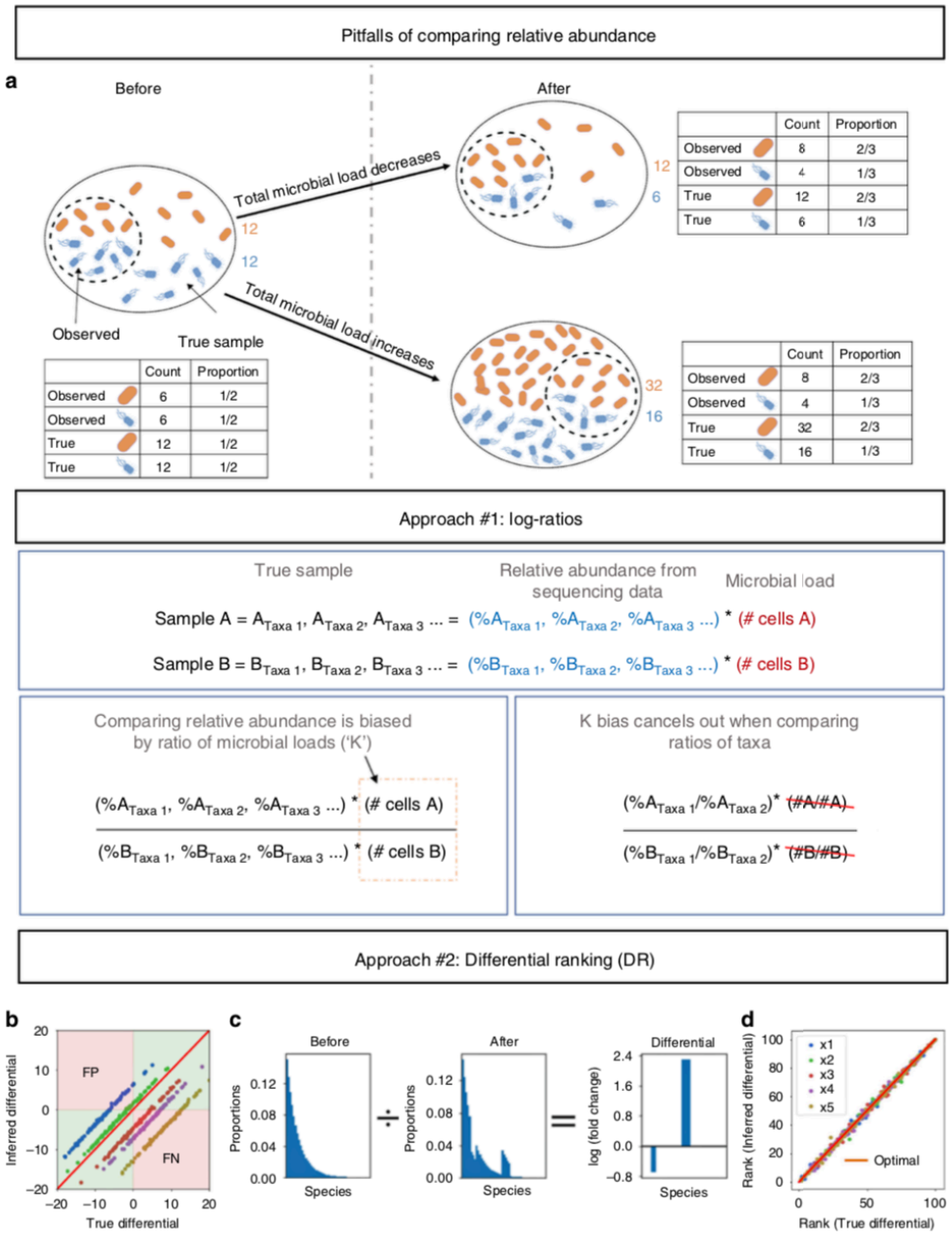

Figure 3.1.1. *Illustration demonstrating statistical limitations inherent in compositional datasets*. **a** Two different biological scenarios can yield the exact same proportions of taxa in samples from a population pre- and post-treatment. **b** Simulated datasets plotting the true differential obtained using absolute abundance data on the x-axis, versus the inferred differential obtained using relative abundance data on the y-axis. Each dot represents a taxon in the dataset, and the colors represent datasets with various ratios of total microbial load (K) between before and after samples. The red line represents the optimal scenario where the samples have equal microbial load. This illustrates the prevalence of either false positives (FP) or false negatives (FN) when performing differential abundance analysis on samples with unequal total microbial load. The presence of either FPs or FNs is dictated by a nonlinear function of the true differential (see online methods). **c** An illustration of differential proportions of bacterial species before and after treatment. **d** Same data as **b** but plotting the rank of the differentials, demonstrating that ranks are equivalent regardless of differences in microbial load.

However, many different scenarios could lead to the same observation. For example, the orange taxon could quadruple and the blue taxon only double. The orange taxon could remain constant, and the blue taxon halve. Or the orange taxon could halve, but the blue taxon could decrease four-fold. Because we only observe relative abundance data, we cannot differentiate among these outcomes, which have markedly different biological significance. Infinite different outcomes produce the same 2:1 ratio of orange to blue, greatly complicating the generation of a meaningful null hypothesis and therefore yielding misleading p-values, as has been previously established [9-11].

Multiple processing steps are required to generate microbiome sequencing data. Samples are col- lected from a much larger population (e.g. fecal material from the gut, or water sample from the ocean). From these samples, a subsample is used for DNA extraction (e.g. a swab from a fecal sample, or an aliquot of a water sample). Even if the same amount of sample is extracted throughout an experiment, many DNA extraction kits are optimized for efficiency and can become saturated, complicating direct correlations of DNA yield and microbial load. A subsample of the extracted DNA is then used as input for PCR, a subset of the resulting amplicon is pooled into a library, and a subset of the library is sequenced.

By the time quality-filtered sequencing data is obtained, the sequences reflect only a small subset of the population and are not an accurate representation of the microbial load in the original sample [12]. Analyzing relative abundance data with inappropriate statistical tools can yield up to 100% false discovery rates [13-14]. Therefore, in addition to relative abundance data, quantitative information about total microbial load is necessary to determine which microbes are changing.

Multiple approaches at each level of sample processing have been proposed to quantify the total microbial load from environmental samples. Adding a known amount of reference DNA as

an internal standard has been used to extrapolate the amount of starting nucleic material [15,16]. Normalization by this method is complicated due to the calibration challenges of choosing the proper amount of internal standard [16]. At the post- extraction level, quantitative PCR (qPCR) of genomic DNA with universal primers against the 16S rRNA gene has been deployed to estimate total microbial load [17]. However, it is impossible to prevent primer bias, resulting in uneven amplification of rRNA genes across species, and the DNA extraction method can influence microbial composition [18-20]. Further, quantification by both spike-in and qPCR is performed on multiple subsets of the original sample.

Quantifying microbial load by flow cytometry is performed on the original sample, and is agnostic to nucleotide sequences. One recent study reported that adding quantitative information obtained by flow cytometry dramatically improved interpretation of 16S rRNA gene amplicon sequencing data [12]. However, flow cytometry requires expensive, relatively low-throughput equipment, and often can only estimate the cell concentration rather than the total microbial load.

The total microbial load of an environmental sample is one dimension of measurement among the hundreds to thousands of dimensions measured by microbial relative abundances. If the absolute abundance of one taxon and the relative abundance of all taxa is known, it is feasible to compute the absolute abundance of all taxa. As such, considerable information rests in relative abundances, and important insights can be gleaned without costly microbial quantification methods. Below we describe two methods to evaluate relative differential abundance independent of microbial load information.

3.1.2 Results

Ratios circumvent bias without microbial load quantification

Computing changes in abundance from compositional data introduces a bias due to the lack of total microbial load (Fig. 1 approach#1). Simulated data in Figure 1b shows how different biases (i.e. ratios between total microbial loads) can cause either false positives or false negatives. By simply comparing the ratio of taxa between samples, the bias constant introduced by unknown microbial load cancels out. Taking the logarithm of this ratio (log-ratio) enforces symmetry around zero, giving equal weight to relative increases and relative decreases [9,10].

A novel approach to rank differential abundance

Comparing ratios of taxa can circumvent the bias introduced by unknown microbial loads. How- ever, choosing taxa for comparison from the thousands in a given sample set can be challenging. Here we provide a way to rank microbes that are changing the most relative to each other. The term "differential" refers to the logarithm of the fold change in abundance of a taxa between two conditions. With microbial load information, one can calculate absolute differentials. Microbiome sequencing datasets provide relative abundances, and thus can only infer relative differentials.

The ranks of relative differentials are identical to the ranks of absolute differentials (Fig. 1d). However, because of the bias described above, we cannot infer if a microbe has changed based on rank alone, and therefore a coefficient of zero does not imply that the microbe has not changed abundance.

Relative differentials can be estimated directly using multinomial regression, which has been proposed previously to handle sampling zeros [21-24]. The coefficients from multinomial

regression analysis can be ranked to determine which taxa are changing the most between samples. We refer to this ranking procedure as Differential Ranking (DR).

Reference frames in compositional data analysis

Analyzing compositional data requires a choice of reference frames for inferring changes in abundance. By "reference frame", we draw on the concept from physics where velocity is measured "relative to" another moving object. As microbial populations change, we can constrain our inferences to how microbial populations change relative to reference frames given by other microbial populations. The denominator in a log-ratio determines the reference frame for inferring changes. In DR, the differential abundance of each taxon serves as a reference to each other when they are ranked numerically. To demonstrate these principles, we confirm the utility of employing reference frames in biological datasets.

DR reveals differentially abundant microbes in saliva

We demonstrate the utility of DR in a sample set with dramatic differences in total microbial load. Unstimulated saliva samples were collected before and after brushing teeth (n=32) and processed in parallel for microbial load quantification with flow cytometry and 16S rRNA gene amplicon sequencing. Importantly, participants were asked to provide unstimulated saliva for exactly 5 minutes. As a result, we obtained a proxy for the total microbial load by taking into account salivary flow rate. As expected, the total microbial load significantly decreased after brushing teeth (Fig. 2a).

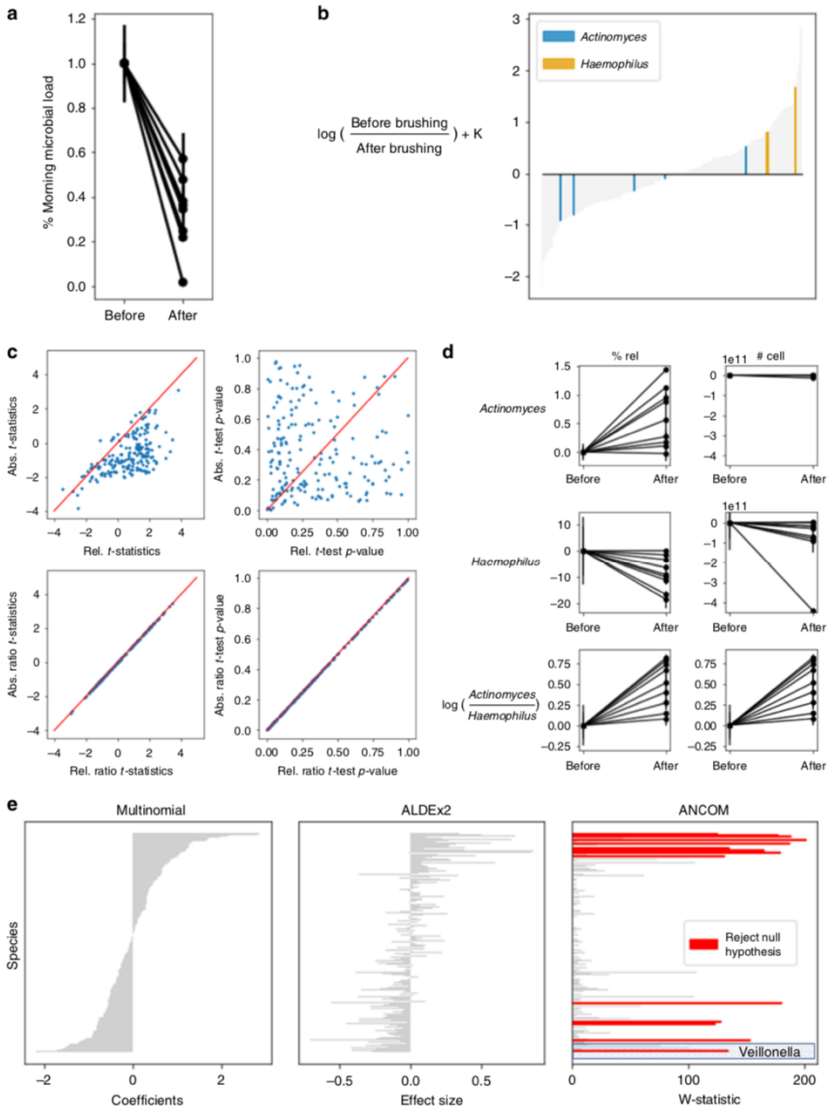

Figure 3.1.2. *Analysis of salivary microbiota before and after brushing teeth*. **a** Flow-cytometry-quantified microbial load in unstimulated saliva collected for 5 min normalized to before brushing teeth. Each line corresponds to a different volunteer. Error bars represent the standard deviation from duplicate flow-cytometry measurements. **b** Microbial ranks estimated from multinomial regression with Actinomyces and Haemophilus highlighted. The y-axis represents the log-fold change that is known up to some bias constant K, and the x-axis numerically orders the ranks of each taxa in the analysis. **c** A comparison of t- statistics (left) and p-values (right) between before and after samples where each dot is an individual taxon (top graphs) or ratio between each taxon to Actinomyces (bottom graphs) calculated from relative abundance data (x-axis) and absolute abundance data (y-axis). The 1-1 correspondence in the ratio graphs is a result of the microbial loads cancelling out, as described in Eq. (3). **d** A comparison of relative abundance vs absolute abundance data of Actinomyces, Haemophilus and log(Actinomyces: Haemophilus) before and after brushing teeth. Error bars represent standard error of the mean. **e** Comparison of the multinomial coefficients used for DR, ALDEx2 and ANCOM outputs. The test statistics generated from ALDEx2 and ANCOM are sorted in the same order as the multinomial coefficients to provide a consistent comparison. All taxa that passed the significance tests are highlighted in red.

We performed paired t-tests to evaluate the change in abundance of each taxon before and after brushing teeth using either relative or absolute abundance data (microbial load multiplied by 16S copy number-corrected relative abundances) (Fig. 2c). Applying t-tests to the relative data had a high false-positive rate, as seen by the disagreements between the relative and absolute t-statistics (Spearman r=0.53). Further, there was no correlation in p-value distribution between the relative and absolute abundance data (Spearman r=0.09), highlighting issues when the null hypothesis is not consistent between the relative abundances and the absolute abundances.

Alternatively, evaluating the ratio between Actinomyces and the remaining taxa produced identical t-statistics and p-values between the relative and absolute abundance data (Spearman r=1.0). Ratio-based analyses are unaffected by microbial load (equation 3 in methods) and result in identical interpretations as one obtains from costly and rate-limiting flow-cytometry measurements.

From the DR analysis (Fig. 2c), we can identify which taxa are changing the most relative to each other). Here, we highlight *Actinomyces* and *Haemophilus* species, which have very different ranks. *Actinomyces* tend to have low ranks and *Haemophilus* have high ranks. The difference in ranks between these taxa correctly suggests that *Haemophilus* taxa are more prevalent relative to other taxa before brushing, and *Actinomyces* taxa are more prevalent relative to other taxa after brushing. From the t-test results on relative abundances it appears that *Actinomyces* significantly increased (t-statistic=3.74, p-value=0.002) after brushing teeth and that *Haemophilus* significantly decreased (t-statistic=−3.67, p-value=0.002). However, absolute abundance data revealed that only *Haemophilus* significantly decreased (t-statistic=−2.155, p-value=0.0478) (Fig. 2d).

The log-ratio of *Actinomyces* and *Haemophilus* between the relative and the absolute abundance data is identical. While we cannot observe the decrease of *Haemophilus* or the consistency of *Acti- nomyces* abundance, with the log-ratio of their relative abundance we can observe the interaction between these two taxa and the increase of *Actinomyces* relative to *Haemophilus* after brushing teeth (t-statistic=5.289, p-value=9.07×10^{-5}).

These results are consistent with our knowledge about oral biogeography. *Haemophilus* is typically found on the periphery of oral biofilms and was likely removed from the biofilm during the brushing process, whereas *Actinomyces* is generally found on the surface of the tooth and acts as an anchor for biofilm attachment[25]. Importantly, this experiment demonstrates the potential fallibility of relying on relative abundance; it is incorrect to conclude that *Actinomyces* increases after tooth brushing despite the increase in relative abundance. As demonstrated by flow cytometry, total microbial load decreases, and while both *Haemophilus* and *Actinomyces* decrease, *Haemophilus* decreases more.

To investigate how other compositional methods perform, we ran ANCOM and ALDEx2 on the same dataset (Fig. 2e). ALDEx2 did not identify any of the microbes to be changing, which contradicts flow-cytometry measurements that show there is a large decrease in the microbial community after tooth brushing. ANCOM identified multiple significantly changing microbes. One of these detected microbes was Veillonella, which conflicts with absolute abundances suggesting that Veillonella is not significantly changing (t-statistic=1.04, p-value=0.315). The false positive detected by ANCOM likely arose due to their choice of reference frame.

Elucidating interkingdom relationships in atopic dermatitis using DR

The tooth brushing example provides ground truth for using log-ratios and DR, but many clinically relevant microbiome questions involve less obvious differences. Using data from patients with atopic dermatitis (AD), an important skin disease, we demonstrate how viewing relative abundances alone can produce false negatives.

AD has a complex etiology. Many microbiome studies performed using next-generation sequencing have focused on bacterial changes associated with AD, especially the pathogen *Staphylococcus aureus*. The yeast genus *Malassezia* has also been implicated in AD, although conflicting results have been published as to which *Malassezia* species are involved and whether they are more or less prevalent in AD [26]. A recent shotgun metagenomic study examined the skin microbiome over time during an AD flare and recovery. The authors observed a decrease in *Staphylococcus aureus* relative abundance in the healthy, recovered skin (non-lesioned) compared to AD flare (lesion), but no significant changes in the relative abundance of *Malassezia* species over time in these AD patients [27].

Applying compositional methods to this dataset revealed new insights. Observing the DR results (Fig. 3a), it is apparent that, compared to lesioned skin, *S. aureus* is one of the taxa to decrease the most relative to all other microbes in the non-lesioned sites, followed by *S. epidermidis*, and *M. globosa*. Consistent with the analysis of relative abundance in Fig. 3b, the ratio of *S. aureus : P. acnes* was significantly increased in flare (t-statistic=3.397, p-value=3.02×10^{-3}) and correlated with SCORAD score, a clinical assessment of AD severity (Pearson=0.603, p- value=3.516×10^{-6}). Contrary to previous findings, both *S. epidermidis : P. acnes* and *M. globosa : P. acnes* were also significantly increased in lesioned skin (t-statistic=4.2297, p-

value=4.53×10^{-4}, and t-statistic=4.297, p-value=3.889 × 10^{-4}, respectively) and correlated with SCORAD score (Pearson r=0.464, p-value=6.975 × 10^{-4}, and Pearson r=0.668, p-value=1.125 × 10^{-7}, respectively) (Fig. 3c).

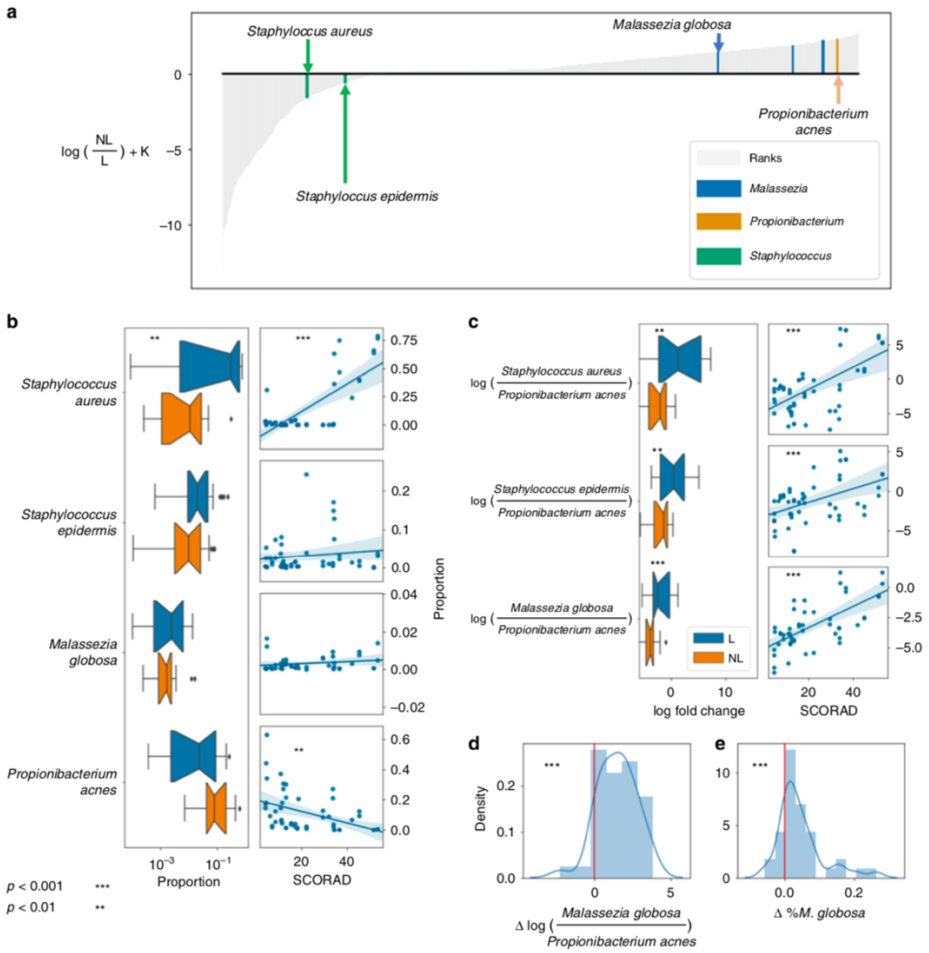

Figure 3.1.3. *DR analysis of skin in two atopic dermatitis studies*. Panels **a–c** represent data from Byrd et al.[27], and panels d, e represent data from Leung et al.[28]. Both studies compare lesioned (L) to non-lesioned (NL) skin. **a** Microbial ranks estimated from multinomial regression applied to shotgun metagenomics from Byrd et al.[27] with key genera highlighted. The y-axis represents the log-fold change that is known up to some bias constant K. **b** Proportions of *S. aureus, S. epidermidis, M. globosa*, and *P. acnes* in lesioned (blue) and non-lesioned (orange) skin (left) and correlation of relative abundance with SCORAD score (right). **c** Log-ratios of (*S. aureus: P. acnes*), (*S. epidermidis: P. acnes*), and (M. globosa: P. acnes) (left) and correlation of ratio with SCORAD score (right). Error bars represent standard deviation across participants (n = 20). **d** Change in log-ratio of (*M. globosa: P. acnes*) from Leung et al. [28]. **e** Change in relative abundance of *M. globosa* between lesioned and non-lesioned skin from Leung et al.[28]. Presented p-values are from paired t-test statistics.

To validate this observation, we analyzed shotgun data from an independent AD dataset [28]. In this dataset, the relative abundance of *M. globosa* significantly increased between lesioned and non-lesioned skin (Fig. 3e, t-statistic=4.135, p-value=0.0001). But the ratio of *M. globosa : P. acnes* increased even more dramatically in lesioned skin (t-statistic=7.298, p-value=9.729×10^{-9}) (Fig. 3d). These results are congruent with a previous report that *M. globosa* was cultivated more successfully from lesioned versus non-lesioned sites in AD [29]. Thus, DR analysis can identify novel, clinically significant microbial changes which can be validated across cohorts by choosing insightful reference frames.

DR across environmental gradients in the Central Park soils

Differential ranks can also be learned for continuously valued data. We demonstrate this with data from the Central Parks Soil experiment [31] which contains more than 1,000 samples and 30,000 taxa sampled across pH and nitrogen gradients. The largest factor driving diversity was pH, and Washburne et al [32] showed that there were lineages of microbes associated with nitrogen when pH was accounted for. Here we applied multinomial linear regression to estimate microbial DR along both nitrogen and pH gradients (Fig. 4).

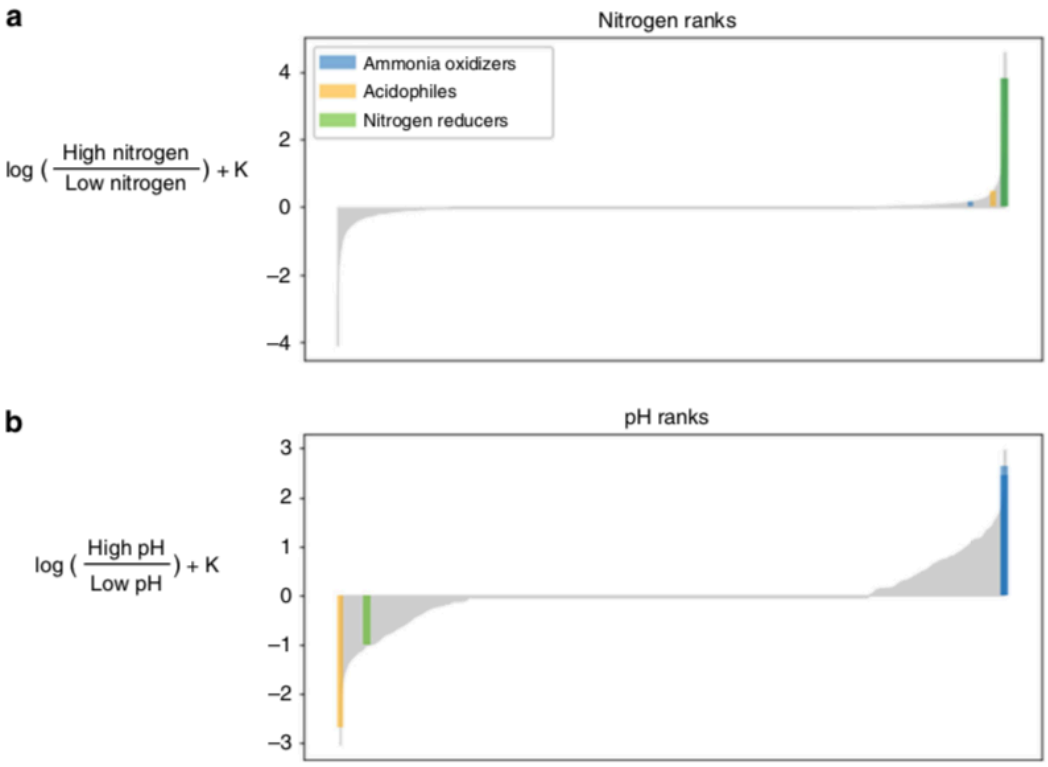

Figure 3.1.4. *DR analysis of the Central Park dataset*. **a** Microbes ranked with respect to their association with nitrogen. **b** Microbes ranked with respect to their association with pH. Putative hits against an acidophile, an ammonia oxidizer and a nitrogen reducer are highlighted.

Of the top 5 and bottom 5 ranked microbes in the nitrogen and pH gradients, only 4 microbes were annotated. The top 4th and 5th microbe that is associated with acidic environments was a putative match against *Candidatus Solibacter* and *Telmatobacter*, which has been found to grow in a pH range of 3.5-6 [33-34]. The top microbe most associated with high pH was *Chryseolinea*, which has been shown to grow between pH range of 5-10. The top third microbe associated with low nitrogen concentration was a putative match against *Gemmatimonas*, which is a known nitrogen reducer [35].

The multinomial regression was able to appropriately identify which organisms were most associated with low pH, high pH, and nitrogen. However, even amongst the highly ranked

organisms there is a major lack of functional annotations. Having the appropriate rankings in place may pro- vide new insights into these organisms and guide experimental validation.

3.1.3 Discussion

Adding information about absolute microbial load between samples can highlight issues inherent in compositional data analysis. However, there are multiple practical and technical challenges in quantifying microbial load. For example, skin swabs are often difficult to use in flow cytometry due to very low microbial load and difficulty in transferring intact cells from swabs into liquid solution. Furthermore, skin samples are notoriously sensitive to 16S rRNA gene primer choice making qPCR quantification challenging[35]. Similarly, for historically collected samples that exist only as DNA in a freezer or as sequences in a database, flow cytometry approaches to determine absolute microbial load are not feasible.

However, absolute abundances of a community are only one degree of freedom; in a community of N species, N-1 degrees of freedom exist in the relative abundances. By using flow cytometry to quantify total microbial load, we validated these analytical tools in 16S rRNA gene amplicon sequencing data from unstimulated saliva. We found evidence of false positives when looking exclusively at changes in relative abundance before and after brushing teeth. By evaluating the ratio of *Actinomyces: Haemophilus*, we reached an identical conclusion to our absolute abundance data without the need for microbial load quantification. The consistency of our results rests in the use of ratios defining reference frames for inferring compositional changes.

Furthermore, we highlighted an example of a false negative in previously generated shotgun metagenomic data from the skin of individuals with AD. We were able to reproduce the findings that *S. aureus*, and to a lesser extent *S. epidermidis*, are differentially abundant in AD

lesions. Additionally, using log-ratios and differential ranking, we were also able to show a more subtle but statistically significant change in *M. globosa* abundance in AD lesions. This same result was obtained in two independent metagenomic studies of AD patients and agrees with previous cultivation-based work quantifying increased colony forming units of *M. globosa* in AD lesions.

Consistency between inferences made based on relative and absolute abundance data is crucial, because in many circumstances it is not possible or practical to estimate total microbial load. The seeming contradictions between absolute and relative abundances does not invalidate data from the existing 100,000+ experiments utilizing 16S rRNA gene amplicon or metagenomic sequencing[36,37]. Importantly, these techniques are not limited to next generation microbiome sequencing, but can be applied to any experiments involving compositional data (e.g. metabolomics, proteomics, etc.).

Although various methods of multinomial-based models have been developed [22–25], the interpretation of the resulting model requires care. A zero valued coefficient does not imply that the corresponding species abundance has not changed, due to the total microbial load bias as discussed in figure 1. DR provides a novel means to correctly interpret the coefficients of these models. By ranking the coefficients we can determine which taxa have changed the most relative to each other. This subtle distinction acknowledges the limits of compositional data analysis, and as demonstrated above can have dramatic impacts on data interpretation.

While there are widespread misconceptions concerning how to interpret microbial abundances, we have shown that misinterpretations stem from a misunderstanding of the reference frame used in analysis. Ongoing efforts at the NIH and EMBL-EBI have already stored petabytes of multi-omics datasets ready to be re-analyzed, and databases, such as Qiita and gcMeta, contain curated data and metadata from hundreds of thousands of samples[36,37]. There is much promise for

resolving outstanding controversies by re-analyzing these datasets using reference frames to make stable inferences of compositional change.

3.1.4. Methods

Cancelling out bias in compositional data.

The change between two samples containing compositions (e.g., absolute abundances of D microbes) $\boldsymbol{A} = (a_1, ..., a_D)$ and $\boldsymbol{B} = (b_1, ..., b_D)$, can be computed as follows:

$$\frac{\boldsymbol{A}}{\boldsymbol{B}} = \left(\frac{a_1}{b_1}, \dots \frac{a_D}{b_D}\right) \quad (1)$$

If we are only able to measure relative abundances, as is the case with next generation amplicon sequencing, we can only estimate the proportion a_i for species i in the sample A. Estimating the true abundance can be done via $a_1 = N_a P_{a1}$, where N_a is the total abundance of sample A. To estimate the true change,

$$\frac{\boldsymbol{A}}{\boldsymbol{B}} = \left(\frac{N_a \times p_{a_1}}{N_b \times p_{b_1}}, \dots \frac{N_A \times p_{a_D}}{N_B \times p_{b_D}}\right) = \frac{\boldsymbol{p_A}}{\boldsymbol{p_B}} \times \frac{N_A}{N_B} \quad (2)$$

To determine if species i abundance has changed between samples A and B, we test to see if $ai/bi = 1$. However, as shown above, we cannot perform this test, since the results of this test would be confounded by the total biomass bias N_A/N_B.

In many cases the total biomass cannot be estimated, so any techniques to identify important species will need to alleviate this bias. One alternative is to use ratios. If we choose species D to be the reference species, it is clear that the total biomass cancels as follows

$$\frac{a_1/a_D}{b_1/b_D} = \frac{p_{a_1}/p_{a_D}}{p_{b_1}/p_{b_D}} \quad (3)$$

Another alternative is to use ranks. Ranks have been shown to be context of microbiome studies [38,39] and have been commonly employed to study species richness in the context of ecology [40]. Since the bias is applied uniformly across the differential, it will not affect the ordering of the species. Hence, ranks are agnostic to the total biomass bias.

$$\operatorname{rank}\left(\frac{\boldsymbol{A}}{\boldsymbol{B}}\right) = \operatorname{rank}\left(\frac{\boldsymbol{p_A}}{\boldsymbol{p_B}} \times \frac{N_A}{N_B}\right) = \operatorname{rank}\left(\frac{\boldsymbol{p_A}}{\boldsymbol{p_B}}\right) \tag{4}$$

Because of the equivalence of ranks between absolute and relative data, it is possible to identify the species that are increasing or decreasing the most. This means that the following statements hold

$$\arg\max\left(\frac{A}{B}\right) = \arg\max\left(\frac{p_A}{p_B}\right) \tag{5}$$

$$\arg\min\left(\frac{A}{B}\right) = \arg\min\left(\frac{p_A}{p_B}\right) \tag{6}$$

The ranks are connected to the log-ratios, the differences between ranks will yield differences in log-ratios given by

$$\log\frac{a_i/N_a}{b_i/N_b} - \log\frac{a_j/N_a}{b_j/N_b} = \log\frac{a_i/b_i}{a_j/b_j} = \log\frac{a_i/a_j}{b_i/b_j} \tag{7}$$

These ranks are still relative; a microbe that is detected to be increasing the most could still be decreasing in absolute abundance. For instance, in the tooth brushing example, ranks identified specific genera of *Actinomyces* to be increasing the fastest, but all of the microbes are depleted, suggesting that *Actinomyces* is just decreasing much less than the other microbes. This differential is also commonly referred to as a perturbation in the context of the compositional literature [10]. It is important to note that this does not justify applying rank-based statistical methods, such as

Spearman correlation or Kruskal–Wallis, to relative abundance data since these tests do not satisfy scale invariance [41,42].

Both the log-ratios and the differential ranking techniques satisfy scale invariance, meaning that both of these techniques are agnostic to the total microbial load. This concept is critical when analyzing relative abundance data, since this is one step closer to maintaining consistent conclusions between the original environment and the observed sequences.

False discovery rates in relative differential abundance.

Attempting to estimate absolute log-fold differentials from relative abundances can result in either false positives (FP) or false-negatives (FN) depending on the distribution of true differential abundance. Whether FNs or FPs are observed depending on a nonlinear relationship involving the true (unobserved) differential abundance. To demonstrate this, let

$$\boldsymbol{\delta} = (\delta_1, \dots, \delta_D) = \left(\log\left(\frac{a_1}{b_1}\right), \dots, \log\left(\frac{a_D}{b_D}\right)\right)$$

denote the absolute differential of the D species between two conditions, *A* and *B*. Further, let

$$\widehat{\boldsymbol{\delta}} = (\hat{\delta}_1, \dots, \hat{\delta}_D) = \left(\log\left(\frac{a_1/N_A}{b_1/N_B}\right), \dots, \log\left(\frac{a_D/N_A}{b_D/N_B}\right)\right)$$

represent the relative differentials from compositional data. By definition, we know the following is true

$$\log\left(\frac{a_i}{b_i}\right) = \log\left(\frac{a_i/N_A}{b_i/N_B}\right) + \log\left(\frac{N_A}{N_B}\right) \qquad (8)$$

If $\log(N_A/N_B) > 0$, then that will mean that for every microbe i, $\delta i > \hat{\delta}_i$. This implies that there is increased microbial load in *A* compared to *B*, and that this increase will give rise to FNs. This is because the overall community increase will not be captured from the relative abundance data.

In contrast, if $\log(N_A/N_B) < 0$, then for every microbe i, $\delta i < \hat{\delta}_i$. This means that there is a decrease in the absolute microbial load in *A* compared to *B*. This decline in the total community will not be captured from the relative abundances, and some of the species will be detected to be increasing, giving rise to FPs. An example of this was shown in the saliva microbiota study (Fig. 2).

Multinomial regression

To perform the differential ranking (DR) analysis, we used multinomial regression. Multinomial regression and related count regression models are commonly used in the context of microbiome analysis. Here, we use the multinomial regression model since these models can reliably estimate the means and can be easily reinterpreted in the context of compositional data analysis. Counts from the multinomial regression can be formulated using additive log-ratio transformation (alr) in the following generative model

$$\beta_{jk} \sim \mathcal{N}(0, \mu_\beta) \tag{9}$$

$$\boldsymbol{\eta_i} = \text{alr}^{-1}(\boldsymbol{X_i}\boldsymbol{\beta}) \tag{10}$$

$$\boldsymbol{Y_i} \sim \text{Multinomial}(\boldsymbol{\eta_i}), \tag{11}$$

where $\boldsymbol{Y_i}$ represents the measured microbial load for sample *i*. $\boldsymbol{\beta}$ represents the coefficients of the model across all measured covariates indexed by *k*. These coefficients can be interpreted as a relative differential discussed in the examples above. $\boldsymbol{X_i}$ represents the vector metadata covariates for sample *i*. These metadata covariates can represent both continuous and categorical variables, where categorical variables are represented as binary variables. A normal prior centered around zero was placed on the coefficients $\boldsymbol{\beta}$ to serve as regularization to combat issues associated with high dimensionality. The *j*th component of the $\boldsymbol{\beta_k}$ coefficient vector represents the jth alr

coordinate, which can be interpreted as a log concentration using one of the microbes as a reference. It does not matter which microbe is used as a reference, since the proportions η_i will be identical regardless of reference microbe. The inverse alr function is commonly used in the context of compositional data, given as follows

$$\text{alr}^{-1}(x) = C[\exp(0, x_1, \ldots, x_{D-1})] \quad C[x] = \left[\frac{x_1}{\sum_{i=1}^{D} x_i}, \ldots, \frac{x_D}{\sum_{i=1}^{D} x_i}\right]. \qquad (12)$$

This is also referred to as a degenerate softmax function, which is commonly used in the context of neural networks. This function is isomorphic between $R^{\text{D-1}}$ and S^{D} (the space of proportions), so this will ward against identifiability issues when estimating these model parameters. The alr function is defined as

$$\text{alr}(x) = \left(\log\frac{x_2}{x_1}, \log\frac{x_3}{x_1}, \ldots, \log\frac{x_D}{x_1}\right) \qquad (13)$$

The models were estimated using a maximum a posteriori priori (MAP) estimation using stochastic gradient descent. Multinomial regression was implemented using Tensorflow [43] and can be found in https://github.com/biocore/songbird.

Interpreting ranks

Supplementary Fig. 2 outlines how to draw hypotheses using the proposed ranking procedure. First the relative differentials need to be computed, preferably using a count-based regression model such as the multinomial regression described above. As noted in the introduction, the coefficients can be represented as centered log ratio (clr) coordinates as follows

$$\beta_k^{(\mathrm{clr})} = \mathrm{clr}(\mathrm{alr}^{-1}(\beta_k)) \tag{14}$$

$$\mathrm{clr}(x) = \left(\log\left(\frac{x_1}{g(x)}\right), \ldots, \log\left(\frac{x_D}{g(x)}\right)\right) \tag{15}$$

where *g(x)* represents the geometric mean. These coordinates are typically centered around zero, meaning that the chosen reference frame is the center-of-mass, or in other words the average microbe. This is the same reference frame that ALDEx2 and sometimes ANCOM uses. Once the relative differentials are estimated there are two possible analyses. It is possible to construct compositional biplots to visualize all of the regression coefficients and determine how microbes are clustered and driven by metadata covariates. This procedure is outlined in the Songbird tutorial on github.

The other possibility is to identify candidate differentially abundant microbes. To this end, one can construct rank plots (e.g., Figs. 2b and 3a). The rank plots show the ordering of all of the taxa with respect to how much they are associated with a particular metadata covariate, and specific taxa can be highlighted to show their ranks as positions on the rank plot. From the ranks, one can focus on taxa that have very high ranks or very low ranks, since those are the ones that are increasing/decreasing the most relative to each other, and are likely to be important contributors.

These ranks can also help inform which taxa can be used for a suitable reference frame since the difference between the relative differentials can approximate the effect size that those two microbes will have. As a result, microbes that have very different ranks can be suitable candidates for a log-ratio test. An ideal reference microbe is present across most samples, since this will allow the denominator in a log-ratio to be defined. This was one of the reasons why Actinomyces and P. acnes were chosen as reference microbes in the case studies. Furthermore, if a microbe is anticipated to be stable across experimental conditions, this could provide additional motivation to select that microbe as the reference microbe.

Zeros will remain to be problematic when comparing log-ratios of taxa among conditions—the procedure used here was to treat zeros as missing data and drop them from the analysis. However, this approach may not be optimal, for instance if two microbes never occur together in the same sample, but one microbe has a very high rank, and the other microbe has a very low rank. The two microbes may have significant explanatory power, but it will not be possible to perform a log-ratio test without imputing the zeros. In scenarios such as this, it may be more appropriate to utilize presence/absence procedures.

While the above procedures provide some recommendations on how to pick an appropriate reference frame, picking a reference frame for hypothesis testing is still an outstanding challenge. Since a reference frame can be defined as the average of a set of microbes, there are 2^N possible reference frames for N microbes. Rivera–Pinto proposed one approach towards finding an optimal reference frame [44]; however, this solution maybe suboptimal. Furthermore, it is not clear what properties an optimal reference frame should satisfy, or how false discoveries could be controlled. More theoretical work will need to be done in order to understand statistical properties of these reference frames.

Simulated benchmarks comparing ANCOM2, ALDEx2, and DR.

We used simulated data to benchmark DR to the output of ALDEx2 [45] and ANCOM [46]. Details can be found in the simulation-benchmarks ipynb at https://github.com/knightlab-analyses/reference-frames. Here, we compared the linear mixed effects model in ANCOM2, the t-test in ALDEx2 and ranked multinomial regression coefficients in DR. ALDEx2 determined taxa were significant if the FDR corrected p-value fell below 0.05. A taxon was determined to be significant by ANCOM if it passed the 0.9 cutoff.

ALDEx2 and the proposed multinomial regression for DR in this paper use nearly identical models concerning categorical metadata. The major difference is the choice of priors; our model uses a normal prior whereas ALDEx2 uses a Dirichilet prior. As a result, the coefficients from ALDEx2 and the multinomial regression are nearly identical (Supplementary Fig. 1), suggesting that the same ranking procedure can also be applied to the ALDEx2 output. However, ALDEx2 can only handle a single categorical covariate at a time, whereas the multinomial regression proposed can handle multiple covariates, including continuously valued covariates, as shown in the Central Park soils dataset (Fig. 4).

It is important to note that the hypothesis tests that ANCOM and ALDEx2 use may not be consistent with the absolute differentials. Under perfect conditions when the absolute differentials are centered around zero (Supplementary Fig. 1a–d), both ANCOM and ALDEx2 correctly infer that microbes with a differential close to zero are likely not changing. However, if the center of mass changes and the average microbe is now decreasing on average −2 log fold (Supplementary Fig. 1e–h), both ALDEx2 and ANCOM will incorrectly infer that microbes changing −2 log fold are not changing. In this example, the center of mass reference frame is inappropriate, and leads to predictions that microbes are not changing when they are actually changing on an absolute scale. This highlights difficulties when attempting to link information from relative data to absolute data using hypothesis tests. The hypothesis tests that ALDEx2 and ANCOM perform here are not necessarily incorrect, but could be misleading in situations where microbial load differs dramatically among conditions.

Interpreting relative differentials through balances

Balances are ratios of taxa, or groups of taxa, that were previously presented as a valid approach to analyzing compositional data [14]. If we examine the model parameters $\boldsymbol{\beta_k} \in \mathbb{R}^{D-1}$, we reinterpret the quantities given by $\text{alr}^{-1}(\boldsymbol{\beta k})$ as relative differentials as discussed in Fig. 1. It is also worthwhile to note the connection between βk and balances. Since $\boldsymbol{\beta k}$ is expressed in alr coordinates, there is also a direct connection to ilr coordinates, meaning that $\boldsymbol{\beta k}$ can also be transformed into balances. More explicitly, the ilr coordinates of these coefficients can be computed as follows

$$\boldsymbol{\beta}_{\mathbf{k}}^{(\text{ilr})} = \text{ilr}_{\boldsymbol{\Psi}}(\text{alr}^{-1}(\boldsymbol{\beta}_{\boldsymbol{k}})). \tag{16}$$

The resulting coefficients are represented as coordinates given by the orthonormal basis Ψ. An example of such a basis can be derived from bifurcating trees discussed in Morton et al. [14], Silverman et al. [47], and Washburne et al. [31]. This can allow for relative changes in abundances as given by $\text{alr}^{-1}(\boldsymbol{\beta k})$ to inform which balances are changing in ancestral states given by the tree. The multinomial regression serves as an alternative means to compute regression coefficients discussed in PhILR, Phylofactor and Gneiss, while avoiding issues with imputation and zeros.

Saliva microbiota study

Nine volunteers provided unstimulated saliva so that salivary flow rate could be measured according to a standardized protocol [48]. Briefly, individuals were asked to allow saliva to flow for exactly five minutes through a disposable funnel (Simport, SIM F490-2) into a sterile, 15 mL conical tube preloaded with 2 mL sterile glycerol for bacterial preservation. Participants were asked to provide samples before brushing and after brushing teeth in the morning and in the evening. Samples were inverted several times to mix with the glycerol and stored at −20°C

immediately after collection. This study was approved by an Institutional Review Board (IRB# 150275) and written informed consent was acquired before sample collection.

Unstimulated saliva samples were thawed on ice and aliquots were diluted tenfold with sterile, 1x PBS. To remove human cells and salivary debris, samples were filtered using a sterile 5 µm syringe filter (Sartorius Stedim Biotech GmbH). 5 µl 20x SYBR green (SYBR™ Green I Nucleic Acid Gel Stain, Invitrogen) was added to 1 mL of the microbial suspension (0.1x final concentration) and incubated in the dark for 15 min at 37°C. Finally, 50 µl AccuCount Fluorescent Particles (Spherotech, ACFP-70-10) were added for assessment of microbial load. Samples were processed on a SH800 Cell Sorter (Sony Biotechnology) using a 100 µm chip with the threshold set on FL1 at 0.06%, and gain settings as follows; FSC = 4, BSC = 25%, FL1 = 43%, FL4 = 50%. The gating strategy was adapted from Vandeputte et al. [12] Briefly, fluorescent microbial cells were gated from background on a FL1-Fl4 density plot. Aggregates were excluded by taking the linear fraction on a plot of FL1-height versus FL1-area as previously described [49], and remaining background was removed by eliminating large events detected on a FSC-BSC density plot (Supplementary Fig. 3). Negative controls (sterile PBS stained identically to samples) were run between each sample set to exclude cross contamination. Settings were identical among all samples.

DNA extraction and 16S rRNA amplicon sequencing were done using Earth Microbiome Project (EMP) standard protocols (http://www.earthmicrobiome.org/protocols-and-standards/16s). Five hundred microliter of unstimulated saliva was used for gDNA extraction with MagAttract PowerSoil DNA Kit (QIAGEN) as previously described [50]. Amplicon PCR was performed on the V4 region of the 16S rRNA gene using the primer pair 515f to 806r with Golay error-correcting barcodes on the reverse primer. Two hundred forty nanogram of each amplicon

was pooled and purified with the MO BIO UltraClean PCR cleanup kit and sequenced on the Illumina MiSeq sequencing platform.

Demultiplexed fastq files were processed using QIIME2 (https://qiime2.org) [51]. Deblur was used to denoise the sequences [52]. Taxonomy was assigned and 16S rRNA gene copy number-corrected using RDP classifier [53] then collapsed to the genus-level. All taxa reported in the manuscript were validated using the NCBI BLAST database [54]. Absolute abundances were estimated by multiplying the total cell-count estimated by flow cytometry by the copy number-corrected microbial proportions from sequencing as outlined above.

For differential abundance testing (Fig. 2e), ALDEx2 determined taxa were significant if the FDR corrected p-value fell below 0.05. A taxon was determined to be significant by ANCOM if it passed the 0.6 cutoff. Songbird was used to perform multinomial regression and the repository can be found here: https://github.com/ mortonjt/songbird. Paired t-tests were performed to evaluate the differences before and after brushing teeth. All log-ratios that were evaluated to either positive or negative infinity were dropped prior to statistical analysis.

Analyzing the Central Park soils study with reference frames

Data from Ramirez et al. [30] were retrieved from Qiita [55] (https://qiita.ucsd.edu/study/ description/2104). Amplicon sequence variants that appeared in less than 24 samples were filtered out, reducing the number of analyzed taxa to 30, 248 taxa. This filtering criteria was chosen to ensure that each of the six covariates had at least four samples to fit against. The following multinomial linear model was estimated

$$y_i \sim \text{Multinomial}(x_i\beta)$$
$$\beta = [\beta_{\mathbf{0}}, \beta_{\text{water}}, \beta_{\text{nitro}}, \beta_{\text{pH}}, \beta_{\text{carbon}}, \beta_{\text{biomass}}]$$

where $\boldsymbol{y_i}$ represents the microbial relative abundances in sample i, x_i are the measured covariates for sample i, β_{water} are the relative differentials with regard to water content, β_{nitro} are the relative differentials with regard to nitrogen concentration, β_{pH} are the relative differentials with regard to pH, β_{carbon} are the relative differentials with regards to carbon measurements and β_{biomass} are the relative differentials with regards to measured biomass.

Shotgun metagenome studies

We used supplementary data from Byrd et al. [27] and Leung et al. [28]. The provided relative abundances were compared to log-ratios of given taxa from the raw count data. Paired t-tests were performed to evaluate the differences between lesion and non-lesion skin samples. All log-ratios that were evaluated to either positive or negative infinity were dropped prior to statistical analysis. These numerical issues occur when particular microbes are not observed, and we treat them as missing data, respectively.

Data availability

The sequences and biom tables [56] from the saliva microbiota study can be found on Qiita (http://qiita.microbio.me) [55] under study ID 11896 and at EBI under ERP111447.

Code availability

All analyses can be found under https://github.com/knightlab-analyses/reference-frames

3.1.5 Acknowledgements

Chapter II, part 3, in full, is a reprint of previously published material: Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K. *Improving saliva shotgun metagenomics by chemical host DNA depletion*. Microbiome. 2018 Dec;6(1):42. I was the primary investigator and author of this paper. The co-authors listed above supervised or provided support for the research and have given permission for the inclusion of the work in this dissertation.

3.1.5 References

1. Weiss SJ, Xu Z, Amir A, Peddada S, Bittinger K, Gonzalez A, Lozupone C, Zaneveld JR, Vazquez-Baeza Y, Birmingham A, Knight R. Effects of library size variance, sparsity, and compositionality on the analysis of microbiome data. Peer J. 3, e1408 (2015).

2. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).

3. Paulson, J. N., Stine, O. C., Bravo, H. C. & Pop, M. Differential abundance analysis for microbial marker-gene surveys. Nat. Methods 10, 1200–1202 (2013).

4. Russel J, Thorsen J, Brejnrod AD, Bisgaard H, Sorensen SJ, Burmolle M. DAtest: a framework for choosing differential abundance or expression method. bioRxiv. Jan 1:241802 (2018).

5. Hawinkel, S., Mattiello, F., Bijnens, L. & Thas, O. A broken promise: microbiome differential abundance methods do not control the false discovery rate. Brief. Bioinform. 20, 210–221 (2017).

6. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. Microbiome. 2014 Dec 1;2(1):15.

7. Gloor, G. B., Wu, J. R., Pawlowsky-Glahn, V. & Egozcue, J. J. It's all relative: analyzing microbiome data as compositions. Ann. Epidemiol. 26, 322–329 (2015).

8. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome datasets are compositional: and this is not optional. Front. Microbiol. 8, 2224 (2017).

9. Aitchison, J. The statistical analysis of compositional data. J. R. Stat. Soc. Series B Stat. Methodol. 44, 139–160 (1982).

10. Pawlowsky-Glahn, V., Egozcue, J. J. & Tolosana-Delgado, R. Modeling and Analysis of Compositional Data (John Wiley & Sons, 2015).

11. Kumar MS, Slud EV, Okrah K, Hicks SC, Hannenhalli S, Bravo HC. Analysis and correction of compositional bias in sparse sequencing count data. BMC Genom. 19, 799 (2018).

12. Vandeputte D, Kathagen G, D'hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, Wang J, Tito RY, De Commer L, Darzi Y, Vermeire S, Falony G, Raes J. Quantitative microbiome profiling links gut community variation to microbial load. Nature 551, 507–511 (2017).

13. Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R, Peddada SD. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb. Ecol. Health Dis. 26, 27663 (2015).

14. Morton JT, Sanders J, Quinn RA, McDonald D, Gonzalez A, Vázquez-Baeza Y, Navas-Molina JA, Song SJ, Metcalf JL, Hyde ER, Lladser M. Balance trees reveal microbial niche differentiation. mSystems 2, e00162–16 (2017).

15. Smets W, Leff JW, Bradford MA, McCulley RL, Lebeer S, Fierer N. A method for simultaneous measurement of soil bacterial abundances and community composition via 16s rrna gene sequencing. Soil. Biol. Biochem. 96, 145–151 (2016).

16. Tkacz, A., Hortala, M. & Poole, P. S. Absolute quantitation of microbiota abundance in environmental samples. Microbiome 6, 110 (2018).

17. Nadkarni, M. A., Martin, F. E., Jacques, N. A. & Hunter, N. Determination of bacterial load by real-time pcr using a broad-range (universal) probe and primers set. Microbiology 148, 257–266 (2002).

18. Rubin BE, Sanders JG, Hampton-Marcell J, Owens SM, Gilbert JA, Moreau CS. DNA extraction protocols cause differences in 16s rrna amplicon sequencing efficiency but not in community profile composition or structure. Open Microbiol 3, 910–921 (2014).

19. Wagner Mackenzie, B., Waite, D. W. & Taylor, M. W. Evaluating variation in human gut microbiota profiles due to dna extraction method and inter-subject differences. Front. Microbiol. 6, 130 (2015).

20. Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z. & Forney, L. J. Evaluation of methods for the extraction and purification of dna from the human microbiome. PLoS One. 7, e33865 (2012).

21. Silverman, J. D., Durand, H., Bloom, R. J., Mukherjee, S. & David, L. A. Dynamic linear models guide design and analysis of microbiota studies within artificial human guts. Microbiome 6, 202 (2018).

22. Äijö, T., Müller, C. L. & Bonneau, R. Temporal probabilistic modeling of bacterial compositions derived from 16s rrna sequencing. Bioinformatics 34, 372–380 (2017).

23. Grantham, N. S., Reich, B. J., Borer, E. T. & Gross, K. Mimix: A bayesian mixed-effects model for microbiome data from designed experiments. Preprint at bioRxiv https://arxiv.org/abs/1703.07747 (2017).

24. Xia, F., Chen, J., Fung, W. K. & Li, H. A logistic normal multinomial regression model for microbiome compositional data analysis. Biometrics 69, 1053–1063 (2013).

25. Welch, J. L. M., Rossetti, B. J., Rieken, C. W., Dewhirst, F. E. & Borisy, G. G. Biogeography of a human oral microbiome at the micron scale. Proc. Natl Acad. Sci. USA 113, E791–E800 (2016).

26. Glatz, M., Bosshard, P. P., Hoetzenecker, W. & Schmid-Grendelmeier, P. The role of malassezia spp. in atopic dermatitis. J. Clin. Med. Res. 4, 1217–1228 (2015).

27. Byrd AL, Deming C, Cassidy SK, Harrison OJ, Ng WI, Conlan S, Belkaid Y, Segre JA, Kong HH. Staphylococcus aureus and staphylococcus epidermidis strain diversity underlying pediatric atopic dermatitis. Sci. Transl. Med. 9, eaal4651 (2017).

28. Leung DY, Calatroni A, Zaramela LS, LeBeau PK, Dyjack N, Brar K, David G, Johnson K, Leung S, Ramirez-Gama M, Liang B. The nonlesional skin surface distinguishes atopic dermatitis with food allergy as a unique endotype. Sci. Transl. Med. 11, eaav2385 (2019).

29. Falk MH, Linder MT, Johansson C, Bartosik J, Back O, Sarnhult T, Wahlgren CF, Scheynius A, Faergemann J. The prevalence of malassezia yeasts in patients with atopic dermatitis, seborrhoeic dermatitis and healthy controls. Acta Derm. Venereol. 85, 17–23 (2005).

30. Ramirez KS, Leff JW, Barberán A, Bates ST, Betley J, Crowther TW, Kelly EF, Oldfield EE, Shaw EA, Steenbock C, Bradford MA. Biogeographic patterns in below-ground diversity in new york city's central park are similar to those observed globally. Proc. R. Soc. B 281, 20141988 (2014).

31. Washburne AD, Silverman JD, Leff JW, Bennett DJ, Darcy JL, Mukherjee S, Fierer N, David LA. Phylogenetic factorization of compositional data yields lineage-level associations in microbiome datasets. Peer J. 5, e2969 (2017).

32. Ward NL, Challacombe JF, Janssen PH, Henrissat B, Coutinho PM, Wu M, Xie G, Haft DH, Sait M, Badger J, Barabote RD. Three genomes from the phylum acidobacteria provide insight into the lifestyles of these microorganisms in soils. Appl. Environ. Microbiol. 75, 2046–2056 (2009).

33. Pankratov, T. A., Kirsanova, L. A., Kaparullina, E. N., Kevbrin, V. V. & Dedysh, S. N. Telmatobacter bradus gen. nov., sp. nov., a cellulolytic facultative anaerobe from subdivision 1

of the acidobacteria, and emended description of acidobacterium capsulatum kishimoto et al. 1991. Int. J. Syst. Evol. Microbiol. 62, 430–437 (2012).

34. Fahrbach M, Kuever J, Remesch M, Huber BE, Kämpfer P, Dott W, Hollender J. Steroidobacter denitrificans gen. nov., sp. nov., a steroidal hormone-degrading gammaproteobacterium. Int. J. Syst. Evol. Microbiol. 58, 2215–2223 (2008).

35. Meisel JS, Hannigan GD, Tyldsley AS, SanMiguel AJ, Hodkinson BP, Zheng Q, Grice EA. Skin microbiome surveys are strongly influenced by experimental design. J. Invest. Dermatol. 136, 947–956 (2016).

36. Shi W, Qi H, Sun Q, Fan G, Liu S, Wang J, Zhu B, Liu H, Zhao F, Wang X, Hu X. gcMETA: a global catalogue of metagenomics platform to support the archiving, standardization and analysis of microbiome data. Nucleic Acids Res. 47, 637–648 (2018).

37. Gonzalez A, Navas-Molina JA, Kosciolek T, McDonald D, Vázquez-Baeza Y, Ackermann G, DeReus J, Janssen S, Swafford AD, Orchanian SB, Sanders JG, Briswal C, Dorrestein P, Knight R. Qiita: rapid, web-enabled microbiome meta-analysis. Nat. Methods 15, 796–798 (2018).

38. Holmes S, Alekseyenko A, Timme A, Nelson T, Pasricha PJ, Spormann A. Visualization and statistical comparisons of microbial communities using R packages on phylochip data. In Pac. Symp. on Biocomp. 142–53 (NIH Public Access, 2011).

39. Callahan, B. J., Sankaran, K., Fukuyama, J. A., McMurdie, P. J. & Holmes, S. P. Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. F1000Res. 5, 1492 (2016).

40. Magurran, A. E. Ecological diversity and its measurement (Princeton university press, Princeton, New Jersey, 1988).

41. Friedman, J. & Alm, E. J. Inferring correlation networks from genomic survey data. PLoS Comput. Biol. 8, e1002687 (2012).

42. Lovell, D., Taylor, J., Zwart, A. & Helliwell, C. Caution! compositions! can constraints on omics data lead analyses astray. CSIRO 1, 44 (2010).

43. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, Kudlur M. Tensorflow: a system for large-scale machine learning. OSDI 16, 265–283 (2016).

44. Rivera-Pinto J, Egozcue JJ, Pawlowsky-Glahn V, Paredes R, Noguera-Julian M, Calle M. Balances: a new perspective for microbiome analysis. mSystems 3, e00053–18 (2018).

45. Gloor, G. Aldex2: Anova-like differential expression tool for compositional data. ALDEX manual modular 20, 1–11 https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0067019 (2015).

46. Kaul, A., Davidov, O. & Peddada, S. D. Structural zeros in high-dimensional data with applications to microbiome studies. Biostatistics 18, 422–433 (2017).

47. Silverman, J. D., Washburne, A. D., Mukherjee, S. & David, L. A. A phylogenetic transform enhances analysis of compositional microbiota data. eLife 6, e21887 (2017).

48. Navazesh, M. & Kumar, S. K. S., University of Southern California School of Dentistry. Measuring salivary flow: challenges and opportunities. J. Am. Dent. Assoc. 139, 35S–40S (2008).

49. Props R, Kerckhof FM, Rubbens P, De Vrieze J, Sanabria EH, Waegeman W, Monsieurs P, Hammes F, Boon N. Absolute quantification of microbial taxon abundances. ISME 11, 584 (2017).

50. Marotz C, Amir A, Humphrey G, Gaffney J, Gogul G, Knight R. DNA extraction for streamlined metagenomics of diverse environmental samples. Biotechniques 62, 290–293 (2017).

51. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336 (2010).

52. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Xu ZZ, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. Deblur rapidly resolves Single-Nucleotide community sequence patterns. mSystems 2, 00191–16 (2017).

53. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73, 5261–5267 (2007).

54. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA. Database indexing for production megablast searches. Bioinformatics 24, 1757–1764 (2008).

55. Gonzalez A, Navas-Molina JA, Kosciolek T, McDonald D, Vázquez-Baeza Y, Ackermann G, DeReus J, Janssen S, Swafford AD, Orchanian SB, Sanders JG, Knight R. Qiita: rapid, web-enabled microbiome meta-analysis. Nat. Methods 15, 796–798 (2018).

3.2

Creating accessible processing tools for microbiome sequencing datasets

Next generation microbiome sequencing datasets are highly complex, often containing information about thousands of microbial taxa. These datasets are too large be processed with computer programs traditionally used in the biological sciences such as Excel. In order to draw conclusions about the underlying biology from microbiome sequencing, some degree of bioinformatics is required. This chapter highlights three computational tools that I contributed to which both advance our ability to accurately assess microbial community data and provide intuitive results that are accessible for biologists with limited bioinformatics background.

3.2.1 Robust, taxonomy driven beta-diversity analysis (RPCA)

Beta diversity is a dimensionality-reduction technique commonly used in the field of microbiome research. This tool was first described by the ecologist Robert Whitaker in 1960 [1] and is a measurement of the similarity of two environments (or samples) given all the features represented in those environments. This information is usually represented by Principal Coordinates Analysis (PCoA) which plots each sample as a point in space; the closer together two points are the more similar the microbial composition of those two samples. This is extremely useful for understanding the overall structure of a given dataset and is one of the most common tools used in analyzing microbiome sequencing datasets.

However, one major limitation of PCoA is that the axes of the plot abstractly represent percent variability explained, and you cannot retrieve information about which features are

differentiating the samples. Second, many common tools used to calculate sample similarity for PCoA do not take into account the inherent compositionality of a dataset (as described in detail in Chapter 3.1). A third obstacle to dimensionality reduction of microbiome sequencing datasets is sparsity, since there are usually relatively few microbial taxa found across all samples.

Robust Aitchison principal components analysis (RPCA) was developed in order to directly address these limitations [2]. This algorithm harnesses the power of previously developed machine learning algorithms to complete sparse matrices such as the output of microbiome sequencing. The similarity calculation is performed on centered-log ratio transformed data, which allows it to be scale invariant and account for compositionality. Importantly, this tool also allows for identification of the features that drive separation between sample groups. This tool has the ability to highlight hidden data structures not observed with traditional analyses as highlighted in Chapter 4.2 (Figure 4.2.1).

3.2.2 Compositionally aware differential feature ranking (Qurro)

Identifying microbes that are differentially abundant across environments is a major goal of microbiome studies. As described in Chapter 3.1, given the limitations of next generation sequencing ouput the best way to interpret differential abundance is through ranking. Tools such as songbird (Chapter 3.1) and RPCA (Chapter 3.2.1) produce feature rankings that can be used to estimate differential abundance, however the output of these tools is neither intuitive nor easily accessible for efficient searching without extensive bioinformatics background.

Qurro is a novel tool designed to provide an interactive interface through which you can view the output of feature ranking tools such as songbird or RPCA [3]. The output of qurro is an interactive HTML where users can search for specific microbial taxa and within seconds produce

log-ratios of microbial taxa using a differential feature rank plot to help guide microbial selection. The use of this tool requires no computational background and greatly reduces the barrier to applying state-of-the-art differential abundance tools.

3.2.3 Open-source analysis software with interactive visualizations (Q2)

Tools to analyze microbiome sequencing data are constantly evolving and adapting. These tools are written across a variety of programming languages and require various levels of expertise, which can make the analysis of these datasets daunting. Qiime2 (Q2) is an open source system for the reproducible analysis of microbiome sequencing data [4]. The architecture of Q2 is built around the concept of plugins, which allows computational tools developed by individuals across the world to be implemented into one standardized system to maximize accessibility to the microbiome community. Q2 can be used through the command line or through jupyter notebooks. Many of the outputs can be visualized through interactive HTML sites which generate figures that can be downloaded at high quality for publication.

This system has allowed for unprecedented collaboration and has dramatically strengthened the accessibility of analysis tools. In less than a year, the Qiime2 publication [4] has over 200 citations, highlighting the importance of this system for analyzing microbiome sequencining data.

3.2.4 References

1. Whittaker RH. Vegetation of the Siskiyou mountains, Oregon and California. Ecological monographs. 1960 Feb;30(3):279-338.

2. Martino C, Morton JT, Marotz CA, Thompson LR, Tripathi A, Knight R, Zengler K. A novel sparse compositional technique reveals microbial perturbations. mSystems. 2019 Feb 26;4(1):e00016-19.

3. Fedarko M, Martino C, Morton J, González A, Rahman G, Marotz C, Minich J, Allen E, Knight R. Visualizing 'omic feature rankings and log-ratios using Qurro. NARGAB-2019-111.R1. in press

4. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz J, Bittinger K, Brejnrod A, Brislawn C, Titus Brown C, Callahan B, Caraballo-Rodriguez A, Chase J, Cope E, DaSilva R, Diener C, Dorrestein P, Douglas G, Durall D, Duvallet C, Edwardson C, Ernst M, Estaki M, Fouquier J, Gauglitz J, Gibbons S, Gibson D, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley G, Janssen S, Jarmusch A, Jiang L, Kaehler B, Kang K, Keefe C, Keim P, Kelley S, Knights D, Koester I, Kosciolek T, Kreps J, Langille M, Lee J, Liu Y, Loftfield E, Lozupone C, Maher M, Marotz C, Martin B, McDonald D, McIver L, Melnik A, Metcalf J, Morgan S, Morton J, Naimey A, Navas-Molina J, Nothias L, Orchanian S, Pearson T, Peoples S, Petras D, Preuss M, Pruesse E, Rasmussen L, Rivers A, Robeson M, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song S, Spear J, Swafford A, Thompson L, Torres P, Trinh P, Tripathi A, Turnbaugh P, Ul-Hasan S, vanderHooft J, Vargas F, Vazquez-Baeza Y, Vogtmann E, Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber K, Williamson C, Willis A, Zu Z, Zaneveld J, Zhang Y, Zhu Q, knight R, and Caporaso JG. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature biotechnology. 2019 Aug;37(8):852-7.

Chapter 4.

Applying novel tools to better understand the oral microbiome

The final chapter is dedicated to applying the benchtop and computational tools described in Chapters 2 and 3 to gain novel insight into the human oral microbiome. Both sections are currently under peer-review. The first section applies a novel strategy of removing dead cell signal with quantitative flow cytometry in order to count the number of microorganisms present in human saliva from healthy individuals. This first-of-its-kind study revelaed that the human oral microbiome is incredibly dynamic, with the number of microbes present changing by orders of magnitude within an individual over the course of the day, highlighting the need for scale-invariant analysis tools.

The second section is an analysis of a massive dataset generated by our collaborator Dr. Ryan Demmer from the University of Minnesota. In this analysis we searched for early microbial markers of periodontitis, and their possible link to cardiometabolic health. The results presented were only possible with the novel analysis tools described in Chapter 3.

4.1

Quantifying live microbial load in human saliva samples over time reveals stable composition and dynamic load

Human saliva contains a distinct and rich community of microorganisms. While the microbial composition of saliva has been documented, the number of microorganisms present in saliva and how this number changes over time remains unclear. Furthermore, the viability of these microorganisms over time is also unknown. Here we present a novel method to characterize live microbial load in parallel with microbial community sequencing. We apply this method to unstimulated saliva samples collected in two distinct experiments; one collected longitudinally throughout the entire course of an ordinary day and the other collected across an acute perturbation.

4.1.1 Background

The human oral cavity provides a distinct microbial niche and contains a complex community of microorganisms. Because saliva is easy to sample and has relatively high biomass, it has become an increasingly popular environment to study host-microbe interactions. The composition of the human salivary microbiome has been shown to be stable over weeks [1] and months [2]. However, how the salivary microbiome is affected by the acute perturbations of daily life, including dental hygiene and eating habits, is less well understood. Determining these intricate dynamics could have major implications for cross-sectional studies that do not (or cannot) control for these variables.

While we have limited knowledge about how the microbial composition of saliva changes over time, it is also unknown how the number of microorganisms, i.e. the microbial load, changes. This is in part because it is currently relatively easy to estimate the microbial composition of a given sample through sequencing, but more challenging to quantify the microbial load of complex communities. However, elucidating the exact number of cells in and on our body is not only an important component for understanding host-microbe interactions, it can also affect the interpretation and analysis of sequencing data [3–5].

Multiple methods have been proposed for microbial load estimation, including, but not limited to, quantitative sequencing spike-ins, extrapolation of cell number from 16S rRNA gene copy number enumeration by quantitative PCR (qPCR), and flow cytometry. Each method includes advantages and disadvantages as previously described in detail [3,4]. The current gold standard for quantifying live microbial load from complex communities is using flow cytometry [3–7], although it is important to keep in mind that one of the caveats to microbial load estimation by flow cytometry is the potential presence of bacterial aggregates which may artificially deflate the cell count. This method involves staining microbes with a fluorescent DNA dye and using a clever gating strategy to exclude background signal without relying on cell size or density [6].

However, not all detectable DNA comes from living organisms. In 2016 Carini et al., demonstrated that in soil samples an average of 40% of DNA originated from non-intact cells [8]. The presence of extracellular DNA or DNA from non-intact cells, “relic DNA”, has long been recognized in human microbiome samples. For example, evaluating microbial load in sputum samples from patients with cystic fibrosis without taking viability into account resulted in inaccurate clinical conclusions [9]. This is due to the large amount of relic DNA in cystic fibrosis sputum samples [10].

Treating samples with propidium monoazide (PMA) enables the removal of DNA not protected by a cell membrane [11]. PMA, similar to propidium iodide, is a DNA intercalator that cannot pass intact cell membranes. Therefore, it only intercalates in relic DNA. Upon exposure to visible light N_2 is photolytically cleaved from the PMA molecule resulting in a highly reactive nitrene intermediate. This reactive product quickly forms a covalent bond with the DNA to which it is intercalated, inducing a DNA break. Experiments from a chemical cousin of PMA, ethidium monoazide, demonstrated that these molecules can intercalate roughly every 10-80 base pairs [12], effectively excluding the resulting DNA fragments from downstream analysis (i.e. DNA fragmented to this size are not recovered in typical DNA extraction kits, cannot be detected with typical fluorescent readouts, and cannot be amplified by PCR). Importantly, any additional PMA not intercalated in DNA reacts with H_2O and is rendered inert following light treatment. This method has been increasingly employed over the past few years to distinguish signal from live cells from relic DNA [8,9,13–19].

Here, we optimized a technique to quantify live microbial load from saliva samples using PMA and flow cytometry. This technique was applied to unstimulated saliva samples collected throughout the course of a single day and in response to an acute perturbation in order to assess saliva microbial dynamics.

4.1.2 Results

Approach to quantifying live microbial load and composition

We developed a protocol to quantify live microbial load in unstimulated saliva by removing relic DNA with PMA prior to staining for flow cytometry (Fig. 1A). Validation experiments with Gram-positive and Gram-negative bacterial cultures demonstrated a linear relationship between

flow cytometry counts and classic colony-forming units (CFU) enumeration, especially when the sample was pre-treated with PMA to remove relic DNA (Fig 1B). The ability of PMA to remove dead cell signal was further validated by heat-killing a fresh culture of E.coli. Without PMA treatment, heat-killed cells were counted at similar levels to fresh cells by flow cytometry (Fig. S1). Removing relic DNA with PMA prior to flow cytometry removed detection of these dead cells, matching the CFU counts.

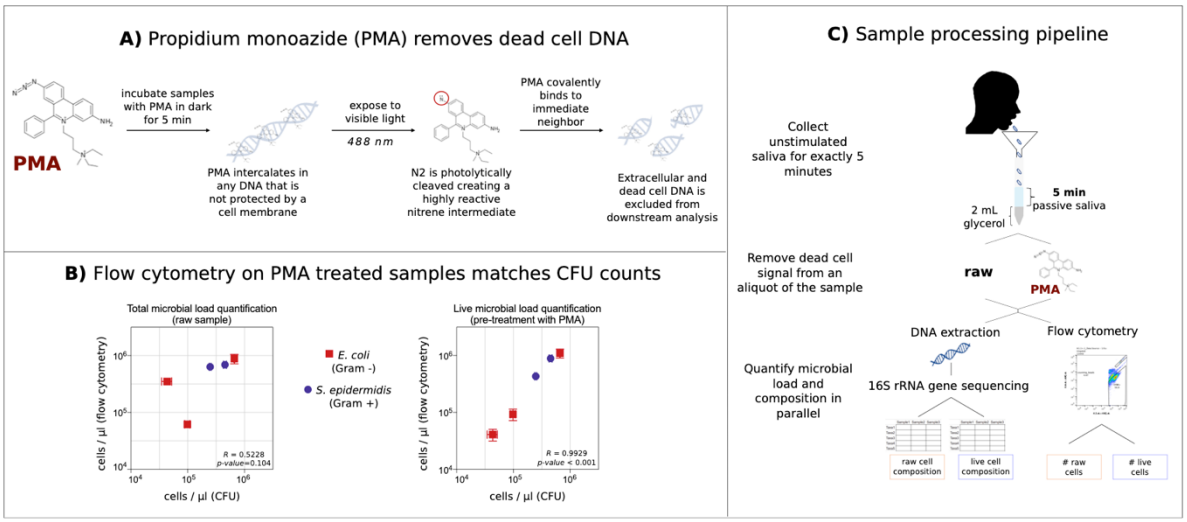

Figure 4.1.1. *Quantifying live microbial load and composition in human saliva*. (**A**) Schematic illustrating how PMA removes DNA not protected by an intact cell membrane. (**B**) Flow cytometric quantification of microbial load matches colony-forming unit (CFU) estimation more closely when the sample is pre-treated with PMA for both a gram-positive bacteria, E.coli, and a gram-negative bacteria, S.aureus. Each point represents technical triplicates of each method with the standard deviation shown for the CFU measurement (x-axis) and flow cytometry measurement (y-axis). (**C**) Schematic of sample collection and processing allowing for measurement of unstimulated salivary flow rate, live and total microbial load, and live and total microbial composition.

With this method, we set out to determine the microbial population dynamics in human saliva by quantifying live microbial load and composition throughout a single day. Specifically, we elucidated if daily perturbations (e.g. brushing teeth, eating a meal) induce ecological shifts that could negatively affect cross-sectional datasets. We collected unstimulated saliva at 9 timepoints throughout the day and in response to an acute perturbation (see summary of participant

demographics in Table 1). From each sample, we measured salivary flow rate, live microbial load, and live microbiome composition via 16S rRNA gene amplicon sequencing (Fig. 1C).

Table 4.1.1. *Summarized participant demographics by study*. For each experiment, the average and standard deviation of the age of the participants is listed, as well as the number of male and female participants and total number of samples collected.

study	treatment	average age	stdev age	# female	# male	# samples per participant	total # samples
daily dynamics	n/a	36.9	18.5	5	5	9	90
acute perturbation	water	26.9	5.1	3	4	3	21
	antiseptic mouthwash	27.9	4.3	3	4	3	21
	alcohol-free mouthwash	32	5.2	3	4	3	21
	soda	30.1	3.9	2	5	3	21
TOTAL / AVERAGE		**30.76**	**7.4**	**16**	**22**		**174**

Human saliva microbial load and viability

In total, we processed 172 samples from 38 individuals across two experiments. The average unstimulated salivary flow rate was 0.48 mL per minute (+/- 0.03 standard error of the mean), and the median flow rate was 0.38 mL per minute (range 0.02 to 1.56), similar to previously published reports [20,21]. There was no significant difference in salivary flow rate between samples from males versus females (independent T-test *p-value* = 0.398), and there was no correlation of salivary flow rate with age (Pearson R correlation = -0.003, *p-value* = 0.964). Previous studies report higher unstimulated salivary flow rate in males [21], and decreasing flow rate with increasing age [20]. The lack of statistical significance for age and gender in this study may be due to the relatively low number of participants enrolled (n=38, Table 1).

The number of live microbial cells collected in 5 minutes varied by more than 3 orders of magnitude across participants and timepoints (Fig. 2a). The percentage of live cells (calculated by dividing the number of cells obtained from the PMA-treated by raw sample) also fluctuated greatly throughout the day and across perturbations, ranging from <1% to 100% (Fig. 2b). Salivary flow rate was negatively correlated with microbial concentration in saliva (Pearson correlation coefficient = -0.326, *p-value* = 0.009) and this relationship was stronger when considering only live cells (Pearson correlation coefficient = -0.377, *p-value* = 0.003) (Fig. 2c).

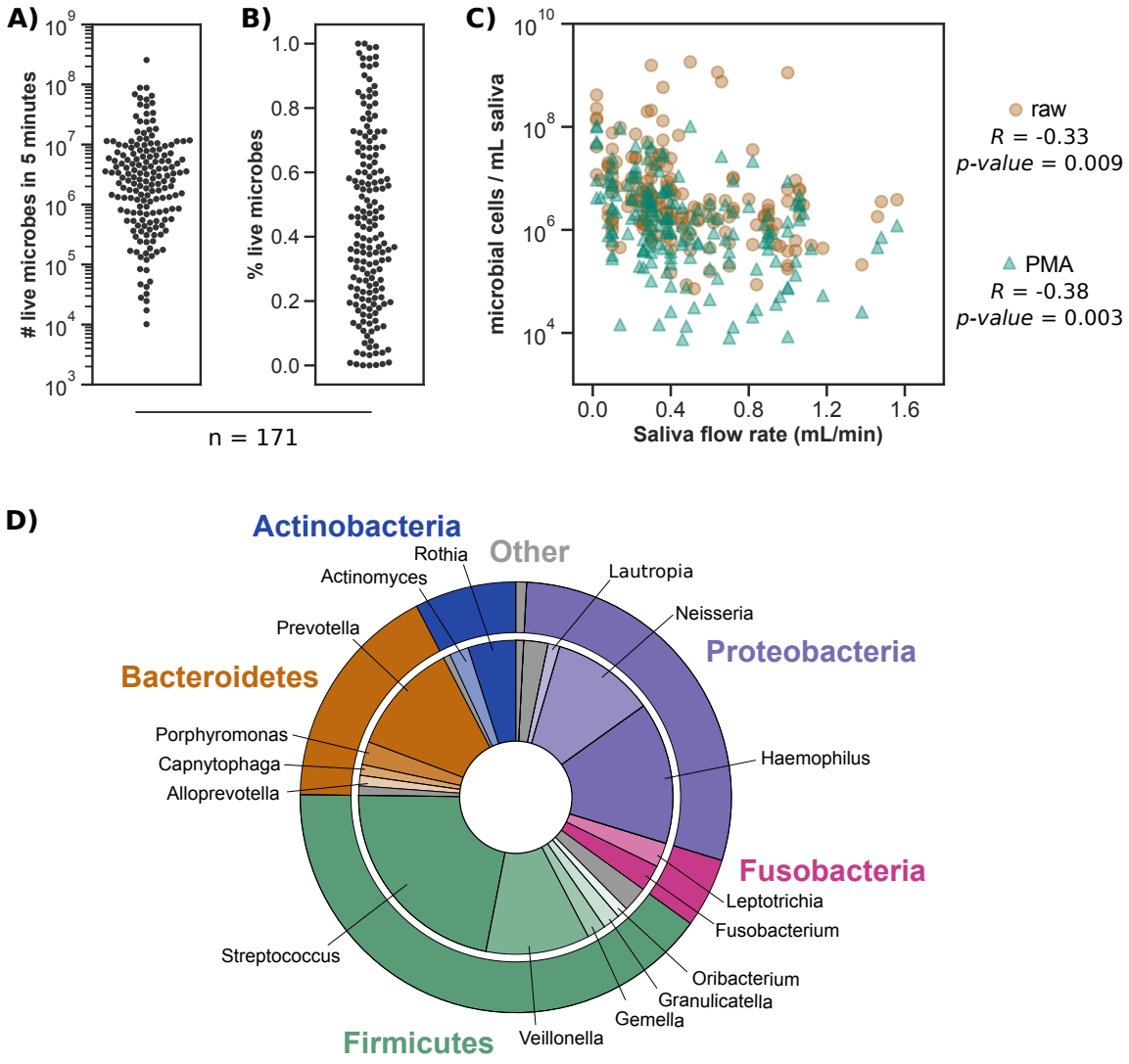

Figure 4.1.2. *Microbial load and viability vary widely across healthy participants and is negatively correlated with salivary flow rate*. Samples from both the daily dynamics and acute perturbation study are shown combined here (n=172). **A**) Total number of live microbial cells detected in unstimulated saliva collected for 5 minutes determined by PMA-treatment and flow cytometry. **B**) Percentage of live cells (number of cells after PMA treatment / number of cells detected in raw sample using flow cytometry) in saliva samples. **C**) Number of microbial cells per mL determined by flow cytometry on the y-axis by salivary flow rate in mL per minute on the x-axis. Statistics represent Pearson's correlation coefficient of log-transformed microbial concentration data (log(number of cells per mL)) and salivary flow rate (mL per min). **D)** Pie chart displaying average relative abundance of the top 5 most abundant phyla (representing >99% of the data) and respective genera with >1% relative abundance (representing >90% of the data).

We also estimated bacterial concentration via qPCR of the 16S rRNA gene (Fig. S2) [22]. Quantification by flow cytometry and qPCR were significantly correlated in both the raw (Pearson's r = 0.272, p=0.011) and PMA treated samples (Pearson's r = 0.333, p=0.002) (Fig. S2a). These results are similar to a previous study by Vandeputte et al., when comparing flow cytometry to qPCR [3]. This relatively low correlation coefficient is not unexpected given that qPCR and flow cytometry have different biases when determining absolute quantification as previously discussed [3,4]. However, salivary flow rate was not significantly correlated with 16S rRNA gene copy number determined by qPCR in either the raw samples (*p-value* = 0.753) or PMA treated samples (*p-value* = 0.106) (Fig. S2b).

Following quality control filtering, the median sequencing depth was 33,999 reads per sample, with a median of 147 amplicon sequence variants (ASVs) per sample. When collapsed across all participants, >99% of the entire dataset belonged to 5 phyla (Fig. 2d). Within these 5 phyla, 16 genera had an average relative abundance across samples >1% and together made up >90% of the data. Shotgun sequencing of the daily dynamic samples (n=88) provided nearly identical results, where the same top 5 phyla made up >99% and the same top 16 genera made up 89% of the taxonomic hits from Metaphlan2. These results are consistent with previous work looking at a range of human oral cavity sites, including tongue, mucosa, and supra- and subgingival plaque, corroborating the hypothesis that saliva microbial composition is a mix of dissociated cells across all these diverse niches of the human oral cavity [23–26]. Furthermore, this suggests that across orders of magnitude in biomass (Fig. 2a), saliva microbial composition is highly conserved across individuals and time.

Daily dynamics of the human saliva microbiome

To assess the effect of daily perturbations on the saliva microbiome, we recruited 10 individuals to collect 9 saliva samples throughout the day (summarized demographics under 'daily dynamics' study in Table 1, full demographics available online: https://github.com/knightlab-analyses/Saliva_quantification_study/tree/master/ipynb). Participants were asked to collect unstimulated saliva first thing in the morning, before brushing their teeth or eating, then again after brushing their teeth in the morning, and then roughly every 2 hours throughout the day. We asked participants to report any additional oral hygiene events and what they ate and drank throughout the day.

The number of live cells collected in 5 minutes changed dramatically even within a given individual (Fig. 3a). When normalized to the first time point, it is clear that the highest microbial load was obtained immediately upon waking (before brushing teeth) (Fig. 3b). Samples collected early in the morning (the first two timepoints collected before and after brushing teeth) tended to have lower salivary flow rates than samples collected later in the day (*p-value* = 0.015), in line with decades of research showing markedly lower salivary flow rates during sleep [27].

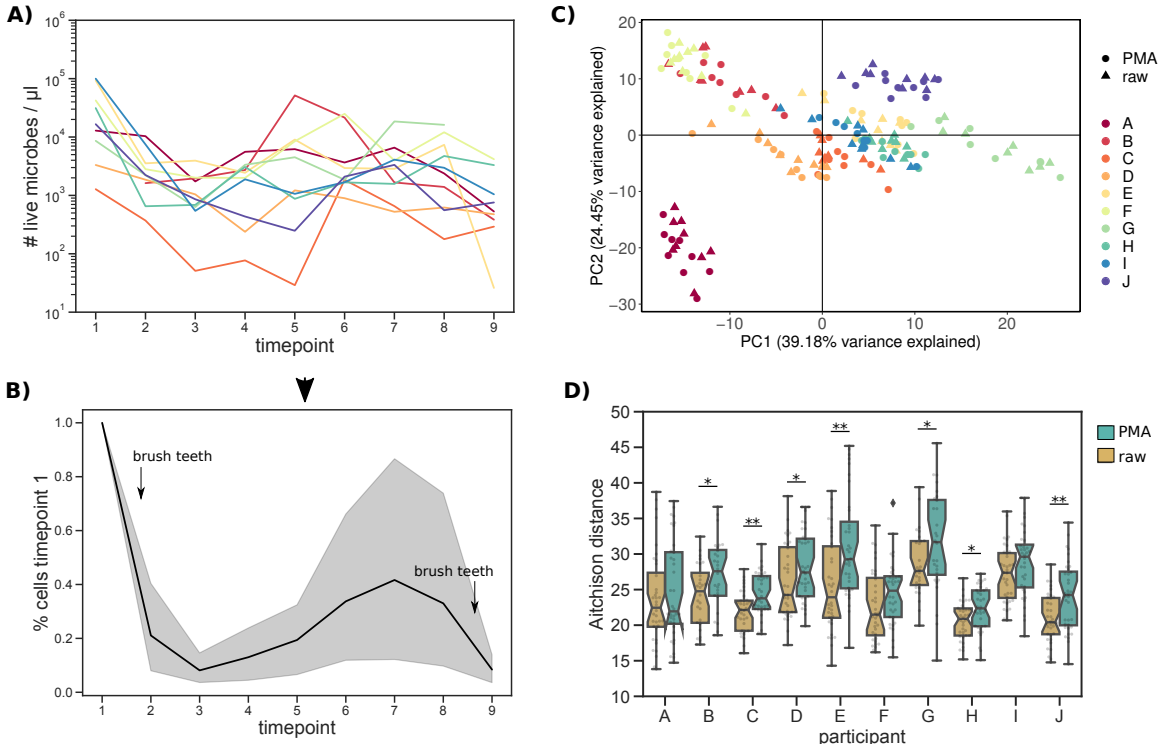

Figure 4.1.3. *Daily dynamics of the human saliva microbiome*. **A**) Live microbial cells collected in unstimulated saliva (for 5 minutes) throughout the course of a day starting right after waking and before going to bed. **B**) Same data as A) normalized per individual to time point 1 representing 100%. **C**) PCoA of the Aitchison distance matrix reveals clustering by individual (colors) rather than treatment (shape). **D**) Aitchison distance across samples from the same individual over time. Statistical significance represents Kruskal-Wallis test between raw and PMA treated samples for each participant; * ≤0.05, **≤0.01, ***≤0.001.

We wanted to identify potential changes in saliva samples collected after eating. We compared saliva samples collected after eating to all other samples and found a significant increase in unstimulated salivary flow rate following eating (0.537 +/- 0.039 vs 0.43 +/- 0.030 mean mL per minute +/- SEM; independent T-test p=0.027). However, we found no significant difference in the number of either live (Kruskal-Wallis *p-value* = 0.524) or total microbial cells (Kruskal-Wallis *p-value* = 0.957) collected in 5 minutes. We also evaluated the percentage of raw 16S rRNA gene amplicon sequencing reads aligning to chloroplast 16S rRNA genes. Nearly half (49%) of the saliva samples collected after eating had chloroplast reads, compared to only 16% of the

remaining saliva samples. Of these chloroplast positive samples, the average relative abundance was 0.6% in samples collected after eating, and only 0.04% in the remaining saliva samples. Furthermore, there were 4 times as many raw samples containing chloroplast reads as PMA-treated samples (n=42 versus n=10, respectively), indicating that most of the chloroplast DNA was extracellular or from non-intact chloroplasts. All chloroplast reads were filtered from the table for downstream analysis.

We next assessed β-diversity in the quality filtered 16S rRNA gene amplicon data using the Aitchison distance metric [28], which calculates distance on centered-log ratio (clr) transformed data making it invariant to differences in total microbial load across samples unlike other traditionally used β-diversity metrics. Aitchison β-diversity was mostly driven by participant, more than gender, processing method (i.e. raw versus PMA-treated samples), or whether the sample was collected after eating (Fig. 3c; PERMANOVA pseudo-F statistic by participant = 26.9, by gender = 13.6, by processing method = 2.9, by food intake = 1.89).

To confirm the robustness of these results, we performed shotgun sequencing on the PMA treated samples since they were processed with our lyPMA protocol and were therefore depleted in host DNA [29]. In line with our previous host depletion experiments in saliva, the median percentage of shotgun sequencing reads aligned to the human genome was 9.8%. Metaphlan2 [30] was used to assign taxonomy to the species level, and β-diversity was assessed on the resulting table again using the Aitchison distance metric. As observed in the 16S rRNA gene amplicon data, the average distance among different individuals at the same timepoint was greater than the average distance within an individual across timepoints (Fig. S3 Kruskal-Wallis *p-value*<0.001), despite large changes in microbial load. We also assessed the functional pathways present across timepoints and participants using HUMAnN2 and the MetaCyc database [31,32]. A total of 355

unique pathways were identified. Once again, the Aitchison distance on the pathway table revealed that samples from the same participant over time are functionally more similar than samples across different participants collected at the same time point (Fig. S3 Kruskal-Wallis *p-value*<0.001). The distance separating samples collected from the same individual over time versus among different individuals was largest in the 16S rRNA gene amplicon data (at the amplicon sequence variant level), smaller in the MetaPhlAn2 data (at the species level), and smallest in the HUMAnN2 data (at the functional pathway level) (Fig. S3). This finding is in line with the ecological theory of functional equivalence (or functional redundancy) underpinning a stable ecosystem. The microbial community in saliva is remarkably stable across daily perturbations and massive changes in microbial load, and the exact organisms in the population fluctuate more than the functional capabilities of the entire ecosystem.

We next sought to determine how removing relic DNA with PMA affected β-diversity using the 16S rRNA gene amplicon sequencing data (since only PMA treated samples were used for shotgun sequencing). We found that raw and PMA-treated samples from the same saliva sample were more similar than samples from the same person processed the same way over time (Kruskal Wallis p<0.001), indicating the relatively subtle effects of relic DNA removal. However, we also found that samples collected from the same individual over time were more dissimilar when relic DNA was removed with PMA compared to the raw samples (Fig. 3D). This mirrors recent findings from Carini et al., who found that relic DNA removal with PMA enhanced their ability to distinguish longitudinally collected soil microbial communities [33]. These findings suggest that relic DNA removal can improve resolution of longitudinally collected microbial communities across a wide range of environments.

Effect of acute perturbation on saliva microbial load and composition

To assess the effect of a controlled, acute perturbation on the saliva microbiome, we recruited 28 healthy individuals and randomly assigned them to 4 groups (summarized demographics under 'acute perturbation' study in Table 1, full demographics available online). Each group was asked to either swish for 30 seconds with water, antiseptic mouthwash, or alcohol-free mouthwash, or to drink a can of soda. Live microbial load and composition was assessed before (timepoint 1), 15 minutes after (timepoint 2), and 2 hours after (timepoint 3) the assigned perturbation.

To identify the effects of each treatment on microbial load, we calculated the number of non-intact ('dead') cells in each sample by subtracting the number of PMA-treated ('live') cells from the number of untreated total cells detected by flow cytometry (Fig. 4A). The group that swished with water and the group that drank a soda had no significant change in microbial load across any of the timepoints (one-way ANOVA with Tukey's multiple comparisons test on zero-centered (log transformed) data *p-value* > 0.05). In the group that swished with antiseptic mouthwash, there was significantly more dead microbial cells at timepoint 3 compared to live microbial cells at timepoint 2 (*p-value* = 0.044). The strongest effect was in the group that swished with alcohol-free mouthwash where there was an insignificant trend towards increased dead microbial load (*p*-value = 0.156) and a significant reduction in the number of live microbial cells directly after treatment (*p-value* = 0.001). Strikingly, live microbial load returned to baseline levels within 2 hours (Fig 4A).

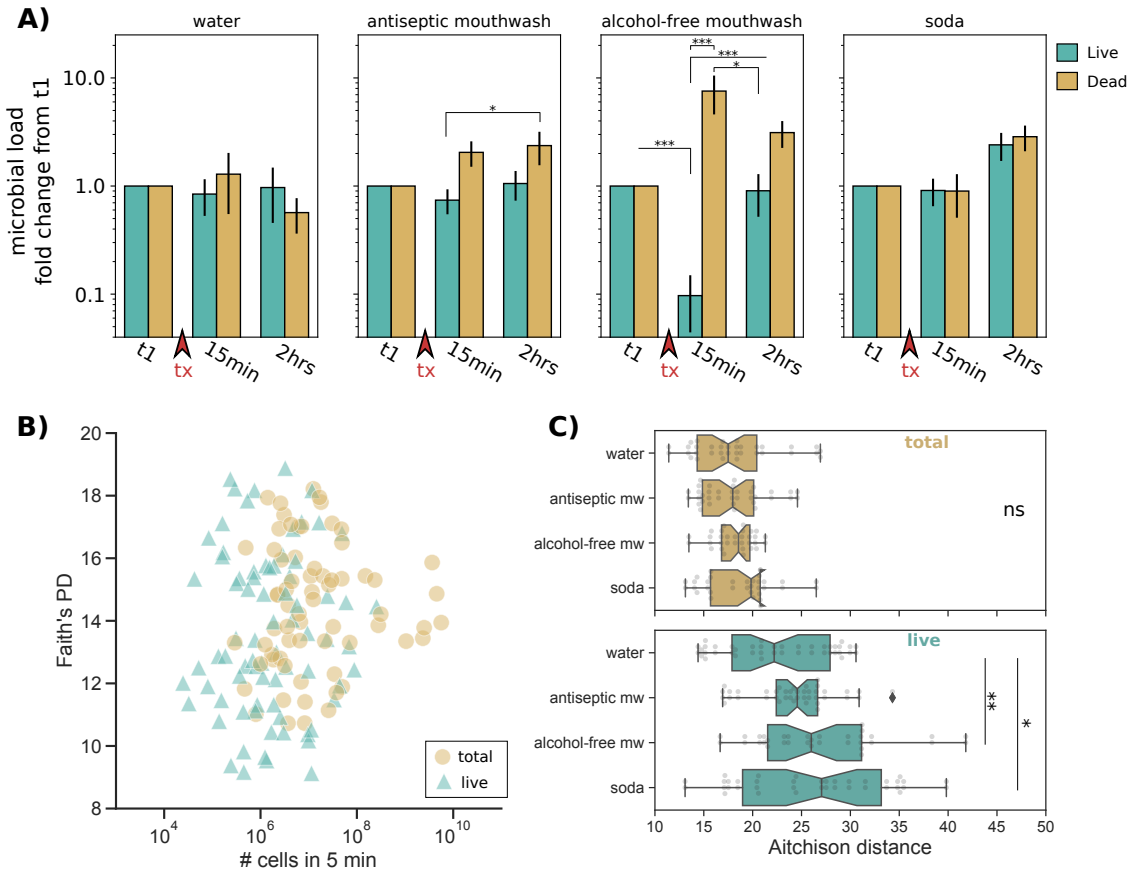

Figure 4.1.4. *The effect of acute perturbation on live and dead microbial load and composition*. **A)** Change in microbial load following treatment over time. The y-axis shows percentage of live (green) and dead (yellow) microbial number detected by flow cytometry normalized per-individual to the first time point (n=7 per group). Significance represents one-way ANOVA with Tukey's multiple comparison test; *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$. **B**) Each dot represents a saliva sample processed with (green) or without (yellow) PMA to remove relic DNA (revealing live and total microbial populations, respectively). The y-axis is Faith's phylogenetic diversity rarefied to 10,000 sequences, and the x-axis is the number of microbial cells collected in 5 min unstimulated saliva detected by flow cytometry (Pearson's correlation coefficient for both live and total populations *p-value* >0.05). **C**) Aitchison distance on 16S rRNA gene amplicon sequencing data between samples from the same individual over time plotted by treatment group. Overall, distance among samples treated with PMA ('live', green, bottom graph) were more dissimilar than the distances between the same samples not treated with PMA ('total', yellow, top graph) (bootstrapped Kruskal-Wallis *p-value* ≤ 0.001). In the raw samples, there was no significant difference among the treatment groups (bootstrapped Kruskal-wallis $p \geq 0.05$). In the PMA-treated samples, participants that swished with alcohol-free mouthwash or drank a soda had significantly greater variation over time compared to the group that swished with water (bootstrapped Kruskal-Wallis *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$).

Throughout these massive fluctuations in microbial load, there was no correlation between alpha-diversity (measured by Faith's phylogenetic diversity) and the number of microbial cells

(Pearson R > 0.2 for both live and total cell counts), again suggesting that although there are major fluctuations in microbial load, the overall saliva microbial composition is relatively stable (Fig 4B).

This finding is supported by the fact that β-diversity was largely driven by participant (Aitchison distance matrix PERMANOVA pseudo-F statistic by participant = 9.1, by gender = 3.6, by processing method = 4.5, by timepoint=0.851). We were interested to see if any of the four treatments groups caused changes in bacterial composition. To assess this, we calculated the average β-diversity (Aitchison distance) among samples collected from a single participant. In the untreated samples, there was no difference in β-diversity across the four treatment groups (Fig 4C). However, when relic DNA was first removed with PMA, there was significantly higher β-diversity distances across samples from participants who swished with alcohol-free mouthwash (one-way ANOVA with Tukey's multiple comparisons test *p=0.007*) or drank a soda (*p*=0.042) compared to the group that swished with water (Fig. 4C). Further, mirroring our observation that microbial load returns to baseline levels within 2 hours, we found that in the group that swished with alcohol-free mouthwash live microbial composition at timepoint 1 was more similar to timepoint 3 than to timepoint 2 (*p-value* = 0.017) (Fig. S4). This suggests that both microbial load and microbial composition are recovered within 2 hours following a major perturbation in microbial load.

4.1.3 Discussion

Here we present a method to perform parallel live microbial load quantification and sequencing. This method combines two established protocols; PMA-treatment to remove relic DNA and microbial flow cytometric quantification. PMA-treatment can be performed without any specialty lab equipment, and the flow cytometry protocol requires a standard flow cytometer equipped with a 488 nm laser and standard detectors. We applied this method to unstimulated saliva samples collected longitudinally.

To our knowledge, this represents the first longitudinal human microbiome dataset with matched 16S rRNA gene amplicon sequencing and flow cytometry quantification, each with total and live (intact) cell evaluation. In addition to providing unprecedented information about the dynamics of the human saliva microbiome, we expect that this dataset will be useful to the community for evaluating novel algorithms [4,7,34,35] to describe and predict ecological shifts in human microbiome samples.

Using this dataset, we demonstrated an inverse relationship between salivary flow rate and microbial load. This could help explain why decreased salivary flow rate is often associated with microbial-derived periodontal diseases, such as caries and gingivitis [36–39]. Interestingly, microbial concentration is also negatively correlated with water content in stool samples [3]. Together, these findings demonstrate the ability of the human body to modulate microbial concentration at mucosal sites.

We found that live microbial load in saliva fluctuates by orders of magnitude throughout a typical day. Despite this fluctuation, taxonomic composition is remarkably consistent across time and within individuals, with more than 90% of all 16S rRNA gene sequences coming from 16 genera across 5 phyla, similar to findings from previous studies [23–26].

The compositional effects of removing relic DNA with PMA were relatively subtle, as this processing step affected taxonomic composition much less than the effects of time or different individuals. However, we did find that PMA treatment revealed greater dissimilarity among samples collected longitudinally from the same participant. Furthermore, we found a clear example (alcohol-free mouthwash) where the assumption that all DNA comes from living organisms can lead to false conclusions; without taking relic DNA into account it would appear that microbial load increases after alcohol-free mouthwash, although this treatment decimated the microbial cell population in saliva. This suggests that while PMA treatment may not greatly influence the results of cross-sectional oral microbiome datasets, it could add greater resolution to longitudinal studies and potentially reveal patterns only detectable in the absence of relic DNA. Interestingly, these findings mirror recent results in a soil microbiome dataset. Carini et al., found that relic DNA removal improved their ability to detect changes in underground microbial community over time [33]. Together these findings highlight how relic DNA can obscure changes in community composition across a broad range of microbial environments.

Participants in saliva microbiome experiments are often asked to avoid eating or oral hygiene anywhere from 2 hours to 24 hours before sampling, based on the assumption that these daily perturbations could influence the oral microbiome. Because of the detailed metadata collected in our longitudinal experiment, we were able to determine the effect of eating and oral hygiene on the saliva microbial load and composition as detailed below.

Salivary flow rate was significantly higher in samples collected after eating, but the total number of cells detected in 5 minutes remained unchanged. We also found an increase in the presence and abundance of chloroplast sequences in saliva samples collected after eating, which are presumably from food sources. Chloroplast and mitochondrial reads are typically filtered out

of 16S rRNA gene amplicon datasets from human microbiomes, and after removing these sequences we found no detectable alteration in the salivary microbiome composition after eating. This suggests that the acute effect of eating has a minimal impact on the saliva microbiome.

While oral hygiene significantly reduced microbial load, the composition remained unperturbed when using analytical tools that account for compositional bias [4]. The decrease in microbial load was detectable after brushing teeth in both the raw and PMA treated samples, suggesting the physical removal or bacterial cells. In contrast, swishing with alcohol-free mouthwash permeabilized cells but did not physically remove them, since the decrease in microbial load was only detectable after removing dead cell signal with PMA. We were surprised that the alcohol-free mouthwash had such a dramatic effect on microbial viability, but not antiseptic mouthwash. This may be due to the presence of natural and artificial compounds with antimicrobial activity found in the alcohol-free mouthwash used in this study (e.g. poloxamer 407, thymol, eucalyptol, menthol, salicylate, lauryl sulfate). This finding highlights the utility of using PMA to remove dead cell signal.

Together, these results showcase the remarkable resilience of the salivary microbiome and highlight the potential of the human salivary microbiome as an easily accessible model to probe complex community organization and response. Despite dramatic perturbations in the number of microbes in saliva, taxonomic compositions remain relatively stable and are strongly host-specific, which lends credence to cross-sectional saliva microbiome studies when samples are collected at different times of day.

4.1.4 Conclusion

Here we present a novel method to assess live microbial composition and load. We identified a negative correlation between salivary flow rate and microbial load, which has implications for oral health. The numbers of microbes in human saliva varied by orders of magnitude throughout a single day and in response to daily perturbations. Furthermore, by removing relic DNA signal with PMA treatment prior to quantification we were able to identify the antimicrobial effect of alcohol-free mouthwash which was missed when quantifying the raw sample. These findings highlight the importance of using analytical tools that account for compositional bias, which is exacerbated by comparing samples with different microbial loads [3–5, 40, 41]. Despite these changes in absolute abundance, we found that microbial composition is extremely stable and host-specific, and not significantly affected by meals or oral hygiene. These results highlight the importance of using scale-invariant tools in the analysis of cross-sectional datasets and demonstrate the ability of relic DNA removal to increase resolution of longitudinal studies.

4.1.5 Methods and Materials

Participant Recruitment and Saliva Collection

Self-described healthy volunteers were recruited in accordance with IRB 150275. Participant demographics are detailed in the supplementary metadata files and summarized in Table 1.

Each participant was given a kit with a funnel (Simport, catalog# F490-2) and 15 mL collection tube (Corning, catalog# 352057) loaded with 2 mL 40% sterile glycerol for each timepoint to be collected. Glycerol was used to preserve the bacterial cells from freezing for

downstream PMA treatment [29]. The final concentration of glycerol varied from 8% to 38%, but this did not correlate with microbial viability (Fig. S5, Pearson R = 0.36), suggesting that variability in final glycerol concentration did not influence our ability to detect live cells in cryopreserved samples. The instructions for unstimulated saliva collection and salivary flow measurement were adapted from Navesh et al. [42]. Briefly, participants were asked to find a comfortable seat, set a timer for 5 minutes, swallow and start the timer. With the head tilted down slightly and lips against the edge of a food-grade funnel, they were asked to relax the jaw, mouth and tongue allowing saliva to pool and eventually drain into the tube. Participants were asked not to swallow during the 5-minute collection. After collection, the lids of the tubes were closed, and the funnel was disposed. The tubes were gently inverted ten times to mix the saliva and glycerol and placed at -20. Samples were brought to the lab the following day on wet ice and stored at -20 until further processing.

In the daily dynamics experiment, 10 individuals were recruited to collect unstimulated saliva throughout a single day. Participants were asked to collect saliva on their own (as detailed above) at 9 time points throughout the day; immediately upon waking (before eating or brushing teeth), soon after brushing teeth, and roughly every 2-3 hours throughout the day. Participants were asked to record the time of each saliva collection and whether they ate or performed oral hygiene since the last sample collection. Of the 90 samples, two were excluded because of missing sample collections, totaling 88 samples.

In the acute perturbation experiment, 28 individuals were recruited to collect unstimulated saliva before, 15 minutes after, and 2 hours after an acute treatment. Participants were asked not to eat or drink (except water) for one hour prior to the first collection. Each participant was blindly assigned to one of 4 groups; rinsing with bottled water (Simple Truth water, Kroger, pH 7.6),

rinsing with mouthwash (Listerine Antiseptic Cool Mint), rinsing with alcohol-free mouthwash (Listerine Zero Alcohol Cool Mint), or drinking a 12 oz. can of Coca-Cola soda. For the rinsing samples, participants were asked to swish 20 mL solution in their mouths for exactly 30 seconds, and to refrain from gargling in the back of their mouths. For the soda samples, participants were asked to drink the 12 oz can within 10 minutes.

Bacterial culturing for proof of concept experiments

Escherichia coli and *Staphylococcus epidermidis* cultures were used for proof of concept experiments to validate the quantitative flow cytometry protocol. Cultures of *S. epidermidis* or *E. coli* were grown overnight in Tryptic Soy Broth (TSB) at 37°C. Bacterial cultures were pelleted by centrifugation at 8,000 g for 5 minutes and resuspended in 1x sterile phosphate-buffered saline (PBS). Resuspended cultures were run across 5 μm syringe filters (Sartorius Stedim Biotech GmbH) and diluted between 100x – 10,000x fold in PBS. Samples treated with PMA were mixed with 10 μM PMA, vortexed briefly, incubated at room temperature protected from light for 5 minutes, then exposed to light by laying on ice <20 cm from a benchtop fluorescent light bulb for 25 minutes. Both raw and PMA samples were stained with 0.1X SYBR green (SYBR™ Green I Nucleic Acid Gel Stain, Invitrogen) and incubated in the dark for 15 minutes at 37°C. Flow cytometry was performed using the Sony SH800 and AccuCount Fluorescent Particles as described below. Colony forming unit (CFU) measurements were performed by plating triplicate 10 μl drops of 10-fold serial dilutions of the bacterial culture on TSB agar plates and incubating at 37°C overnight. The following day dilutions containing between 10 – 50 colonies per drop were counted. For the heat-killed experiment (Fig. S1), overnight *E.coli* cultures were heated to 65°C for 10 minutes and analyzed as described above.

Flow cytometry

Cryopreserved, unstimulated saliva samples were thawed on ice and centrifuged at 3,000g for 1 min to remove any bubbles and allow for accurate volume assessment. The volume (to the closest 0.1 mL) and the weight of each sample was recorded.

Unstimulated saliva samples were diluted tenfold with sterile, 1x PBS. To remove human cells and salivary debris, samples were filtered using a sterile 5 μm syringe filter (Sartorius Stedim Biotech GmbH). Relic DNA was removed from an aliquot of each sample; A final concentration of 10 μM PMA (Biotium) was added to one mL of diluted saliva, vortexed briefly, and incubated at room temperature protected from light for 5 minutes. Samples were lain horizontally on ice <20 cm away from a benchtop fluorescent light bulb for at least 25 minutes, and vortexed briefly every ~5-10 minutes.

An aliquot of diluted, filtered saliva (raw) and the PMA-treated aliquot were then stained in parallel with SYBR green for detection by flow cytometry. 5 μl 20x SYBR green (SYBR™ Green I Nucleic Acid Gel Stain, Invitrogen) was added to 1 mL of the microbial suspension (0.1x final concentration) and incubated in the dark for 15 minutes at 37°C. Finally, 50 μl AccuCount Fluorescent Particles (Spherotech, ACFP-70-10) were added for quantification of microbial load. Samples were processed on a SH800 Cell Sorter (Sony Biotechnology) using a 100 μm chip with the threshold set on FL1 at 0.06%, and gain settings as follows; FSC=4, BSC=25%, FL1=43%, FL4=50%. For the acute perturbation experiment, the threshold was increased to FL1 0.58% to reduce the background signal that was observed in a small number of samples. The gating strategy was adapted from Props et al. [6] and an example is shown in supplementary figure S6. Briefly, fluorescent microbial cells were gated from background on a FL1-FL4 density plot. Aggregates were excluded by taking the linear fraction on a graph of height versus width of the FL1 signal

and remaining background was removed by eliminating large events detected on a forward scatter vs backward scatter density plot. Negative controls (sterile PBS stained identically to samples) were run between each sample set to exclude cross-contamination. The final calculation of cells per μl was performed per the manufacturer's instructions of the AccuCount counting beads and taking into account the dilution factor from the glycerol preservative.

Microbial load estimation with qPCR

To create a standard ladder for 16S rRNA gene copy number extrapolation, gDNA from *Escherichia coli* was amplified with the KAPA HiFi HotStart ReadyMixPCR Kit (cat# KK2602) using the Bakt 341F-805R 16S rRNA gene amplicon primers [43] (Bakt_341f: 5'-CCTACGGGNGGCWGCAG-3', and Bakt_805R 5'-GACTACHVGGGTATCTAATCC-3'). Amplification was performed in triplicate 20 μl reactions containing 10 μl KAPA Readymix, 1 μl primer mix containing 5 μM forward and reverse primers, 2 μl gDNA, and 7 μl H2O. The PCR mix was cycled through the following temperatures on a BioRad CFX instrument: 95 for 5 min, 30 cycles of 95°C for 30 sec and 60°C for 30 sec, then 4°C hold. The PCR product was run on a 1.5% agarose gel and the band was excised and purified using the QIAquick Gel Extraction Kit. Amplicon concentration was quantified in triplicate using Thermo Fisher Qubit Fluorometric Quantitation.

To estimate microbial load, sample gDNA extracted from 200 μl saliva was evaluated in triplicate with KAPA Universal qPCR Master Mix (cat# KK4828) using the Bakt 341F-805R primers listed above. Amplification was performed in triplicate 10 μl reactions each containing 5 μl KAPA MasterMix, 0.5 μl primer mix containing 5 μM forward and reverse primers, 2 μl gDNA, and 2.5 μl H2O. The PCR mix was cycled through the following temperatures on a Roche

LightCycler® 480 instrument: 95 for 5 min, 40 cycles of 95°C for 30 sec and 60°C for 30 sec, then 4°C hold. Triplicate, ten-fold serial dilutions of the standard ladder described above ranging from 1.3 to 1.3E+07 copies were run in parallel and used to extrapolate the number of 16S rRNA gene copies in the saliva samples [22].

PMA treatment and DNA extraction

For the daily dynamics cycle experiment, 500 μl aliquots of each saliva sample was set aside for standard DNA extraction. 1 mL aliquots were used for lyPMA treatment, which depletes human DNA and dead microbial signal [29]; samples were centrifuged at 10,000g for 8 minutes. The supernatant was removed and the cell pellet was resuspended in 200 μl sterile, pure, H_2O and allowed to sit at room temperature for 5 minutes to selectively lyse human cells. PMA was added to a final concentration of 10 μM, vortexed, and left in the dark at room temperature for 5 minutes. Samples were lain horizontally on ice <20 cm away from a benchtop fluorescent light bulb for at least 25 minutes, and vortexed briefly every ~5-10 minutes. The raw and PMA treated aliquots were frozen at 20°C until DNA extraction with the QIAGEN PowerSoil MagAttract DNA kit as previously described [44]. A subset of lyPMA samples with low microbial load failed in sequencing, and the PMA treatment was repeated on samples with up to 1.5 mL of unstimulated saliva. Quality control analysis of the sequencing data showed that these reprocessed samples were similar to the matched raw samples, and this higher volume was used in the follow-up acute perturbation experiment.

For the acute perturbation experiment, samples were processed in a high throughput manner. First, samples were vortexed thoroughly (15 seconds) and 1.5 mL was transferred into a 96 deep-well plate. Cells were pelleted by centrifugation at 3,200g for 15 minutes. 1 mL

supernatant was removed with a multichannel pipette, then a single channel pipette was used to remove the remaining supernatant. Pellets were resuspended in 200 μl sterile, DNAse free H_2O by pipetting up and down ~12 times and left to sit at room temperature for 5 min to allow for selective lysis of mammalian cells. The plate was then briefly centrifuged (~1,000 g for 1 minute) and 200 μL of each sample was transferred to a 96 well round-bottom plate (Greiner Bio-one cat# 650101). 2 μL of 1 mM propidium monoazide (PMA) was added to each sample for a final concentration of 10 μM. The plate was covered with a transparent seal (Bio-rad Microseal 'B' Seal cat# B0443499), vortexed briefly, and left to sit at room temperature in the dark for 5 minutes. The plate was then placed on ice <20 cm away from a fluorescent bulb light source. Light exposure took place for ~30 minutes with vortexing every ~5-10 minutes. Immediately after light exposure, the samples were briefly centrifuged (~1,000 g for 1 minute) and transferred to a QIAGEN Magattract 96 well plate and extracted as previously described [45].

16S rRNA gene amplicon sequencing and analysis

Sequencing: gDNA was processed for 16S rRNA gene amplicon sequencing using primers against the V4 region of the 16S rRNA gene 515F-806R according to the Earth Microbiome Project protocol [46]. The pooled library was sequenced on the Illumina MiSeq with a paired-end 150 V2 kit.

Quality-control: Data was processed using qiime2 [47]. Demultiplexed sequences were quality filtered for q-score with default settings and processed with deblur [48] trimmed to 150 bp. Samples with less than 1,000 quality-filtered sequences were dropped from downstream analysis. Following quality control, 6 samples dropped out of the acute perturbation experiment (n=78). In order to remove mitochondrial and chloroplast reads, sequences were aligned to the

GreenGenes database [49] and all sequences aligning to mitochondrial and chloroplast reads were filtered out using using the taxonomy-based filtering command '*qiime taxa filter-table*'.

Taxonomic assignment: Microbial taxonomy was assigned to the quality filtered sequences using the Human Oral Microbiome Database database v15.1 (HOMD) [50] with the '*qiime feature-classifier classify-sklearn'* command on a scikit-learn classifier created from the HOMD [51].

α and β-diversity analysis: Qiime2 was used to calculate α and *β*-diversity using the '*qiime diversity'* commands. For α-diversity, an α-rarefaction curve was generated to determine the appropriate subsampling sequencing depth (Fig. S7).

Shotgun sequencing and analysis

For the daily dynamics study, gDNA from lyPMA treated samples [29] was quantified with Quant-iT™ PicoGreen™ dsDNA Assay Kit (ThermoFisher Scientific), and 1 ng of input DNA was used in a 1:10 miniaturized Kapa HyperPlus protocol. For samples with less than 1 ng DNA, a maximum volume of 3.5 µl input was used. Equimolar amounts of each sample was pooled and the library was size selected for fragments between 300 and 700 bp on the Sage Science PippinHT. The pooled library was sequenced as a paired-end 150-cycle run on an Illumina HiSeq2500 v2 run at the UCSD IGM Genomics Center. Demultiplexed reads were quality filtered with TrimGalore v0.4.2 [52]. Reads aligning to the host genome (GRCh38.p7) were identified using Bowtie 2 v2.3.0 [53] with parameters set by the flag -very-sensitive-local and removed from the analysis (median percent of host aligned reads 9.8%, similar to previously reported numbers using the lyPMA method [29]). Samples with less than 10,000 microbial reads were excluded from the analysis, leaving a total of 71 samples with a median of 348,242 quality-filtered microbial reads per sample.

Taxonomic assignment was performed with MetaPhlAn v2.0 [30] using the default parameters. Functional assignment was performed with HUMAnN2 [31] using the default parameters.

Statistical analyses

All quality filtered tables and the code written to produce the figures and statistical tests presented in this manuscript can be viewed and reproduced using Jupyter iPython Notebooks through github at: https://github.com/lisa55asil/Saliva_quantification_studies.

For significance testing based on distances from sequencing data, a permutation test was used (perm_test.py in github repository). This was chosen since univariate statistical tests often assume that observations are independently and identically distributed, which is not the case with distance calculations. Similar to PERMANOVA, the group labels were shuffled, and a Kruskal-Wallis test was applied. P-values were calculated by (#(K > Kp) + 1) / (number of permutations + 1) where K is the kruskal-wallis statistic on the original statistic and Kp is the Kruskal-Wallis statistic computed from the permuted grouping. 1000 permutations were used for the permutation test.

4.1.6 Declarations

Ethics approval and consent to participate

Self-described healthy volunteers were recruited in under IRB 150275, approved by the UCSD Human Research Protections Program under Federal wide Assurance number, FWA00004495.

Consent for publication

Consent from each participant was obtained in accordance with IRB 150275.

Availability of data and material

All raw sequencing data collected in this study are available through Qiita [54] under study ID 11896 (daily dynamics) and 11899 (acute perturbation) and the European Nucleotide Archive under accession number ERP111447 (daily dynamics), and accession number ERP117149 (acute perturbation).

Additionally, all metadata and the processed, quality filtered sequencing tables as described in the methods are available in the supplementary tables, and all analyses for figure generation and statistical analyses are publicly available through github: https://github.com/lisa55asil/Saliva_quantification_study

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded in part by the Army Research Office under grant number W911NF1810158. CM was funded by NIDCR NRSA F31 Fellowship 1F31DE028478-01.

Authors' contributions

CM, RK, and KZ designed the experiment. CM optimized the protocol, ran the experiment, analyzed the data, and wrote the manuscript. JM assisted in the analysis. PN performed proof of concept experiments.

4.1.7 Acknowledgements

We would like to thank Cristal Zuniga for assistance with data visualization, and Julia Toronczak, MacKenzie Bryant, and Elaine Guo for assistance processing the flow cytometry experiments.

Chapter IV, part 1, is currently under review for publication: Marotz C, Morton J, Navarro P, Knight R, Zengler K. *Quantifying live microbial load in human saliva samples over time reveals stable composition and dynamic load*. I am the primary investigator and author of this paper. The co-authors listed above supervised or provided support for the research and have given permission for the inclusion of the work in this dissertation.

4.1.8 Supplemental Figures

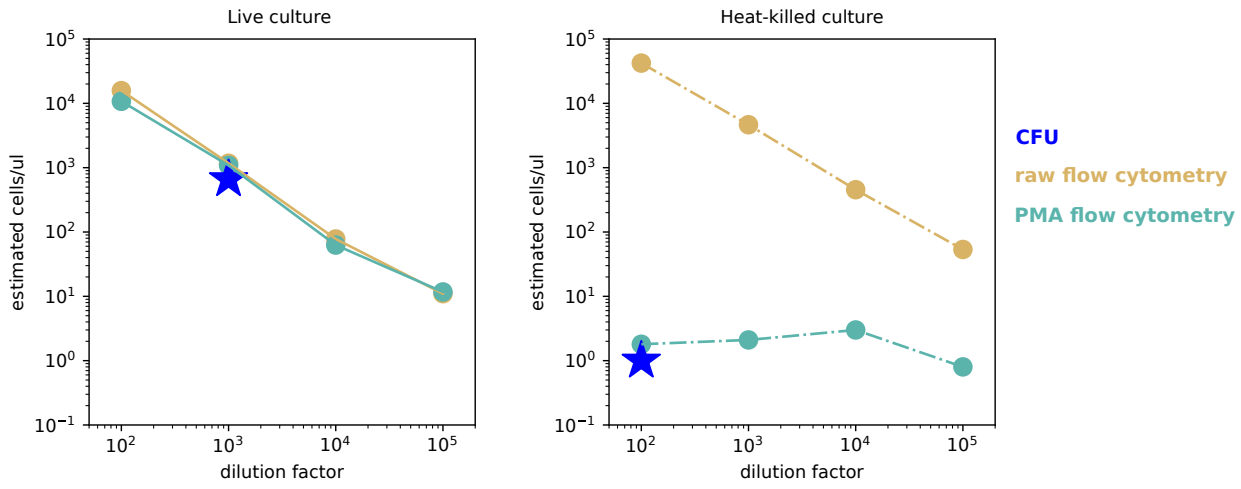

Figure 4.1.S1. *PMA removes dead cell signal from flow cytometry*. An overnight culture of E.coli was divided into 2 aliquots and either left untreated (A) or heated to 65°C for 10 minutes (B). The aliquots were serially diluted and quantified by CFU counts on an agar plate (blue star) and by flow cytometry both without (orange) and after PMA treatment (red).

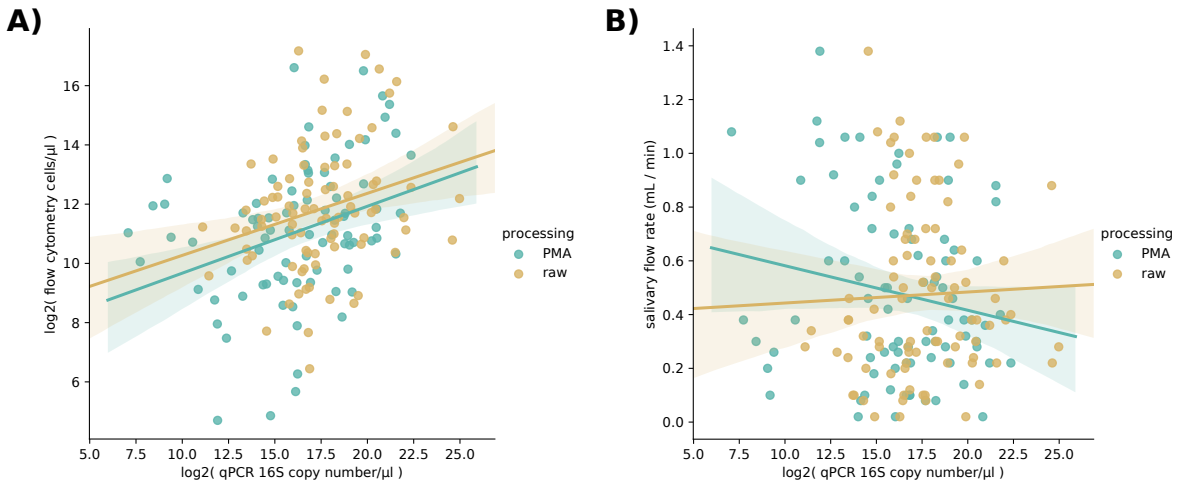

Figure 4.1.S2. *Correlation between microbial load quantification with qPCR and flow cytometry, and correlation with salivary flow rate*. (**A**) 16S rRNA gene copies per 2 ul sample determined by qPCR (a-axis) versus microbial cell count determined by flow cytometry (y-axis). Samples treated with PMA are in blue (Pearson's r = 0.314, p=0.003) and raw samples are in orange (Pearson's r = 0.287, p=0.007). (**B**) 16S rRNA gene copies per 2 ul sample determined by qPCR (a-axis) does not correlate with salivary flow rate (mL per minute). Samples treated with PMA are in blue (Pearson's r = -.176, p=0.107) and raw samples are in orange (Pearson's r = 0.036, p=0.742).

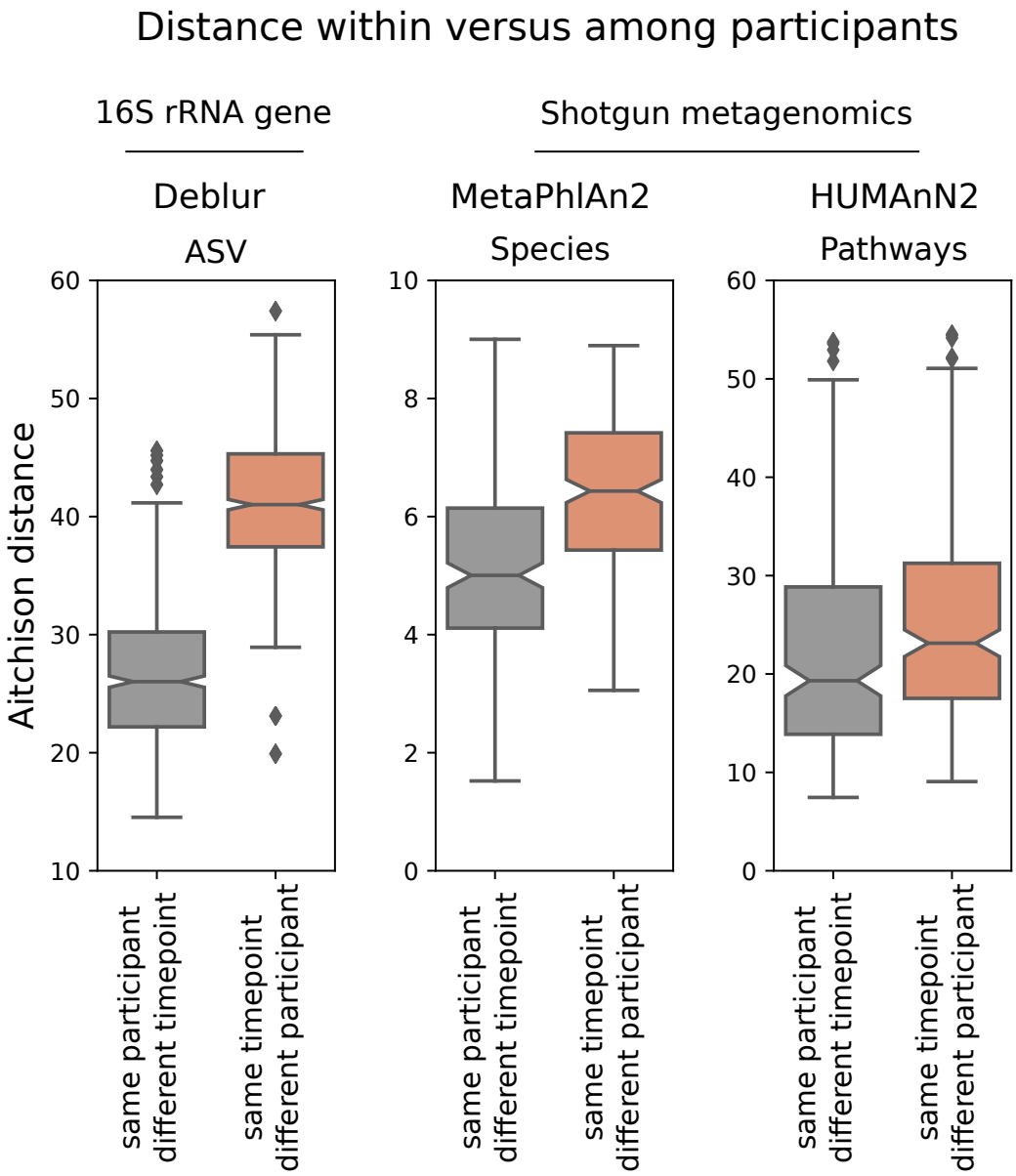

Figure 4.1.S3. *Aitchison distance within participants over time versus between participants across multiple sequencing outputs*. Saliva samples from the daily dynamics experiment were processed for both 16S rRNA gene amplicon sequencing and shotgun metagenomic sequencing (n=88). Aitchison distance was calculated between all samples in each dataset. The average distance among samples from the same individual over time (gray) were significantly smaller than the average distance among different participants collected at the same timepoint (pink) in Deblur output from 16S rRNA gene amplicon data at the amplicon sequence variant level (left), MetaPhlAn2 shotgun metagenomic data at the species level (middle), and HUMAnN2 shotgun metagenomic data at the pathway level (right). Statistical significance assessed with bootstrapped Kruskal-Wallis and *p-value* <0.001 for all datasets.

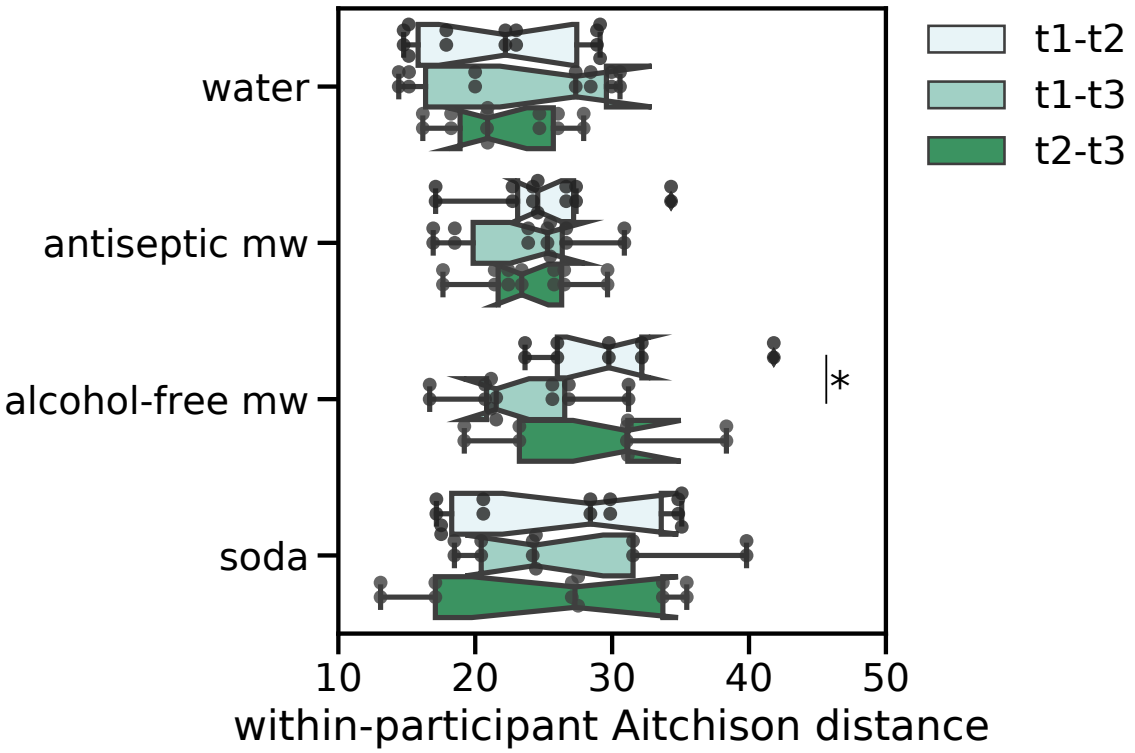

Figure 4.1.S4. *Aitchison distance within participants in the acute perturbation study*. Only Aitchison distances among samples collected from the same participant were used for this analysis. The average distance between samples collected at timepoints 1 versus 2, timepoints 1 versus 3, and timepoints 2 versus 3 were calculated for each treatment group and statistical significance between timepoint comparisons was assessed with bootstrapped Kruskal-Wallis on each treatment group. The only statistically significant difference was that samples collected directly following alcohol-free mouthwash treatment were more dissimilar than samples collected 2 hours later compared to the baseline timepoint (p=0.009).

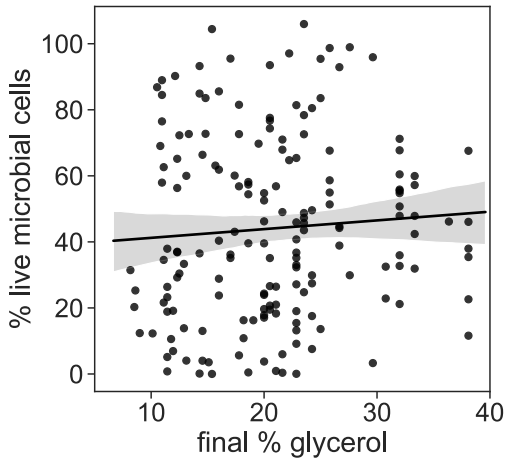

Figure 4.1.S5. *Effect of sample preservation on viability*. Participants were asked to passively drool on top of 2 mL 40% glycerol; because the salivary flow rate differed among individuals and across time points the final percentage of glycerol before freezing varied. However, the final percentage of glycerol had no correlation with the percentage of live cells in the sample (Pearson correlation coefficient R = 0.071, p=0.359) suggesting that final glycerol percentage did not influence viability

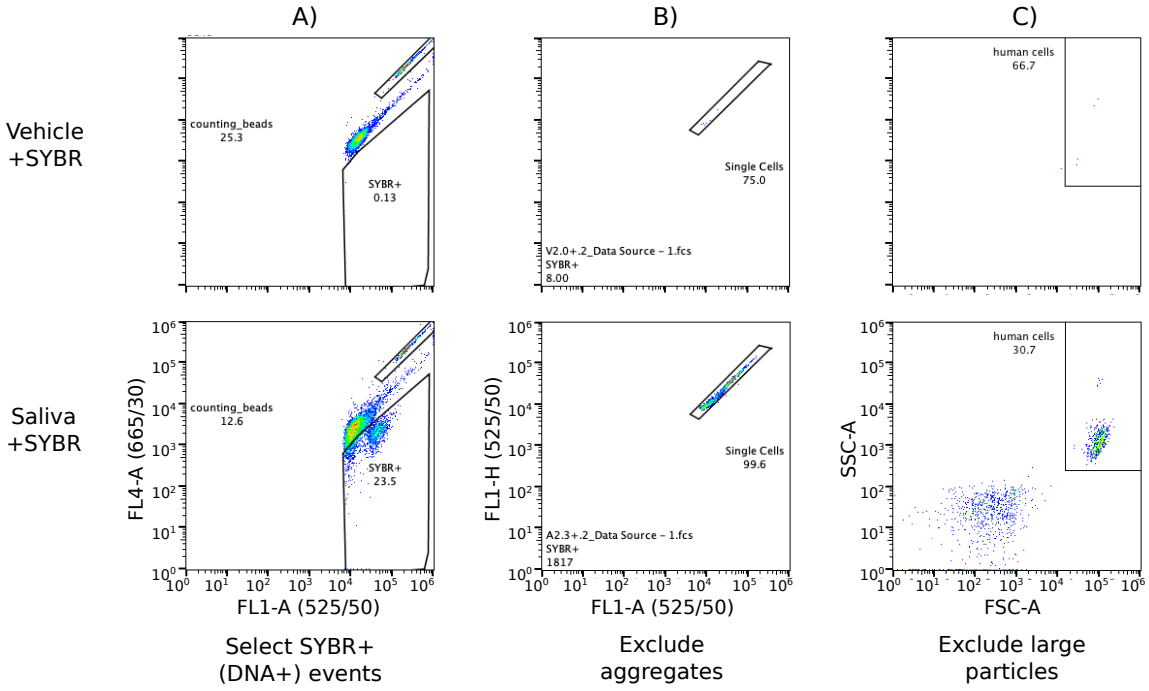

Figure 4.1.S6. *Flow cytometry gating strategy adapted from Props et al [Props 2016]*. Unstimulated saliva samples were diluted ten-fold in sterile PBS, filtered across a 5μm filter to remove human cells, stained with SYBR green, and processed on a Sony SH800 with Spherotech counting beads. The threshold was set on the FL1 detector. (**A**) The first gate selects for events with enhanced 525 nm specific emission to select DNA positive events. (**B**) Doublets were excluded by selecting only events following a linear trend between FL1 height and area. (**C**) Human cells are excluded by their large size on forward (FSC-A) and side (SSC-A) scatter area.

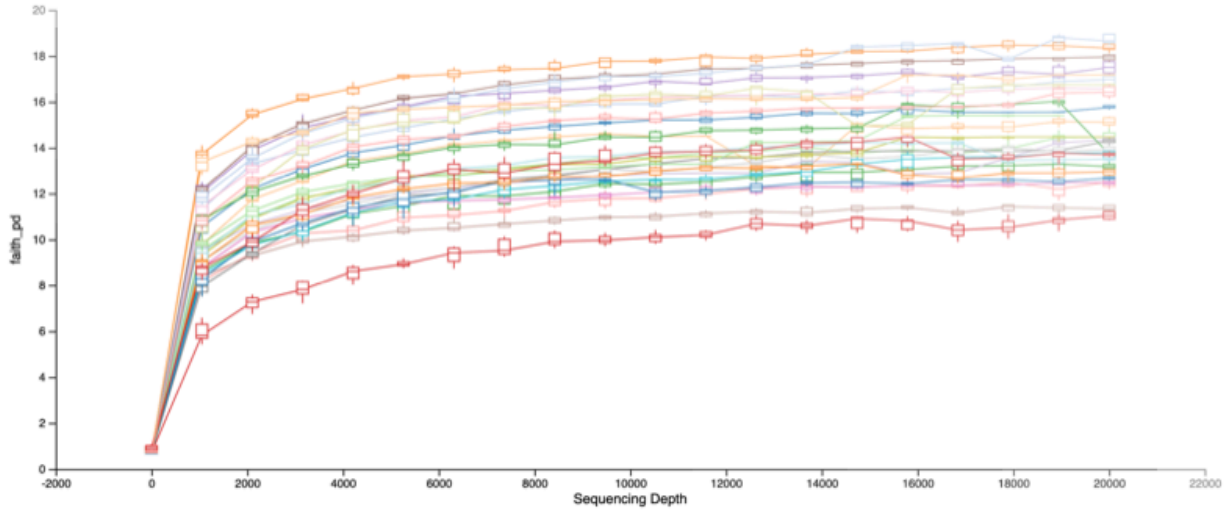

Figure 4.1.S7. *Faith's phylogenetic diversity alpha rarefaction curve by participant*. Qiime2 was used to generate and visualize an alpha-diversity rarefaction plot; 1 to 20,000 reads from each sample were randomly selected and Faith's PD was calculated on ten iterations of each sequencing depth (20 steps total) and plotted with the standard error of the mean.

4.1.9 References

1. Cabral DJ, Wurster JI, Flokas ME, Alevizakos M, Zabat M, Korry BJ, Rowan AD, Sano WH, Andreatos N, Ducharme RB, Chan PA, Mylonakis E, Fuchs BB, Belenky P. The salivary microbiome is consistent between subjects and resistant to impacts of short-Term hospitalization. Sci Rep. 2017;7(1):1–13.

2. Cameron SJS, Huws SA, Hegarty MJ, Smith DPM, Mur LAJ. The human salivary microbiome exhibits temporal stability in bacterial diversity. FEMS Microbiol Ecol. 2015;91(9):1–9.

3. Vandeputte D, Kathagen G, D'Hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, Wang J, Tito RY, De Commer L, Darzi Y, Vermeire S, Falony G, Raes J. Quantitative microbiome profiling links gut community variation to microbial load. Nature. 2017;551(7681):507–11.

4. Morton JT, Marotz C, Washburne A, Silverman J, Zaramela LS, Edlund A, Zengler K, Knight R. Establishing microbial composition measurement standards with reference frames. Nat Commun. 2019;10(1).

5. Props R, Kerckhof FM, Rubbens P, Vrieze J De, Sanabria EH, Waegeman W, Monsieurs P, Hammes F, Boon N. Absolute quantification of microbial taxon abundances. ISME J. 2017;11(2):584–7.

6. Props R, Monsieurs P, Mysara M, Clement L, Boon N. Measuring the biodiversity of microbial communities by flow cytometry. Methods Ecol Evol. 2016;7(11):1376–85.

7. Li C, Chng KR, Av-Shalom T V., Tucker-Kellogg L, Nagarajan N. An expectation-maximization-like algorithm enables accurate ecological modeling using longitudinal metagenome sequencing data. Microbiome. 2019;(7):118.

8. Carini P, Marsden PJ, Leff JW, Morgan EE, Strickland MS, Fierer N. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. Nat Microbiol. 2016;2(0):16242.

9. Rogers GB, Marsh P, Stressmann AF, Allen CE, Daniels TVW, Carroll MP, Bruce KD. The exclusion of dead bacterial cells is essential for accurate molecular analysis of clinical samples. Clin Microbiol Infect. 2010;16(11):1656–8.

10. Nelson MT, Pope CE, Marsh RL, Wolter DJ, Weiss EJ, Hager KR, Vo AT, Brittnacher MJ, Radey MC, Hayden HS, Eng A, Miller SI, Borenstein E, Hoffman LR. Human and Extracellular DNA Depletion for Metagenomic Analysis of Complex Clinical Infection Samples Yields Optimized Viable Microbiome Profiles. Cell Rep. 2019;26(8):2227-2240.e5.

11. Nocker A, Sossa-Fernandez P, Burr MD, Camper AK. Use of propidium monoazide for live/dead distinction in microbial ecology. Appl Environ Microbiol. 2007;73(16):5111–7.

12. Soejima T, Iida K, Qin T, Taniai H, Seki M, Takade A, Yoshida S. Photoactivated ethidium monoazide directly cleaves bacterial DNA and is applied to PCR for discrimination of live and dead bacteria. Microbiol Immunol. 2007;51(8):763–75.

13. Chu ND, Smith MB, Perrotta AR, Kassam Z, Alm EJ. Profiling living bacteria informs preparation of fecal microbiota transplantations. PLoS One. 2017;12(1):1–16.

14. Banihashemi A, Van Dyke MI, Huck PM. Application of long amplicon propidium monoazide-PCR to assess the effects of temperature and background microbiota on pathogens in river water. J Water Health. 2017;15(3):418–28.

15. Extercate RAM, Zaura E, Brandt BW, Buijs MJ, Koopman JE, Crielaard W, ten Cate JM. The effect of propidium monoazide treatment on the measured bacterial composition of clinical samples after the use of a mouthwash. Clin Oral Investig. 2015;19(4):813–22.

16. Cuthbertson L, Rogers GB, Walker AW, Oliver A, Green LE, Daniels TWV, Carroll MP, Parkhill J, Bruce KD, Van Der Gast CJ. Respiratory microbiota resistance and resilience to pulmonary exacerbation and subsequent antimicrobial intervention. ISME J. 2016;10(5):1081–91.

17. Klebanoff MA, Zhang J, Nansel TR, Yu K-F, Andrews WW, Brotman RM, Schwebke JR. Race of male sex partners and occurrence of bacterial vaginosis. Sex Transm Dis. 2010;37(3):184–90.

18. Zhao F, Liu H, Zhang Z, Xiao L, Sun X, Xie J, Pan Y, Zhao Y. Reducing bias in complex microbial community analysis in shrimp based on propidium monoazide combined with PCR-DGGE. Food Control. 2016;68:139–44.

19. Stinson LF, Keelan JA, Payne MS. Characterization of the bacterial microbiome in first-pass meconium using propidium monoazide (PMA) to exclude nonviable bacterial DNA. Lett Appl Microbiol. 2019;68(5):378–85.

20. Sawair FA, Ryalat S, Shayyab M, Saku T. The Unstimulated Salivary Flow Rate in a Jordanian Healthy Adult Population. J Clin Med Res. 2009;1(4):219–25.

21. Idrees M, Nassani M, Kujan O. Assessing the association between unstimulated whole salivary flow rate (UWSFR) and oral health status among healthy adult subjects: A cross-sectional study. Med Oral Patol Oral Cir Bucal. 2018;23(4):e384–90.

22. Gao Z, Perez-Perez GI, Chen Y, Blaser MJ. Quantitation of major human cutaneous bacterial and fungal populations. J Clin Microbiol. 2010;48(10):3575–81.

23. Keijser BJF, Zaura E, Huse SM, Van Der Vossen JMBM, Schuren FHJ, Montijn RC, Ten Gate JM, Crielaard W. Pyrosequencinq analysis of the oral microflora of healthy adults. J Dent Res. 2008;87(11):1016–20.

24. Zaura E, Keijser BJ, Huse SM, Crielaard W. Defining the healthy "core microbiome" of oral microbial communities. BMC Microbiol. 2009;9:1–12.

25. Welch JLM, Rossetti BJ, Rieken CW, Dewhirst FE, Borisy GG. Biogeography of a human oral microbiome at the micron scale. Proc Natl Acad Sci U S A. 2016;113(6):E791–800.

26. Proctor DiM, Fukuyama JA, Loomer PM, Armitage GC, Lee SA, Davis NM, Ryder MI, Holmes SP, Relman DA. A spatial gradient of bacterial diversity in the human oral cavity shaped by salivary flow. Nat Commun. 2018;9(1).

27. Dawes C. Circadian rhythms in human salivary flow rate and composition. J Physiol. 1972;220(3):529–45.

28. Aitchison J, Barceló-Vidal C, Martín-Fernández JA, Pawlowsky-Glahn V. Logratio Analysis and Compositional Distance. Math Geol. 2000;32:271–5.

29. Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K. Improving saliva shotgun metagenomics by chemical host DNA depletion. Microbiome. 2018;6(1):1–9.

30. 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.

31. 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 [Internet]. 2018;15(11):962–8. Available from: http://dx.doi.org/10.1038/s41592-018-0176-y

32. Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Midford PE, Ong Q, Ong WK, Paley S, Subhraveti P, Karp PD. The MetaCyc database of metabolic pathways and enzymes. Nucleic Acids Res. 2018;46(D1):D633–9.

33. Carini P, Delgado-Baquerizo M, Hinckley E-L, Brewer TE, Rue G, Vanderburgh C, McKnight D, Fierer N. Effects of spatial variability and relic DNA removal on the detection of temporal dynamics in soil microbial communities. bioRxiv [Internet]. 2018;11(1):402438. Available from: https://www.biorxiv.org/content/10.1101/402438v1.abstract

34. Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R, Peddada SD. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb Ecol Heal Dis. 2015;26(0):1–7.

35. Gloor GB, Macklaim JM, Fernandes AD. Displaying Variation in Large Datasets: Plotting a Visual Summary of Effect Sizes. J Comput Graph Stat. 2016;25(3):971–9.

36. Márton K, Madléna M, Bánóczy J, Varga G, Fejérdy P, Sreebny LM, Nagy G. Unstimulated whole saliva flow rate in relation to sicca symptoms in Hungary. Vol. 14, Oral Diseases. 2008. p. 472–7.

37. Samnieng P, Ueno M, Shinada K, Zaitsu T, Wright FAC, Kawaguchi Y. Association of hyposalivation with oral function, nutrition and oral health in community-dwelling elderly Thai. Vol. 29, Community Dental Health. 2012. p. 117–23.

38. Cunha-Cruz J, Scott JA, Rothen M, Mancl L, Lawhorn T, Brossel K, Berg J, Northwest Practice-based REsearch Collaborative in Evidence-based DENTistry. Salivary characteristics and dental caries: evidence from general dental practices. Vol. 144, Journal of the American Dental Association (1939). 2013.

39. Shimazaki Y, Fu B, Yonemoto K, Akifusa S, Shibata Y, Takeshita T, Ninomiya T, Kiyohara Y, Yamashita Y. Stimulated salivary flow rate and oral health status. 2017;59(1):55–62.

40. Aitchison J. Reducing the dimensionality of compositional data sets. J Int Assoc Math Geol. 1984;16(6):617–35.

41. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: And this is not optional. Front Microbiol. 2017;8(NOV):1–6.

42. Navazesh M, Kumar SKS. Measuring salivary flow. J Am Dent Assoc. 2008;139:35S-40S.

43. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 2013;41(1).

44. Marotz C, Amir A, Humphrey G, Gaffney J, Gogul G, Knight R. DNA extraction for streamlined metagenomics of diverse environmental samples. Biotechniques. 2017 Jun 1;62(6):290–3.

45. Marotz C, Amir A, Humphrey G, Gaffney J, Gogul G, Knight R. DNA extraction for streamlined metagenomics of diverse environmental samples. Biotechniques. 2017;62(6):290–3.

46. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vázquez-Baeza Y, González A, Morton JT, Mirarab S, Zech Xu Z, Jiang L, Haroon MF, Zhao H, et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017 Nov 1;551(7681):457.

47. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Caporaso JG, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37:848–57.

48. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Xu ZZ, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems. 2017 Apr 21;2(2):e00191-16.

49. McDonald D, Price MN, Goodrich J, Nawrocki EP, Desantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J [Internet]. 2012;6(3):610–8. Available from: http://dx.doi.org/10.1038/ismej.2011.139

50. Escapa IF, Chen T, Huang Y, Gajare P, Dewhirst FE, Lemon KP. New Insights into Human Nostril Microbiome from the Expanded Human Oral Microbiome Database (eHOMD): a Resource for the Microbiome of the Human Aerodigestive Tract. mSystems. 2018;3(6):1–20.

51. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Gregory Caporaso J. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome. 2018;6(1):1–17.

52. Krueger F. "Trim galore." A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. 2015.

53. Langmead B, Trapnell C, Pop M, Salzberg S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.

54. Gonzalez A, Navas-Molina JA, Kosciolek T, McDonald D, Vázquez-Baeza Y, Ackermann G, DeReus J, Janssen S, Swafford AD, Orchanian SB, Sanders JG, Shorenstein J, Holste H, Petrus S, Robbins-Pianka A, Brislawn CJ, Wang M, Rideout JR, Bolyen E, Dillon M, Caporaso JG, Dorrestein PC, Knight R. Qiita: rapid, web-enabled microbiome meta-analysis. Nat Methods [Internet]. 2018;15(10):796–8. Available from: http://dx.doi.org/10.1038/s41592-018-0141-9

4.2

Early microbial markers of periodontitis in the Oral Infections Glucose Intolerance and Insulin Resistance Study (ORIGINS)

Nearly 50% of Americans suffer from periodontitis, and 10% are diagnosed with diabetes. The high-comorbidity rate of these diseases suggests a shared etiology, at least in part. Changes in oral microbial communities have been documented in the context of severe periodontitis and diabetes, both independently and together. However, much less is known about the early microbial markers of these diseases. We used a subset of the ORIGINS project dataset, which collected detailed periodontal and cardiometabolic information from 787 healthy individuals, to identify early microbial markers of periodontitis and its association with markers of cardiometabolic health. Using state-of-the-art compositional data analysis tools, we identified the log ratio of *Treponema* to *Corynebacterium* bacteria to be a novel Microbial Indicator of Periodontitis (MIP), and found that this MIP correlates with poor periodontal health and cardiometabolic markers in both subgingival plaque and saliva.

4.2.1 Introduction

The human oral cavity hosts hundreds of unique microbial taxa. Within the oral cavity, there are multiple distinct niches that contain different compositions of microbial taxa. For example the supra- and subgingival tooth surface, tongue, cheek, and roof of mouth each have reproducibly distinct microbial communities [1]. In the context of severe periodontal disease, the composition of microbial taxa in the supra and subgingival plaque (SubP) undergo dramatic changes [2–4]. Periodontitis-associated subgingival biofilms often contain a climax community

dominated by *Porphyromonas gingivalis*,*Treponema denticola*, and *Tannerella forsythia,* referred to as the 'red complex' [5]. The red complex has been studied in depth for its ability to negatively affect host physiology through virulence factors and expedite gingival deterioration in severe disease. However less is known about the early microbial markers of periodontitis, or when compositional changes in plaque biofilms occur relative to disease onset.

There is strong evidence for a link between oral health and cardiometabolic health [6]. While it has long been recognized that individuals with diabetes are at higher risk for periodontitis, less is known about the etiology of these diseases and their interplay during the early stages of disease onset.

Next-generation sequencing has enabled the high-throughput collection of detailed microbial information from thousands of samples. However, this data is inherently compositional, and care must be taken to draw reproducible conclusions from these datasets [7]. In this analysis, we use a suite of recently developed compositional data analysis tools to identify differentially abundant bacteria in the early stages of periodontitis and cardiometabolic disease.

We applied these tools to a subset of data collected through the ORIGINS project (Oral Infections, Glucose Intolerance and Insulin Resistance Study) [8]. This dataset contains information from a large group of healthy individuals including a comprehensive periodontal examination, quantitative cardiometabolic markers, as well as 16S rRNA gene amplicon sequencing from saliva and SubP from both healthy and diseased sites, processed separately, which allowed us to look for site specific microbial markers.

4.2.2 Results

Overview of the cohort

Wave2 of the ORIGINS project recruited 787 healthy volunteers. Each participant underwent extensive periodontal examination and metabolic measurements as previously described [8]. Subgingival plaque (SubP) was collected from healthy and diseased sites (as applicable) following a standardized protocol totaling 1,107 samples (Fig 1). Saliva samples were collected in parallel and processed for a subset of 282 participants. Both saliva and SubP samples were processed for 16S rRNA gene amplicon sequencing.

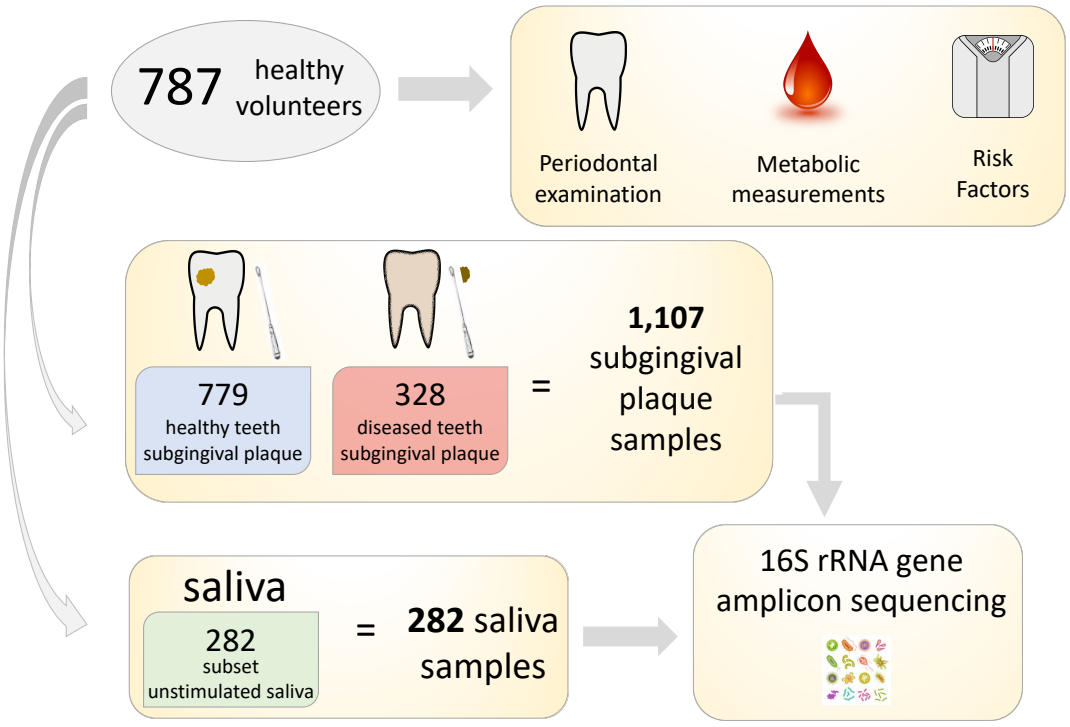

Figure 4.2.1. *Experimental design*. 787 healthy volunteers were recruited to participate in the ORIGINS project. Each participant underwent an extensive periodontal examination, metabolic assessment, and completed a detailed questionnaire including demographic and risk factor information. SubP samples were collected from teeth with periodontal pockets <4mm depth (healthy) and teeth with periodontal pockets >4mm depth (diseased) where applicable. In parallel, unstimulated saliva was collected and processed for a subset of individuals. In total, 16S rRNA gene amplicon sequencing data from 1,107 SubP samples and 282 saliva samples was generated for analysis.

Periodontal pocket depth drives microbial diversity

We used the robust Aitchison Principal Components Analysis (RPCA) method [9] to assess beta-diversity across SubP samples. The first axis of separation showed distinct clusters of SubP collected from shallow versus deep periodontal pockets (Fig 2A). Surprisingly, the distance between healthy and diseased samples from the same individual were larger than the distance between healthy samples from different people or diseased samples from different people (Fig 2B). This finding was only identified using RPCA, which accounts for the inherent sparsity and compositionality of next-generation sequencing experiments [7,9].

We performed an effect size redundancy analysis (RDA) to determine which factors explained the variation in microbial composition across samples [10]. Thirteen factors were included in the RDA including six periodontal metrics (shallow or deep periodontal pocket depth, percent of sites bleeding on probing, whole mouth periodontal score, average whole mouth pocket depth, average whole mouth attachment loss, and percent of sites with attachment loss greater than 3), three demographic factors (participant, sex, and age), three metabolic factors (fasting insulin, prediabetes status, and average systolic blood pressure) and one lifestyle factor (tobacco smoking status). Of these thirteen factors, eight were found to have a significant effect size (p-value <0.05).

By far, the factor which explained the most variation was whether the plaque sample came from a deep or shallow periodontal pocket (Fig 2C). Overall, periodontal metrics accounted for most of the explained variance (20%) followed by demographic factors (1.7%) and lastly metabolic factors (0.6%). Tobacco smoking did not have a significant effect on SubP microbial composition in this analysis.

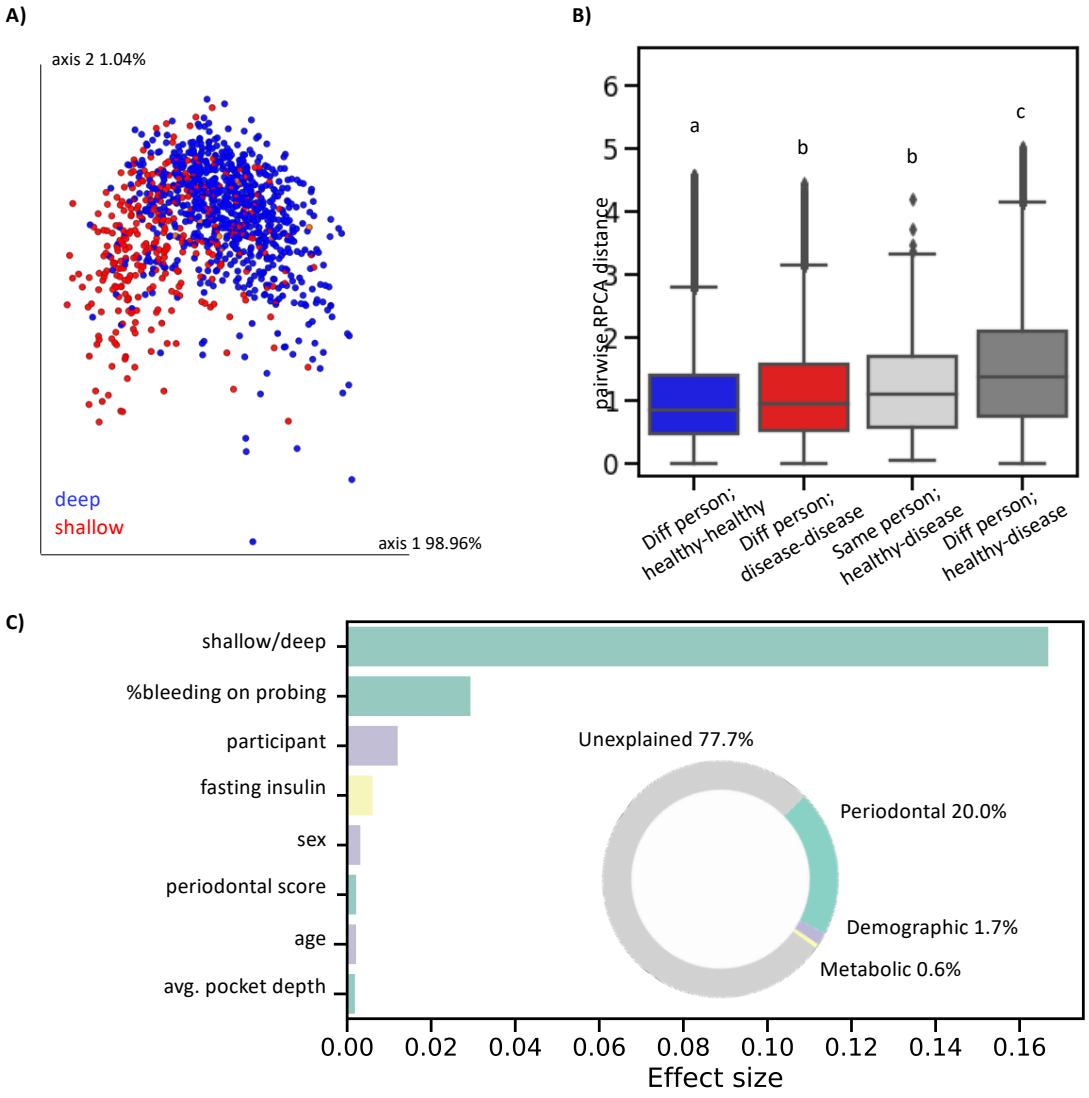

Figure 4.2.2. *Beta-diversity and redundancy analysis in subgingival plaque*. **A**) RPCA colored by periodontal pocket depth. Permanova pseudo-F statistic = 397.062, p-value<0.001. **B**) RPCA distance among pairwise samples; healthy samples from different people (n=308,505), diseased samples from different people (n=54,285), healthy versus disease samples from the same person (n=322), healthy versus disease samples from different people (n=259,058). Each group is significantly different from all other groups (one-way ANOVA with Tukey's multiple corrections, p<0.05). **C**) Redundancy analysis (RDA) estimates the percent microbial diversity explained by each variable. Inset donut chart sums effect sizes by category; periodontal variables explained the majority of microbial variation (20.0%), followed by demographic variables (1.7%) and metabolic variables (0.6%).

Identifying an early microbial indicator of periodontitis in subgingival plaque

Differential ranking [11] was used to identify differentially abundant microbes in shallow versus deep periodontal pockets. Amplicon sequence variants (ASVs) aligned to the genus *Treponema* were more associated with deep periodontal pockets, whereas ASVs aligned to the genus *Corynebacterium* were more associated with shallow periodontal pockets (Fig 3A). To generate a microbial indicator of periodontitis (MIP), we used *Corynebacterium* as a 'reference frame' and calculated the log-ratio of all *Treponema* counts to all *Corynebacterium* counts. This MIP was significantly higher in deep compared to shallow periodontal pockets (paired T-test <0.0001) and revealed that the ratio of *Treponema* to *Corynebacterium* is roughly even in deep periodontal pockets, whereas the ratio in shallow periodontal pockets is heavily skewed towards *Corynebacterium* (Fig 3B). Importantly, because this is relative abundance data we cannot conclude whether this finding is due to an increase in *Treponema* or a decrease in *Corynebacterium*, but the ratio of these two organisms is a consistent biomarker of periodontal health.

To validate the robustness of the MIP, we tested the ability of the MIP to classify samples from shallow versus deep periodontal pockets. When using the entire dataset (1,832 ASVs) samples were classified with an accuracy of 0.89 +/- 0.04. When using just the subset of data used to generate the MIP (164 ASVs, or less than 10% of the total data) samples were classified with an accuracy of 0.84 +/- 0.05% (Fig 3C), supporting the ability of the MIP to predict which samples will develop periodontitis.

Interestingly, even in SubP samples from shallow sites (n=779), the MIP was significantly associated with the percent of sites bleeding on probing across the whole mouth (Pearson

correlation = 0.243, p-value = $8.06e^{-12}$) (Fig 3D). This indicates that even before the detection of disease, there are microbial changes in the SubP.

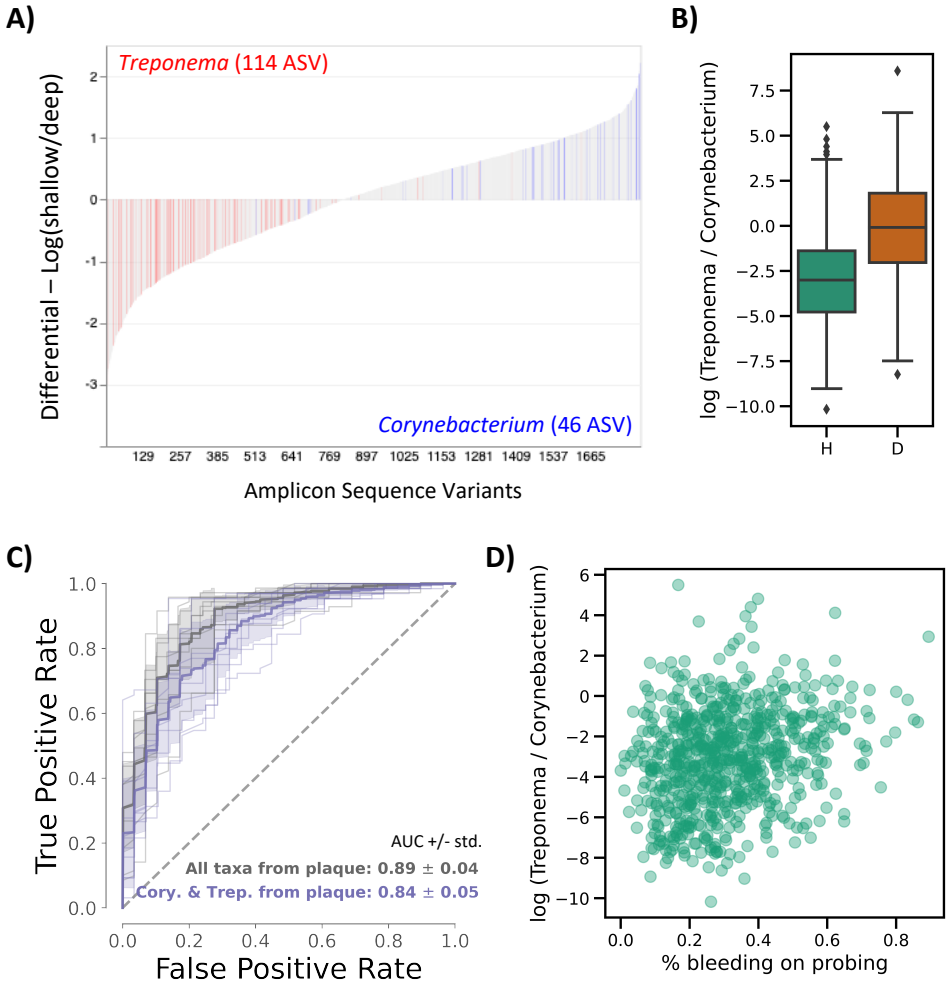

Figure 4.2.3. *The ratio of Treponema:Corynebacteria is an early Microbial Indicator of Periodontitis (MIP) in subgingival plaque*. **A**) Differential ranking with Songbird revealed that *Treponema* sequences were associated with deep SubP, whereas *Corynebacterium* sequences were associated with shallow SubP. **B**) The log ratio of *Treponema*:*Corynebacterium* significantly distinguishes shallow from deep periodontal pockets and is used as a Microbial Indicator of Periodontitis (MIP). **C**) ROC curve displaying the accuracy of a Random Forest classifier trained on the full dataset (blue) versus trained only on *Treponema* and *Corynebacterium* sequences and log-ratio (green) shows similar accuracy at predicting shallow versus deep periodontal pocket depth. **D**) In plaque collected from shallow (healthy) subgingival pockets (n=779), MIP was positively correlated with the percent of sites bleeding on probing (Pearson correlation = 0.243, *p-value* = $8.06e^{-12}$), indicating that microbial changes occur in plaque before clinically detectable disease.

Evaluating the microbial indicator of periodontitis in saliva

Collection of SubP is not trivial and requires clinically trained professionals. Saliva is much easier and economical to collect. We performed 16S rRNA gene amplicon sequencing on a subset of saliva samples (n=282) from the same cohort. Because the saliva and SubP samples were processed with different sequencing strategies at different institutions, we first assigned taxonomy to the Human Oral Microbiome Database (HOMD), a well-curated database of full-length 16S rRNA gene amplicon sequences found in the human oral cavity [12]. We collapsed the SubP and saliva datasets to the species level and merged together. Beta-diversity analysis of the merged table revealed that saliva and SubP had compositionally distinct microbial communities (Fig 4A). The majority of microbial taxa in the merged table were found in both SubP and saliva, although each niche also contained distinct microbiota, with SubP being more diverse than saliva (Fig 4B).

Effect size analysis using RDA of just the saliva table revealed different factors drive microbial diversity in saliva compared to SubP (Fig 4C). Eight factors were included in the RDA including two demographic factors (participant and age), three metabolic factors (mean systolic blood pressure, BMI, and prediabetes status), two periodontal factors (average whole-mouth attachment loss and average whole-mouth periodontal pocket depth) and one lifestyle factor (tobacco-smoking status). Overall, the percent explained was much lower in saliva compared to SubP (5.8% versus 22.3%, respectively). The only significant factors in the saliva RDA were smoking-status and participant (Fig 4C). However, despite clear differences in microbial community composition between saliva and SubP, the MIP was significantly correlated (Pearson $R = 0.387$, p-value = $3.97E^{-11}$) (Fig 4D).

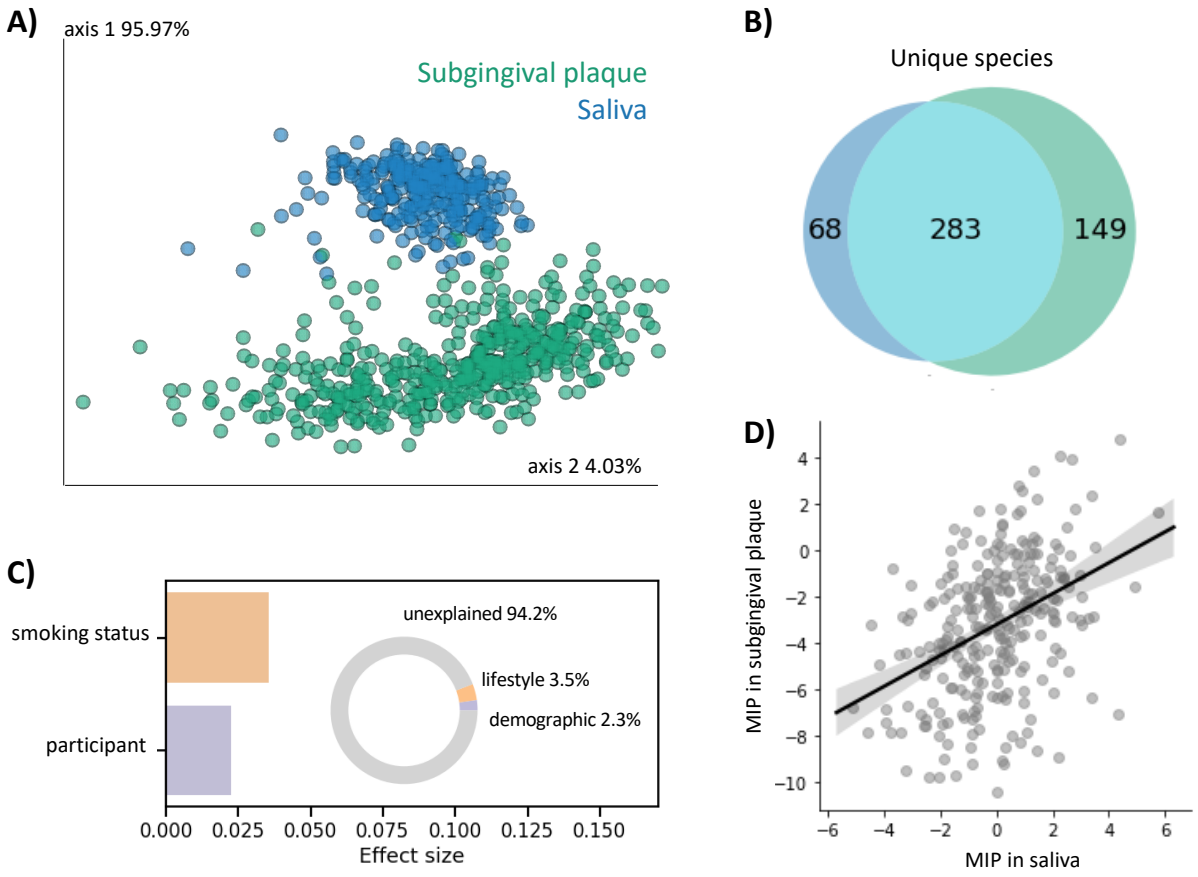

Figure 4.2.4. *Plaque and saliva are compositionally distinct but have correlated MIP*. **A**) Beta-diversity analysis with RPCA show distinct clustering of saliva vs subgingival plaque samples (PERMANOVA<0.001). **B**) Venn diagram of 16S rRNA gene amplicon sequencing data collapsed to the species level shows a majority of microbial species were identified in both saliva and SubP, and that SubP was more diverse. **C**) Redundancy analysis (RDA) estimates the percent microbial diversity explained by each variable. Inset donut chart sums effect sizes by category; unlike SubP, saliva microbial diversity is driven by lifestyle or demographic variables and is not significantly explained by metabolic or periodontal metrics. **D**) Microbial indicator of periodontitis (MIP) was significantly correlated between SubP and saliva samples, despite having been processed at different institutes with different sequencing parameters and being driven by different variables (Pearson R = 0.387, p-value = 3.97E-11).

Microbial Indicator of Periodontitis Correlates with Periodontal Metrics

We assessed the correlation of the MIP with various whole-mouth periodontal metrics (Table 1). We found that SubP MIP was positively, significantly correlated with the percent of sites bleeding on probing, average pocket depth, and average attachment loss. SubP MIP was also very highly correlated with faith's phylogenetic diversity, a measurement of alpha diversity that

accounts for phylogenetic relatedness [13]. All of these correlations held true when looking at only healthy participant SubP, again suggesting that microbial changes precede detectable disease.

Saliva MIP was significantly correlated with the percent of sites bleeding on probing and average pocket depth, but not attachment loss. This held true when looking only at participants with moderate or severe periodontitis, but not healthy participants, suggesting that microbial changes in the subgingival pocket precede microbial changes in saliva. Saliva MIP was also strongly correlated with Faith's phylogenetic diversity in all patients regardless of periodontal status.

Table 4.2.1. *Microbial Indicator of Periodontitis (MIP) is correlated with multiple metrics of periodontal disease in both saliva and subgingival plaque*. Perbop = percent of sites bleeding on probing; meanpd = average periodontal pocket depth; meanaloss = average attachment loss; faith_pd = Faith's phylogenetic diversity. Bolded values represent statistically significant Pearson correlations ($p<0.05$).

subgingival plaque									covariate	saliva								
all samples			no periodontitis			moderate to severe perio				all samples			no periodontitis			moderate to severe perio		
Pearson r	p-value	n	Pearson r	p-value	n	Pearson r	p-value	n		Pearson r	p-value	n	Pearson r	p-value	n	Pearson r	p-value	n
0.193	**1.21E-10**	**1096**	**0.140**	**1.52E-04**	**728**	**0.235**	**5.01E-06**	**368**	perbop	**0.150**	**1.26E-02**	**276**	0.078	2.96E-01	183	**0.229**	**2.76E-02**	**93**
0.251	**3.19E-17**	**1097**	**0.151**	**4.32E-05**	**729**	**0.295**	**8.31E-09**	**368**	meanpd	**0.250**	**2.66E-05**	**276**	0.085	2.55E-01	183	**0.405**	**5.56E-05**	**93**
0.175	**5.32E-09**	**1097**	**0.088**	**1.74E-02**	**729**	**0.179**	**5.48E-04**	**368**	meanaloss	0.049	4.17E-01	276	-0.095	1.99E-01	183	0.059	5.77E-01	93
0.703	**1.25E-165**	**1107**	**0.671**	**1.25E-96**	**729**	**0.744**	**5.75E-66**	**368**	faith_pd	**0.462**	**5.22E-16**	**276**	**0.374**	**2.74E-07**	**178**	**0.650**	**2.37E-12**	**92**

Microbial Indicator of Periodontitis Correlates with Metabolic Metrics

The overarching goal of the ORIGINS project is to identify associations between oral microbes, oral health and cardiometabolic health. To this end we evaluated the correlation of the MIP with various cardiometabolic health metrics (Table 2). SubP MIP was positively, significantly correlated with body mass index (BMI), average systolic and diastolic blood pressure, fasting glucose levels and fasting insulin levels. When taking into account only healthy participants, SubP was also correlated with hemoglobin A1c, but not fasting glucose. When taking into account participants with moderate or severe periodontitis, only BMI and systolic/diastolic blood pressure

were significantly correlated with SubP MIP. Together these results suggest that the microbial changes underlying periodontal health are also influencing cardiometabolic health.

Saliva MIP was positively, significantly correlated with BMI and systolic blood pressure, but this correlation did not hold true when subsetting out healthy participants or participants with moderate to severe periodontitis. This lends further support to the hypothesis that microbial changes originating in the SubP are only later detectable in saliva.

Table 4.2.2. *Microbial Indicator of Periodontitis (MIP) is correlated with multiple markers of cardiometabolic health in both saliva and subgingival plaque*. HbA1C = hemoglobin A1c; bmi = body mass index; meansbp = mean systolic blood pressure; meandbp = mean diastolic blood pressure; glucosecrc = fasting glucose; hsinsulin = fasting insulin. Bolded values represent statistically significant Pearson correlations ($p<0.05$).

subgingival plaque									covariate	saliva								
all samples			no periodontitis			moderate to severe perio				all samples			no periodontitis			moderate to severe perio		
Pearson r	p-value	n	Pearson r	p-value	n	Pearson r	p-value	n		Pearson r	p-value	n	Pearson r	p-value	n	Pearson r	p-value	n
0.040	1.82E-01	1107	**0.134**	**2.82E-04**	**729**	-0.085	1.05E-01	368	HbA1c	0.023	7.04E-01	276	0.087	2.42E-01	183	-0.064	5.45E-01	93
0.163	**6.25E-08**	**1086**	**0.143**	**1.28E-04**	**713**	**0.129**	**1.42E-02**	**363**	bmi	**0.120**	**4.89E-02**	**271**	0.100	1.83E-01	180	0.081	4.45E-01	91
0.145	**1.28E-06**	**1101**	**0.125**	**7.83E-04**	**723**	**0.126**	**1.55E-02**	**368**	meansbp	**0.121**	**4.43E-02**	**276**	0.143	5.40E-02	183	0.056	5.92E-01	93
0.150	**5.70E-07**	**1101**	**0.106**	**4.39E-03**	**723**	**0.168**	**1.18E-03**	**368**	meandbp	0.109	7.00E-02	276	0.110	1.39E-01	183	0.072	4.94E-01	93
0.079	**8.97E-03**	**1104**	0.064	8.40E-02	726	0.031	5.48E-01	368	glucosecrc	0.081	1.84E-01	274	0.071	3.44E-01	181	-0.002	9.87E-01	93
0.133	**8.61E-06**	**1107**	**0.132**	**3.70E-04**	**729**	0.088	9.05E-02	368	hsinsulin	0.097	1.08E-01	276	0.101	1.72E-01	183	0.028	7.90E-01	93

4.2.3 Methods

Sample collection

ORIGINS is an occupation-based cohort study among members of the Service Employees International Union 1199 designed to investigate the relationship between oral microbial community composition and glucose metabolism. Periodontal examination, subgingival plaque and saliva collection were performed as previously described [8]. In summary, 1,188 subgingival plaque samples (4 samples from 297 participants) were collected from the most posterior tooth per

quadrant (excluding third molars) via sterile curettes after removal of the supragingival plaque [18]. Unstimulated saliva was collected from each participant in parallel.

DNA extraction and 16S rRNA gene sequencing

DNA was extracted from subgingival plaque and saliva samples by The Forsyth Institute. 16S rRNA gene amplicon sequencing was performed on subgingival plaque samples by The Forsyth Institute using primers targeting variable region 4; Forward- CCTACGGGAGGCAGCAG and Reverse- CAAGCAGAAGACGGCATACGAGAT. Sequencing was performed on a MiSeq using a Paired End 250 cycle kit.

16S rRNA gene amplicon sequencing libraries on DNA extracted from saliva was performed at UC San Diego using the Earth Microbiome Project protocol [19,20]. Sequencing was performed on a MiSeq using a Paired End 150 cycle kit.

Sequence analysis

Raw reads were analyzed with QIIME2 [21]. Demultiplexed sequences were quality filtered with default parameters in qiime quality-filter q-score, namely, reads were trimmed after the first appearance of 3 basecalls with a PHRED score of 4 or less, and the entire read was removed if the read was truncated to less than 75% of the input sequence. Quality filtered forward-read sequences were denoised using Deblur [22] with the default parameters. Samples with less than 1,000 quality filtered reads were removed from downstream analysis. In order to remove reads aligned to chloroplast or mitochondrial genes, sequences were aligned using a classifier pretrained

on the GreenGenes v13_8 database with 99% sequence homology using sklearn [23]. Sequences aligned to mitochondria or chloroplast were removed using filter-table --p-exclude (0.005% of the entire dataset). A phylogenetic tree was created using fragment insertion via SEPP [24]. Taxonomy was assigned using sklearn [23] against the HOMD database version 15.1 [12]. All features not present in at least 1% of samples were excluded from downstream analysis.

The final quality-filtered subgingival plaque table contained 43,709,128 reads across 1,107 samples with a total of 1,832 amplicon sequence variants (ASV). The final quality-filtered saliva table contained 4,892,251 reads across 282 samples with a total of 859 ASVs.

Differential Abundance Testing

To determine which taxa are associated with which phenotypes in our dataset, we used the concept of Reference Frames [11]. This tool accounts for the compositional nature of next-generation sequencing experiments [7]. In brief, comparing relative abundances among sample groups can be misleading when the total microbial load is unknown, as is the case in this dataset. To avoid these pitfalls, we used the machine learning tool Songbird (https://github.com/biocore/songbird) to perform multinomial regression and then ranked each ASV by it's coefficient in the regression model to determine each taxon's relative differential across a given phenotype. Periodontal pocket depth was used as the formula in the model. The number of random test samples held back for validation in the model was 111 (10%). We used a batch size of 10 with 500 epochs (number of passes through the entire dataset to train the model), a learning rate of 0.001 and a differential prior of 10. The resulting ranks (differentials.qza) were visualized with Qurro [25] and allowed us to prioritize which taxa were most associated with a given phenotype.

To identify taxa associated with shallow versus deep periodontal pockets, we browsed the highest and lowest ranked microbes in this category using Qurro [25]. ASVs assigned to the genus *Corynebacterium* were mostly associated with shallow pockets, whereas ASVs assigned to the genus *Treponema* were mostly associated with deep pockets. To generate a microbial indicator of periodontitis (MIP), we used *Corynebacterium* as a 'reference frame' and calculated the log-ratio of all *Treponema* counts to all *Corynebacterium* counts.

Classification

A Random Forests (RF) [26] model was trained to predict disease status based on shallow versus deep periodontal pockets. The RF model was trained using a Stratified K-Folds cross-validation (CV) with 10-Fold CV splits. On each CV split a RF model with 500 estimators was trained and RF probability-predictions were compared to the test set using the Receiver Operating Characteristic (ROC). The mean and standard deviation from the mean were calculated for the area under the Area Under the Curve (AUC) across the 10-fold CV. This classification was performed on the whole ASV level data table and compared to the table filtered for only members of *Treponema* and *Corynebacterium* concatenated with the log-ratio of *Treponema* to *Corynebacterium.* All classification was performed through Scikit-learn (v. 0.22.2) [23].

4.2.4 Discussion

In a cohort of 787 healthy individuals we were able to identify early microbial markers of periodontitis. Microbial diversity in SubP was most strongly explained by periodontal metrics such as subgingival pocket depth and percent of sites bleeding on probing. RPCA beta-diversity analysis revealed that periodontal niche (i.e. whether the sample was obtained from a shallow versus deep periodontal pocket) was a more important indicator of microbial composition than individual variation. This was not confirmed by metrics that do not account for compositionality (e.g. UniFrac, Bray-Curtis), which can be greatly affected by microbial load [Morton 2019]. Since previous studies have shown increased microbial burden in subgingival pockets with periodontitis, it is likely that microbial load varied greatly across the samples in this dataset and therefore it is crucial to use scale-invariant analyses.

Redundancy analysis (RDA) revealed that saliva microbial communities were influenced by different factors compared to SubP. For instance, while tobacco smoking did not have a significant effect size in SubP microbial composition it had the biggest effect size in saliva microbial composition. This is in line with previous reports showing that microbial composition in oral washes was affected by smoking status [14], while SubP is not greatly affected by smoking status [15]. This finding highlights that saliva and SubP microbial communities are driven by different environmental factors.

We used the factor with the highest effect size on microbial diversity in SubP, whether the sample came from a deep or shallow pocket, to identify a microbial indicator of early periodontal disease. Using reference frames, we calculated the log-ratio of *Treponema*:*Corynebacterium* and found that it significantly differentiated healthy from diseased periodontal pocket sites. This log-ratio was used as a Microbial Indicator of Periodontitis (MIP). SubP MIP was significantly

correlated with poor periodontal health across a wide range of metrics when only looking at healthy plaque samples, suggesting that microbial communities change before disease is clinically detectable. Red complex organisms canonically associated with periodontitis in the literature were also positively correlated with periodontal disease status, but they were not as widely prevalent across samples which complicates scale-invariant analyses.

Despite the fact that saliva has a compositionally distinct microbiome compared to SubP, is driven by different metadata variables, and was sequenced independently with different parameters, we found that the MIP was significantly correlated between plaque and saliva. Saliva MIP was also correlated with poor periodontal health, and a subset of cardiometabolic markers.

Remarkably, phylogenetically-informed alpha diversity was strongly correlated with MIP in both saliva and SubP across all periodontal status categories. Both *Treponema* and *Corynebacterium* species have been identified as microbial scaffolds in plaque biofilms. In the context of healthy periodontal plaque, reproducible biofilms with a specific taxonomic organization, referred to as 'hedgehog' biofilms, are widely prevalent [16]. In the context of severe periodontitis, *Treponema* taxa have been found in the deepest sections of the periodontal pocket, and form close associations with diverse rod-like bacteria [17]. In light of our finding that the ratio of *Treponema* to *Corynebacterium* increases in periodontal disease, this suggests that the biofilm structure shifts from being scaffolded primarily by *Corynebacterium* to *Treponema*, where *Treponema* biofilms are more phylogenetically diverse than *Corynebacterium* biofilms.

Importantly, these microbial community composition transitions appear to occur early in disease, before periodontitis can be diagnosed. The results from this analysis also suggest that these microbial changes occur first in plaque, and as disease progresses can be identified in saliva. Future longitudinal sampling will allow us to more definitively determine if the

Treponema:Corynebacterium ratio increase precedes periodontitis, and this MIP could be used as an early marker of periodontitis that could help guide therapy to prevent periodontal deterioration.

4.2.5 Acknowledgements

Chapter IV, part 2, is currently being prepared for submission: Marotz C, Martino C, Knight R, Demmer R. *Early microbial markers of periodontitis in the Oral Infections Glucose Intolerance and Insulin Resistance Study (ORIGINS).* I am the primary investigator and author of this paper. The co-authors listed above supervised or provided support for the research and have given permission for the inclusion of the work in this dissertation.

4.2.6 References

1. Xu X, He J, Xue J, Wang Y, Li K, Zhang K, Guo Q, Liu X, Zhou Y, Cheng L, Li M, Li Y, Li Y, Shi W, Zhou X. Oral cavity contains distinct niches with dynamic microbial communities. Environ Microbiol. 2015;17(3):699–710.
2. Abiko Y, Sato T, Mayanagi G, Takahashi N. Profiling of subgingival plaque biofilm microflora from periodontally healthy subjects and from subjects with periodontitis using quantitative real-time PCR. J Periodontal Res. 2010;45(3):389–95.
3. Heller D, Silva-Boghossian CMI, Do Souto RM, Colombo APV. Subgingival microbial profiles of generalized aggressive and chronic periodontal diseases. Arch Oral Biol. 2012;57(7):973–80.
4. Kageyama S, Takeshita T, Asakawa M, Shibata Y, Takeuchi K, Yamanaka W, Yamashita Y. Relative abundance of total subgingival plaque-specific bacteria in salivary microbiota reflects the overall periodontal condition in patients with periodontitis. PLoS One. 2017;12(4):1–12.
5. Holt SC, Ebersole JL. Porphyromonas gingivalis, Treponema denticola, and Tannerella

forsythia: The “red complex”, a prototype polybacterial pathogenic consortium in periodontitis. Periodontol 2000. 2005;38:72–122.

6. Preshaw PM, Alba AL, Herrera D, Jepsen S, Konstantinidis A, Makrilakis K, Taylor R. Periodontitis and diabetes: A two-way relationship. Diabetologia. 2012;55(1):21–31.

7. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: And this is not optional. Front Microbiol. 2017;8(NOV):1–6.

8. Demmer RT, Jacobs DR, Singh R, Zuk A, Rosenbaum M, Papapanou PN, Desvarieux M. Periodontal Bacteria and Prediabetes Prevalence in ORIGINS. J Dent Res. 2015;94(9):201S-211S.

9. Sparse AN, Technique C, Microbial R. crossm Perturbations. 2019;4(1):1–13.

10. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, Kurilshikov A, Bonder MJ, Valles-Colomer M, Vandeputte D, Tito RY, Chaffron S, Rymenans L, Verspecht C, Sutter L De, Lima-Mendez G, D’hoe K, Jonckheere K, Homola D, Garcia R, Tigchelaar EF, Eeckhaudt L, Fu J, Henckaerts L, Zhernakova A, Wijmenga C, Raes J. Population-level analysis of gut microbiome variation. Science (80-). 2016;352(6285):560–4.

11. Morton JT, Marotz C, Washburne A, Silverman J, Zaramela LS, Edlund A, Zengler K, Knight R. Establishing microbial composition measurement standards with reference frames. Nat Commun. 2019;10(1).

12. Escapa IF, Chen T, Huang Y, Gajare P, Dewhirst FE, Lemon KP. New Insights into Human Nostril Microbiome from the Expanded Human Oral Microbiome Database (eHOMD): a Resource for the Microbiome of the Human Aerodigestive Tract. mSystems. 2018;3(6):1–20.

13. Faith DP, Baker AM. Phylogenetic Diversity (PD) and Biodiversity Conservation: Some Bioinformatics Challenges. Evol Bioinforma. 2006;2:117693430600200.

14. Wu J, Peters BA, Dominianni C, Zhang Y, Zhiheng P, Yang L, Yingfei M, Purdue M, Jacobs E, Gapstur SM, Huilin L, Alexander A, Hayes R, Ahn J. Cigarette smoking and the oral microbiome in a large study of American adults. ISME J. 2016;10:2435–2446.

15. Lanza E, Magan-Fernandez A, Bermejo B, de Rojas J, Marfil-Alvarez R, Mesa F. Complementary clinical effects of red complex bacteria on generalized periodontitis in a caucasian population. Oral Dis. 2016;22(5):430–7.

16. Welch JLM, Rossetti BJ, Rieken CW, Dewhirst FE, Borisy GG. Biogeography of a human oral microbiome at the micron scale. Proc Natl Acad Sci U S A. 2016;113(6):E791–800.

17. Wecke J, Kersten T, Madela K, Moter A, Göbel UB, Friedmann A, Bernimoulin JP. A novel technique for monitoring the development of bacterial biofilms in human periodontal pockets. FEMS Microbiol Lett. 2000;191(1):95–101.

18. Desvarieux M, Demmer RT, Rundek T, Boden-Albala B, Jacobs DR, Sacco RL, Papapanou PN. Periodontal microbiota and carotid intima-media thickness: The Oral Infections and Vascular Disease Epidemiology Study (INVEST). Circulation. 2005;111(5):576–82.

19. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vázquez-Baeza Y, González A, Morton JT, Mirarab S, Zech Xu Z, Jiang L, Haroon MF, Zhao H, et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017 Nov 1;551(7681):457.

20. Marotz L. Earth Microbiome Project (EMP) high throughput (HTP) DNA extraction protocol. 2018;(2):4–9.

21. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz J, Bittinger K, Brejnrod A, Brislawn C, Titus Brown C, Callahan B, Caraballo-Rodriguez A, Chase J, Cope E, DaSilva R, Diener C and CJ. Reproducible , interactive , scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(August).

22. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Xu ZZ, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems. 2017 Apr 21;2(2):e00191-16.

23. Pedregosa F, Michel V, Grisel O, Blondel M, Prettenhofer P, Weiss R, Vanderplas J, Cournapeau D, Varoquaux G, Gramfort A, Thirion B, Grisel O, Dubourg V, Passos A, Brucher M, Perrot andÉdouardand M, Duchesnay A. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12:2825–30.

24. Janssen S, Mcdonald D, Gonzalez A, Navas-molina JA, Jiang L, Xu Z. Phylogenetic Placement of Exact Amplicon Sequences. mSystems. 2018;3(3):e00021-18.

25. Fedarko M, Martino C, Morton J, González A, Rahman G, Marotz C, Minich J, Allen E KR. Visualizing 'omic feature rankings and log-ratios using Qurro. Nucleic Acids Res Genomics Bioinforma. 2020;

26. Breiman L. Random Forests. Mach Learn. 2001;45:5–32.