UNIVERSITY OF CALIFORNIA,
IRVINE

From a remote Amazon village to the undergraduate lab: How diet and lifestyle shape the human microbiome

DISSERTATION

submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Andrew Oliver

Dissertation Committee:
Dr. Katrine Whiteson, Chair
Dr. Jennifer Martiny
Dr. Elizabeth Bess

2021

DEDICATION

To

My family:
Dan, Evelyn, and Kevin Oliver

TABLE OF CONTENTS

	Page
LIST OF FIGURES	v
LIST OF TABLES	viii
ACKNOWLEDGEMENTS	ix
CURRICULUM VITAE	xi
ABSTRACT OF THE DISSERTATION	xv
INTRODUCTION	1
CHAPER 1: Cervicovaginal microbiome composition is associated with metabolic profiles in healthy pregnancy	7
CHAPTER 2: High fiber, whole foods dietary intervention alters the human gut microbiome but not fecal short-chain fatty acids	47
CHAPTER 3: Characterizing microbiomes of a remote village within the Ecuadorian Amazon	88
SUMMARY AND FUTURE DIRECTIONS	117
REFERENCES	120

LIST OF FIGURES

Figure 1.1 Cervicovaginal microbiome study outline

Figure 1.2 Taxonomy and alpha diversity of vaginal microbiomes during pregnancy

Figure 1.3 Ordination of vaginal microbiomes during pregnancy

Figure 1.4 Relationship between vaginal microbes and metabolites

Figure 1.5 Ordination of functional pathways within the vaginal microbiome

Figure 1.6 Proposed vaginal microbial community model

Figure S1.1 Taxonomic characterization using shotgun metagenomes

Figure S1.2 Fungal taxonomy of cervicovaginal samples

Figure S1.3 A case for mannitol production by *L. crispatus*

Figure S1.4 Metabolic and microbial functional pathway signatures of vaginal microbial communities

Figure S1.5 Ordination of GC-TOF and LC-MS/MS metabolites from samples of saliva and urine from mother and offspring

Figure S1.6 Top GC-TOF and LC-MS/MS saliva metabolites by abundance for mother and offspring

Figure S1.7 Average Bray Curtis similarity of urine and saliva metabolomes between related and unrelated individuals

Figure 2.1 Intervention timeline and sample collection

Figure 2.2 Microbiome community composition through a dietary fiber intervention

Figure 2.3 *GroEL* amplicon analysis of Bifidobacterium during the fiber intervention

Figure 2.4 Genes involved in carbohydrate degradation and SCFA metabolism within metagenomes

Figure 2.5 GC-FID measurements of fecal volatile SCFAs during intervention

Figure S2.1 Comparisons of richness and evenness obtained using different databases for taxonomic assignments

Figure S2.2 Correlations of microbial abundance and fiber intake

Figure S2.3 Mean abundance of *Bifidobacterium* spp. using MIDAS taxonomic counts

Figure S2.4 *Bifidobacterium* phylogenetic analysis

Figure S2.5 Genera that significantly correlate with the abundance of *Bifidobacterium*

Figure S2.6 Abundance of reads mapping to CAZymes

Figure S2.7 Abundance of the inositol degradation pathway across samples

Figure S2.8 Technical variation of SCFAs seen in a subset of samples run in duplicate

Figure 3.1 Description of the Conambo

Figure 3.2 The microbiomes of the Conambo people

Figure 3.3 The influence of cohabitation on the Conambo microbiome

Figure 3.4 Composition and function of Conambo microbiomes in the context of industrialized and non-industrialized microbiomes

Figure S3.1 Richness of Conambo samples using Metaphlan3 and CAT-BAT

Figure S3.2 Taxa bar plots of Conambo samples using Metaphlan3

Figure S3.3 Taxa bar plots of Conambo samples using CAT-BAT

Figure S3.4 Comparing average fecal taxonomies between Metaphlan3 and CAT-BAT

Figure S3.5 Relative abundance of *Streptococcus salivarius* and *S. mitis* in oral Conambo samples

Figure S3.6 Percent of unknown reads in eight industrialized and non-industrialized cohorts

Figure S3.7 NMDS ordination of glycoside hydrolase and polysaccharide lyase abundances in eight industrialized and non-industrialized cohorts

LIST OF TABLES

Table 1.1 Demographics of mothers in cervicovaginal study

Table 2.1 Abundances of four SCFAs in samples pre- and post- intervention

ACKNOWLEDGEMENTS

Firstly, I would like to thank my doctoral advisor Dr. Katrine Whiteson, for fostering my growth as a scientist. Over the years, I have been on the receiving end of many of your questions, and occasionally they would frustrate me! Now, at the end, I find myself asking the same questions of others, and offering similar advice you have always given me. It is on the other side of this degree that I have seen how much I have learned from you. Thank you for taking a chance on me as a graduate student and for creating a research group of some of the best scientists and people I have ever met.

I am also incredibly thankful for my committee, for their thoughtful advice throughout my time at UC Irvine. Specifically Drs. Jennifer Martiny, Elizabeth Bess, Donovan German, and Ilhem Messaoudi. You were all helpful and supportive towards my work, and I appreciate it so much. Additional thanks to Julia Massimelli Sewall for making the fiber project a ton of fun, and being an epic example of a teacher who cares. A special thanks to Dr. Martiny, for being an excellent mentor for me as well. You are an example of a scientist, leader, colleague, and counselor that I will not forget, and certainly will strive to embody more as I continue in science.

A sincere thanks to my lab mates, I could not have done this without you. You are all lifelong friends. Stephen, Tara, Joann, Whitney, Julio, and Jason…I have spent most of the last five years with you and I would certainly do it all over again if you were a part of it.

Friends at UC Irvine I must give a shout out to are Rachel (a billion shout outs are not enough to you, you are wonderful and taught me a PhD amount about being a good person), Jasper, Zane, and Megan. Sloan and Nick, I hope we still keep getting beers. Furthermore, Nick, we did CSUN together, and UCI. You are critical and have made me better at every step. Certainly if you stop following me, I will have to create a Nick voice in my head. One other

thank you to Claudia Weihe, you have been an amazing mentor and friend. I will miss you all terribly.

Thank you to my friends and colleagues from California State University Northridge. Dr. Kerry Cooper, if it was not for you, I would not be receiving my doctorate. I owe you a great debt. The graduate students in the Department of Microbiology at CSUN were like a family to me for many years, thank you.

Thank you to my old friends who have stuck with me all these years. Matt, we have been friends since the first grade, and you have certainly shaped many of my good qualities. Unfortunately I cannot write about everyone, or even list everyone, but a few must appear here: Jenna, Kevin, Jenny, Brad, Erin, Katie, Alex, and Lee.

Finally, the largest thank you to my family: Dan, Evelyn, and Kevin. I am the person I am today because of you. I'm not perfect, but from where I stand, you did an amazing job. I am so lucky to be your son and brother. You are so openly proud of me, and it pushes me to excel. My wins are never too small, never go unnoticed by you. From the bottom of my heart, thank you.

Curriculum vitae of Andrew Oliver

RESEARCH INTERESTS

My overarching research interests are to understand factors that affect the vast taxonomic and functional diversity of human associated microbial and viral communities, especially their contributions to health and disease.

EDUCATION

Ph.D. Molecular Biology and Biochemistry — Aug 2016 – June 2021
University of California Irvine, *Irvine, CA*

M.S. Microbiology — Aug 2013 – May 2016
California State University Northridge, *Northridge, CA*

B.S. Biology — Sep 2008 – May 2012
California Lutheran University, *Thousand Oaks, CA*

RESEARCH EXPERIENCE

Ph.D. Graduate Research — Aug 2016 - Present
University of California, Irvine
Advisor: Katrine Whiteson, PhD

- Investigating the role of microbes in health and disease using high throughput sequencing, metabolomics, and multivariate statistical analysis
- Studied the role of the microbiome during pregnancy and the importance of metabolites in the structure of the community
- Assessed the effect of a high-fiber diet intervention on the microbiome and abundance of short-chain fatty acids
- Collaborated with Dr. John Patton at CSU Fullerton examining the microbiome of the Conambo people of the Amazonian Ecuador.
- Contributed to studying the dynamics of bacteriophage and bacterial co-evolution
- Collaborated with clinical researchers on projects studying blood cancer, Alzheimer's disease, and urinary tract infections

M.S. Graduate Research — Jan 2014 – Aug 2016
California State University Northridge
Advisor: Kerry Cooper, PhD (Currently at U of Arizona)

- Surveillance of foodborne-pathogens, antimicrobial resistance & pathogenesis
- Led a sequenced based and genomic project for an understudied genus of bacteria named *Sporosarcina*
- Pioneered work at CSUN looking at bacterial epigenetics.

PUBLICATIONS

(11) Parker C.T., Huynh S., Alexander A., Oliver A.S., Kerry K Cooper. Genomic epidemiology of Salmonella Typhimurium DT104 strains associated with ground beef**.** *Pathogens.* 2021.

(10) **Oliver, A.,** Chase, A.B., Riedel, S.F., Hendrickson, C., M Lay, Weihe, C., Sewall, J.M., Martiny, J.B.H., Whiteson, K.L. 2019. High fiber, whole foods dietary intervention alters the human gut microbiome but not associated fecal short-chain fatty acids. *mSystems.* 2020.

(9) Adams E. , **Oliver A**., Gille, A., Alaniz, N., Jamie, C., Patton, J., Whiteson, K. Comparing Fecal, Saliva, and Chicha Microbiomes Between Mothers and Children in an Indigenous Ecuadorian Cohort. *Bioarxiv.* 2020. doi: https://doi.org/10.1101/2020.10.02.323097

(8) **Oliver, A**., Lamere, B., Weihe, C., Wandro, S., Lindsay, K.L., Wadhwa, P.D., Mills, D.A., Pride, D., Fiehn, O., Northen, T. de Raad, M., Lynch, S., Martiny, J.B.H., Whiteson., K. Cervicovaginal microbiome composition drives metabolic profiles in healthy pregnancy. *Mbio.* 2020.

(7) Jeney, S.ES., Lane, F., **Oliver, A.**, Whiteson, K., Dutta, S. Fecal microbiota transplantation for the treatment of refractory recurrent urinary tract infection. *Obstetrics & Gynecology.* 2020.

(6) Massimelli Sewall J., **Oliver A.**, Denaro K., Chase AB., Weihe C., Lay M., Martiny JBH, Whiteson K. Fiber Force: A fiber diet intervention in an advanced Course-based Undergraduate Research experience (CURE) course. *J Microbiol Biol Educ*. 2020.

(5) Edwards, RA, *et al*. Global phylogeography and ancient evolution of the widespread human gut virus crAssphage. *Nature Microbiology.* 2020.

(4) Gallagher, T., Phan, J., **Oliver, A.,** Chase, A.B., England, W.E., Wandro, S., Hendrickson, C., Whiteson, K. 2018. Cystic fibrosis-associated *Stenotrophomonas maltophilia* strain-specific adaptations and responses to pH. *Journal of Bacteriology.* 2019.

(3) Phan, J., Gallagher, T., **Oliver, A.,** England, W., Whiteson, K. Fermentation products in the cystic fibrosis airways induce aggregation and dormancy-associated expression profiles in a CF clinical isolate. *FEMS Microbiology Letters.* 2018.

(2) Wandro, S., **Oliver, A.,** Gallagher, T., Weihe, C., England, W., Martiny, J.B.H., Whiteson, K. 2018. Predictable molecular adaptation of coevolving *Enterococcus faecium* and a widespread lytic phage. *Frontiers in Microbiology.* 2018.

(1) **Oliver, A.**, Kay, M., Cooper, K.K. Comparative genomics of cocci-shaped *Sporosarcina* strains with diverse spatial isolation. *BMC Genomics.* 2018.

SUBMITTED

Oliver, A., Haunschild, C., El Aloui, K., Weihe, C., Whiteson, K., Martiny, JBH., Flieschman, A. The role of the gut microbiome in the inflammatory state of MPN patients. *Submitted to Blood Advances.*

PUBLICATIONS IN ADVANCED PREPARATION

Kurbessoian, T., **Oliver, A.,** Cooper, K.K., Baresi, L. Comparative Genomics and Metabolomics of *Sporosarcina* reveals novel spore-forming cocci species.

TEACHING EXPERIENCE

Teaching Associate, General Microbiology & Lab, UC Irvine — Jan 2018 – Jan 2020
Teaching Associate, Microbiology M122, UC Irvine — April 2018 – June 2018

Teaching Associate (*Instructor of Record*), Introduction to Microbiology Lab (BIOL 215L&315L), CSUN	Jul 2014 – Aug 2016
Graduate Assistant, Applied Microbiology Lab, CSUN	Jan 2014 – Jun 2014
Graduate Assistant, Introduction to Microbiology Lab, CSUN	Aug 2013 – Jun 2014
Department Assistant, Anatomy and Physiology, CLU	Sep 2011 – May 2012
Department Assistant, Microbiology, CLU	Jan 2011 – May 2011

SELECTED PRESENTATIONS

Oral Presentation/Lectures

1. *May the Fiber Be with You: Assessing Gut Microbiome Response to a High Fiber Diet Intervention within a Course-Based Undergraduate Research Experience (CURE). 2019 American Society for Microbiology Microbe Conference, San Francisco, CA*
2. *Lend me your sugar, I am your neighbor: Vaginal microbiome clusters drive metabolic profiles in healthy pregnancy. 2018 Department of Molecular Biology and Biochemistry Student seminars, University of California Irvine, Irvine, CA.*
3. *The Human Microbiome and Antibiotic Resistance. 2017 Osher Lifelong Learning Institute* Lecture (OLLI) Series, Irvine, CA
4. Introduction to Genomic Analysis of Hospital Acquired MRSA. 2017 UC Irvine Medical Center, Irvine, CA
5. Coevolution of *Enterococcus* with lytic phage V12. 2017 San Diego State University, *San Diego, CA.*
6. Comparative Genomics and Epigenetics of Six Strains of *Sporosarcina ureae*. 2016 Biology Student Research Symposium, California State University Northridge, *Northridge, CA*.
7. Comparative Genomics and Epigenetics of Six Strains of *Sporosarcina ureae*. 2015 Sigma Xi Student Symposium, California State University Northridge, *Northridge, CA*.

Poster

1. Parker, C.T., Huynh, S., **Oliver, A.S**., Cooper, K.K. Whole Genome Sequence Comparisons of Salmonella Enteritidis Strains Isolated from Samples Associated with Almond Outbreaks. 2014 American Society for Microbiology, Ernest N. Morial Convention Center, *New Orleans, LA.*
2. Cooper, K.K., **Oliver, A.S.,** Sams, C., Korlach, J., Clark, T.A., Loung, K., Huynh, S., Parker, C.T., Mandrell, R.E., Carter, M.Q. Role of DNA methylation in enterohemorrhagic *Escherichia coli* curli expression.
3. **Oliver, A.S.,** Kurbessoian, T., Pregerson, B., Cooper, K.K. Comparative Genomics and Epigenetics of Six Strains of *Sporosarcina ureae*. 2016 American Society for Microbiology MICROBE Meeting, *Boston, MA*.
4. **Oliver, A.S.,** Kurbessoian, T., Pregerson, B., Cooper, K.K. Comparative Genomics and Epigenetics of Six Strains of *Sporosarcina ureae*. 2016 CSU Biotechnology Symposium, *Anaheim, CA*.
5. Wandro, S., **Oliver, A.,** Gallagher, T., Weihe, C., England, W., Martiny, J.B.H., Whiteson, K. *In preparation.* Coevolution between *Enterococcus faecium* and lytic phage V12. 2017 Evergreen Phage Meeting, *Evergreen, WA*
6. **Oliver, A.,** LaMere, B., Wandro, S., Martiny, J., Lindsay, K., Wahdwa, P., Pride, D., Lynch, S., Whiteson, K. Vaginal Microbiome clusters drive metabolic profiles in healthy pregnancy. 2018 Lake Arrowhead Microbial Genomics, *Lake Arrowhead, CA* (Awarded poster)

FELLOWSHIPS AND AWARDS

Rose Hill Fellowship	2020
Edward K. Wagner Award in Virology	2020

NIH T32 Training Grant: Microbiology and Infectious Diseases	2019
American Society of Microbiology Travel Grant	2019
Department Poster Award, MBB Retreat	2019
Lake Arrowhead Microbial Genomics Best Abstract & Poster	2018

PROFESSIONAL SOCIETY MEMBERSHIPS

American Microbiology Society
Tri-Beta Biological Honors Society

SKILLS

BIOINFORMATICS

- Data analysis: metagenomics, comparative genomics, metabolomics
- Packages/Programs: Tidyverse, Renv, lme4, nlme, RFpermute, Biobakary, MIDAS, Geneious, iTOL, Anvio
- Unix/Linux, Bash, R, Python, remote computing, bioinformatic pipeline experience, GitHub
- Multi-variate statistical analysis

LABORATORY

- Gas-chromatography (GC-FID)
- Preparing NGS sequencing libraries, culturing microbes (aerobic and anaerobic, BSL1 and BSL2), microscopy and PCR
- Bacteriophage isolation and propagation; host-range assay
- Next-generation sequencing: Knowledge in all steps from taking a microorganism from nature to DNA, and from DNA to genome sequence.
- Basic laboratory and microbiological skills (media making, culturing, aseptic technique)

ABSTRACT OF THE DISSERTATION

Investigating human microbiomes in health and disease

by

Andrew Oliver
Doctor of Philosophy in Biological Sciences
University of California, Irvine, 2021
Dr. Katrine Whiteson, Chair

The lifespan of people living in westernized countries is increasing, suggesting we are getting healthier; however, we are now struggling with a new health problem: the increasing incidence of chronic disease. We have only recently begun to appreciate the association of microbial colonization with health, and how changes in our lifestyles (i.e. diet, urbanization, sanitation) are affecting this association. Changes in microbial community composition have been linked with numerous acute and chronic disease states. Despite the microbiome's strong correlation with health, there is still much work to be done in understanding the ideal structure and function of these human associated microbial communities.

Humans first experience with microorganisms occurs during gestation, when we are exposed to the metabolites of the maternal microbiome. Maternal microbes and their metabolic by-products critically influence the structuring and functioning of the early infant microbiome and immune system; however, global assessments of maternal microbes and metabolites are lacking. In Chapter 1, my co-authors and I address this knowledge gap by sequencing cervicovaginal fluid from 18 healthy mothers longitudinally throughout their pregnancy. Parallel to microbiome sequencing, we examined the vaginal metabolomes using GF-TOF and LC-MS/MS to broadly capture metabolites within this system. Additionally, we analyzed the metabolomes of offspring through their first year of life. We report strong metabolic biomarkers of microbiome composition

within the cervicovaginal tract, which may ultimately be useful in future diagnostic and therapeutic endeavors for vaginal diseases like bacterial vaginosis (BV). We show a strong correlation between mannitol and the presence of *Lactobacillus crispatus* and speculate as to why a homofermentative vaginal lactobacillus may be generating mannitol via fructose reduction for redox purposes. In chapter 1, I focus on describing the early microbes and metabolites infants are exposed to during pregnancy and birth.

Throughout our lives, our microbiomes are under constant selection by intrinsic (e.g., phage predation) and extrinsic (e.g., antibiotic use) factors. One strong architect of gut microbial composition is diet. Dietary fiber is a favored substrate of microbes deep within the colon, which is able to resist digestion by endogenous enzymes in order to feed the dense bacterial communities throughout the gut. In chapter 2, we conducted a two-week diet intervention study, asking whether an increase in dietary fiber intake through whole foods was capable of changing the gut microbial communities and associated metabolome. We show on a short time scale, increasing dietary fiber in the gut alters the microbial composition significantly, but does not change the abundance of metabolic by-products of fiber fermentation (short-chain fatty acids). While more work needs to be done to disentangle the relationship between how diet influences bacterial metabolites, our work illustrates how malleable gut microbial communities are to a change in diet.

Finally, in chapter 3, I contributed context to a growing body of microbiome research that is woefully lacking microbiome characterizations from non-industrialized populations. We teamed up with anthropologists to study the microbiome of a remote community of people deep within the Ecuadorian Amazon. Our findings show that the microbiomes from these Conambo people are distinct from the microbiomes of industrialized individuals, containing a high abundance of VANISH taxa (volatile and/or associated negatively with industrialized societies of humans).

Notably, nearly 40% of the average fecal metagenome consisted of reads mapping to *Prevotellaceae.* The Conambo people have unique diets, with a significant portion of their caloric intake coming from a cultural variant of cassava beer ("chicha"). Chicha is made by mastication and re-mastication of yuca root by women within each household. We show that while the initial inoculums of chicha are household specific, over the course of fermentation the communities converge on a more similar microbiome irrespective of household origin. These data contribute vitally towards the open question: how has industrialization shaped the microbiome?

INTRODUCTION

Communities of microbes have established themselves in every environment around and within us. Humans are host to billions of microbes, collectively called the microbiome, and these cells are estimated to be at least as numerous as the eukaryotic cells that make up the human body itself (1). From a functional perspective it is estimated that more than half of the ~500,000 metabolites within the human body are made or modified by microorganisms (2). Although initially overwhelming in its complexity, it has been suggested that studying the microbiome can be boiled down to the basic tenets of community ecology, specifically: dispersal, diversification, selection, and drift (3). Indeed, perhaps the most visible study in this field, The Human Microbiome Project (HMP) (4), suggests that these ecological principles can be used to answer fundamental questions about our relationship with microbes, such as (1) what is the stability and resilience of the microbiome, (2) what factors drive microbiomes to look more similar or different from each other, and (3) how do people develop and transmit their microbiomes? The following chapters aim to contribute to these still open questions.

Understanding how individuals acquire and develop their microbiomes in early life is critical, as this developmental trajectory may play a large role in the overall health of the child and young adult. Infants begin physiologically responding to microbes *in utero* and during birth, when they are exposed to maternal microbes and metabolic products (5). In a study of 169 women from a Norwegian birth cohort (NoMIC), Stanislawski et al. found associations between the gut microbiota of some mothers and the microbes that initially colonized their children (6). Another cohort study showed regular prenatal and post-natal farm exposure, i.e., contact with a diversity of microbes, in infants reduced the incidence of chronic health diseases such as asthma and atopy (7). Recent research has supported the idea of fetal programming, where maternal

antibodies prepare the infant gut for the onslaught of colonizing microbes (8). However, it is unclear to what extent the infant metabolome is molded by the microbiome and metabolome of its mother.

Even after the microbiome is acquired in early life, it undergoes constant selection by various lifestyle factors ranging from cohabitation (9) and host genetics (10), to the immune system (2). Changes in an individual's diet can also have a significant effect on the gut microbiota, by potentially altering pH and the availability of various carbon substrates (11). Diet induced changes to the microbiome may have profound impacts on health. One study showed that the high fiber diets of rural South Africans significantly suppressed the risk factors associated with colon cancer, perhaps due in part to increased butyrogenesis (12). Butyrate, and its sister short chain fatty acids (SCFA) acetate and propionate, are produced by the gut microbiota via the degradation of dietary fiber in the colon (13). Colonocytes use the high levels of butyrate in the colon as their primary energy source (reviewed in Clemente et al., 2012). SCFAs have been shown to have beneficial effects ranging from increased intestinal mucus, upregulated production of tight junctions, both of which play roles in immune homeostasis (reviewed in (15)). Butyrate (5 g/kg tributyrin) specifically has been experimentally shown to bind to the PPAR-γ receptor in the colon epithelia of mice, which downregulates nitrate production and helps maintain a hypoxic environment in the gut, suppressing the growth of facultative anaerobes from the family *Enterobacteriaceae* that would otherwise lead to dysbiosis (16). Elegant work has been done to show the beneficial effects of SCFAs on health; however, their benefits may be contextual (i.e. if they are present in the gut vs blood serum). Recent work has shown that fecal SCFAs may be associated with obesity and hypertension (17). Propionate has also specifically been shown to increase glucagon and fatty acid–binding protein 4 in the

plasma of mice, and in humans, higher plasma levels were indicative of poor insulin sensitivity (18). Since these SCFAs result from the bacterial metabolism of dietary fiber, and are potent mediators of health, studying how diet can affect the microbiome will help us better understand the therapeutic value of nutrition and prebiotics.

Despite a growing body of research that has unmasked many factors that shape the composition and function of the human microbiome, far less has been done to put that in the context of evolution of the human lifestyle (e.g., the advent of industrialization and modern medicine). Moreover, due to the increasing expansion of western diet and culture, it is becoming challenging to study non-industrialized microbiomes, which will give insight into how urbanization has altered the human microbiome. A few cohorts exist that illustrate the profound differences between industrialized and non-industrialized microbiomes (19–24). Many of these studies show that rural microbiomes are different in their composition from western microbiomes and are enriched for bacteria such as *Prevotellaceae* and *Succinivibrionaceae* (23). One hypothesis for the contributions to the increasing incidence of chronic disease may arise from differences between microbiomes associated with industrialization compared to non-industrialized microbiomes (25). The following research investigates the non-industrialized microbiome using a new cohort of individuals living in the Conambo River Valley, deep within Amazonian Ecuador (26–28). Interestingly, the main source of calories for the people of Conambo comes from chicha, a fermented beverage made, in part, by inoculating saliva into a root pulp. In this community, chicha is made only by that maternal line within a household. This work attempts to contribute to the characterization of pre-modern microbiomes by sequencing the bacteria, fungi, and other understudied microbes of the Conambo people, which will further highlight the effects of industrialization on the microbiome.

Ultimately it remains a challenge to uncover what factors shape the microbiome. In a large cross-sectional study of 4,000 individuals, researchers tracked 69 covariates and found they explained about 8% of the variation in the human associated microbial communities (29). Putting the individualistic nature of the microbiome another way, Ed Yong for The New York Times writes "The microbiome is the sum of our experiences throughout our lives: the genes we inherited, the drugs we took, the food we ate, the hands we shook." The more variables we measure, the better we might grasp whether the microbiome behaves in predictable ways, or if there is a certain stochasticity to it. For example, the Anna Karenina principle suggests that "All healthy microbiomes are similar; each dysbiotic microbiome is dysbiotic in its own way" (30). How much of that stochasticity is a reflection of our struggle to find meaningful and measurable covariates for the microbiome remains a forefront problem in the field. To that end, studies designed to measure which factors contribute to the composition of the microbiome and how it is transmitted will be continually useful.

Goals and scope of this dissertation

Because the microbiome is so personalized, and vast in terms of diversity and genetic information, it remains a current challenge to identify measurable covariates that give us insight into the structure of the microbial communities present. Just as our human genomes make us unique, it is likely that no single (or several) covariate will be generalizable to the system. The goal of this work is to characterize the microbiome in health, and put the microbiome in context of covariates, such as metabolites, diet, and lifestyle.

Aim 1: Examine the vaginal microbiome and metabolome through pregnancy (Chapter 1)

We used metagenomic and amplicon sequence for comprehensive characterization of the vaginal microbial communities during pregnancy. This allowed us to better understand what

strains are present during pregnancy and what is the functional potential of these communities. Gas chromatography – mass spectrometry and liquid chromatography – mass spectrometry was used to characterize the metabolomes that are present during pregnancy and the early life of the infant. We originally hypothesized that there are metabolic biomarkers that are highly correlated with bacterial community composition in the vagina during pregnancy. Furthermore, as pregnancy has previously been shown to stabilize the vaginal microbiome (31), we examined whether that holds true with our longitudinal data, and if the pregnancy metabolome is also stable during this period. This work has revealed highly specific biomarkers for the vaginal microbiome, and suggests that the abundance and types of vaginal host sugars contribute importantly to the structure and maintenance of the vaginal communities

Aim 2: Study the effect of a high fiber diet intervention on the gut microbiome of healthy individuals (Chapter 2)

One major outcome of the westernization of diet is a decrease in dietary fiber. Most Americans get only ~15 grams/day of fiber, well under the suggested 30 grams/day (32). Since diet is known to alter the microbiome (11), we sought to determine how an increase in dietary fiber changes the composition of the microbiome. Dietary fiber is particularly important, because it is a carbohydrate that persists through the small intestine and into the colon, where it feeds the dense microbial population residing there. Using students enrolled in a course-based undergraduate research experience (CURE) undergraduate microbiology course as subjects, we assessed how a two-week diet intervention consisting of 40+ grams/day of fiber alters the microbiome. Moreover, because microbes use dietary fiber to produce short chain fatty acids (SCFAs), molecules generally implicated in health benefits, we measured the SCFA profile of these individuals. We showed that increasing dietary fiber, a nutrient present in high abundance

in the diets of traditional societies, is a tractable method for altering microbiome composition, but not necessarily the abundance or composition of fecal SCFAs.

Aim 3: Characterize the microbiome composition within the Conambo tribe of Ecuador (Chapter 3)

The extent to which environmental factors and host genetics shape the microbiome remains a critical question in understanding host associated microbial communities. Cohabiting family members have been shown to share significant portions of their microbiomes with one another (9), and even unrelated roommates share more of their viruses with each other than they do with members of different households (33). We sought to examine this question in the context of the Conambo tribe from the Ecuadorian Amazon. Specifically, we wish to address three questions: 1) What is the microbiome composition of people of the Amazonian Ecuador, who have little to no contact with urban civilization, 2) how does the microbiomes of rural Amazon differ from westernized microbiomes, and 3) does the Conambo people's main source of calories, chicha, help mediate the passing of strains between mothers and infants and between relatives.

CHAPTER 1:

Cervicovaginal microbiome composition is associated with metabolic profiles in healthy pregnancy

Authors: Andrew Oliver, Brandon LaMere, Claudia Weihe, Stephen Wandro, Karen L. Lindsay, Pathik D. Wadhwa, David A. Mills, David Pride, Oliver Fiehn, Trent Northen, Markus de Raad, Huiying Li, Jennifer B.H. Martiny, Susan Lynch, Katrine Whiteson

**All supplemental tables contain datasets which can be found at:*
DOI: *10.1128/mBio.01851-20*

Abstract

Microbes and their metabolic products influence early-life immune and microbiome development, yet remain understudied during pregnancy. Vaginal microbial communities are typically dominated by one or a few well adapted microbes, which are able to survive in a narrow pH range and are adapted to live on host-derived carbon sources, likely sourced from glycogen and mucin present in the vaginal environment. We characterized the cervicovaginal microbiomes of 16 healthy women throughout the three trimesters of pregnancy. Additionally, we analyzed saliva and urine metabolomes using GC-TOF and LC-MS/MS lipidomics approaches for samples from mothers and their infants through the first year of life. Amplicon sequencing revealed most women had either a simple community with one highly abundant species of *Lactobacillus* or a more diverse community characterized by a high abundance of *Gardnerella,* as has also been previously described in several independent cohorts. Integrating GC-TOF and lipidomics data with amplicon sequencing, we found metabolites that distinctly associate with particular communities. For example, cervicovaginal microbial communities dominated by *Lactobacillus crispatus* also have high mannitol levels, which is unexpected given the characterization of *L. crispatus* as a homofermentative *Lactobacillus* species. It may be that

fluctuations in which *Lactobacillus* dominate a particular vaginal microbiome are dictated by the availability of host sugars, such as fructose, which is the most likely substrate being converted to mannitol. Overall, using a multi-'omic approach, we begin to address the genetic and molecular means by which a particular vaginal microbiome becomes vulnerable to large changes in composition.

Importance

Humans have a unique vaginal microbiome compared to other mammals, characterized by low diversity and often dominated by *Lactobacillus* spp.. Dramatic shifts in vaginal microbial communities sometimes contribute to the presence of a polymicrobial overgrowth condition called bacterial vaginosis (BV). However, many healthy women lacking BV symptoms have vaginal microbiomes dominated by microbes associated with BV, resulting in debate about the definition of a healthy vaginal microbiome. Despite substantial evidence that the reproductive health of a woman depends on the vaginal microbiota, future therapies which may improve reproductive health outcomes are stalled due to limited understanding surrounding the ecology of the vaginal microbiome. Here, we use sequencing and metabolomic techniques to show novel associations between vaginal microbes and metabolites during healthy pregnancy. We speculate these associations underlie microbiome dynamics, and may contribute to a better understanding of transitions between alternative vaginal microbiome compositions.

Introduction

Vaginal microbes sustain important physiologies and produce metabolites that can directly or indirectly affect maternal health and infant development during pregnancy. Perturbations to early-life microbiomes and associated metabolic dysfunction have been linked with allergy and auto-immune diseases such as asthma (34–37). For example, regular pre-natal

and post-natal farm exposure, i.e., contact with a diversity of microbes during pregnancy and infancy, have been shown to reduce the incidence of chronic health diseases such as asthma and atopy (7). Moreover, recent research has supported the idea of fetal programming, a term describing the process by which the maternal microbiota, as well as maternal antibodies, prepare the infant immune system for the post-natal onslaught of colonizing microbes (8). Others have shown in mice that vaginal dysbiosis, induced by maternal stress, has the potential to negatively affect offspring metabolic profiles (38). Thus maternal microbes, particularly those of the vaginal tract, are some of the first microbes the offspring will encounter and may be central to the study of early-life microbiome and immune development (39–43). Indeed, a recent large-scale study of 2,582 women, over 600 of whom were pregnant (a subset of whom were longitudinally sampled), provided evidence for vaginal microbiome restructuring during pregnancy toward a *Lactobacillus*-dominated community (44). This occurred early in gestation and was associated with a reduced vaginal microbiome metabolic capacity. Post-partum, irrespective of the mode of delivery, the vaginal microbiota resembled that of a gastrointestinal microbiome, likely due to microbial mixing during the birthing process (40), suggesting that both vaginal and gastrointestinal microbial seeding of the neonate occurs.

The human vaginal microbiome maintains low diversity in low pH conditions, and depends on host sugars as carbon sources, with less access to dietary and exogenous nutrients compared to the gut or the skin or the oral cavity. Historically, vaginal microbial communities have been stratified based on hierarchical clustering of the taxa composition (43). Keystone species include *Lactobacillus crispatus* and *L. gasseri* which have been associated with maintenance of a simple vaginal microbiome by their production of bacteriostatic and bactericidal compounds (e.g. lactic acid and hydrogen peroxide) and maintenance of a low pH

(45–47), numerically and functionally dominating their respective vaginal communities. A closely related species, *L. iners*, has been associated with health promoting benefits; however, its genome also encodes the capacity to promote microbiome perturbation by increasing vaginal pH and producing species specific virulence factors (45, 48–50). Bacterial vaginosis (BV), the most common gynecological condition in reproductive aged women (51), is characterized by the presence of a more diverse vaginal microbiome and associated with adverse pregnancy outcomes including preterm birth (52), endometritis (53, 54) and spontaneous abortion (55–58). Recently, vaginal microbial transplants have been successfully implemented as a treatment for intractable BV (59). Despite *L. crispatus* generally being regarded as a highly beneficial and dominant microbe throughout pregnancy, healthy women from different ethnic groups have markedly different species dominating the vaginal microbiome (46). In fact, many healthy women that lack BV symptoms have vaginal microbiomes dominated by microbes that are associated with BV (60), suggesting that taxonomy alone is insufficient to predict health outcomes and that microbial activities, including metabolic productivity may offer a more contemporary view of microbiome function.

An untargeted, more global assessment of microbiomes and associated metabolites during pregnancy and early-life is lacking. To address this gap in knowledge, we collected saliva, urine, and cervical vaginal fluid (CVF) from 18 mothers during each trimester of pregnancy and saliva and urine from offspring through their first year of life. Specifically, we were interested in how maternal CVF microbiome profiles are associated with metabolomic assessments of the same samples. Furthermore, we had the opportunity to examine whether maternal saliva and urine metabolome profiles relate to those of the infant in the first year of life. Here we present DNA sequencing (amplicon and shotgun) and untargeted metabolomics to

characterize microbial and metabolic features of the CVF microbiome throughout pregnancy to determine the composition of the vaginal microbiome from a cohort of healthy Caucasian and Hispanic women, longitudinally sampled throughout a healthy pregnancy.

Methods

Subject Information. Eighteen women were selected from a larger cohort recruited to address how maternal stress affects child development (61, 62) (Table 1.1.1). Inclusion criteria for the larger cohort included >18 years of age, singleton, intrauterine pregnancy, and non-diabetic. Additional inclusion criteria for this study were: normal pre-pregnancy body mass index, vaginal delivery, full-term pregnancy, breastfeeding, and no antibiotics for mother or baby. Generally, these eighteen women and their children represented healthy subjects with the most complete sample sets.

Sample collection. Samples were collected at each trimester of pregnancy for women and through the first year of life for infants. At each timepoint, maternal saliva, urine and cervical vaginal fluid were collected. For infants, urine was collected at birth, six months, and twelve months whereas saliva was sampled at six months and twelve months of age. Maternal saliva was collected using a salivette collection kit, including a small cotton roll contained in a plastic container (Salimetrics, Carlsbad, CA). Mothers were instructed to place the cotton rolls in their mouth until saturated with saliva (approximately 1-3 minutes), and then reseal the swabs in plastic salivette tubes. Infant saliva was collected using Weck-Cel spears and a swab extraction tube system. Infants were allowed to suck on the spear for two minutes, ensuring saturation. Maternal urine was collected using a sterile collection cup. Infant urine was collected using an adhesive u-bag attached to the genital region of the infant. A minimum of 2 ml of urine was collected. Cervical vaginal fluid (CVF) was collected by placing three Dacron swabs into the

cervix for ten seconds to achieve saturation. Each swab was then placed in a plastic vial with 500 µl of sterile PBS. All samples were initially stored at -20°C and then saliva, CVF, and infant urine were subsequently moved to -80°C storage.

Metabolomics. Prior to processing, samples were thawed from -80°C storage. Fifty microliters of each sample were subjected to gas chromatography time-of-flight mass spectrometry (GC-TOF) (63) and liquid chromatography accurate mass mass-spectrometry (LC-MS/MS, lipidomics). Urine, saliva, and CVF from each time point were sent to the West Coast Metabolomics Center (WCMC) for untargeted metabolomics. GC-TOF metabolites were extracted with a mixture of 3:3:2 acetonitrile:isopropyl alcohol:water according to standard operating procedures from the Fiehn Lab at the WCMC (64). LC-MS/MS samples were extracted using a variant of the Matyash method (65). Data were acquired for complex lipids in positive and negative electrospray mode on a Waters CSH column and an Agilent 6530 QTOF mass spectrometer (65). Metabolites were identified by retention time MS/MS matching using MassBank of North America (http://massbank.us) and NIST17 libraries.

High-throughput metabolomics. All urine, saliva and CVF were analyzed using Matrix Assisted Laser Desorption Ionization (MALDI) Imaging Mass Spectrometry (MSI) for high-throughput untargeted metabolomics. Extracted samples in 3:3:2 acetonitrile:isopropyl alcohol:water were diluted 1:2 in water in 384 well plates. Next, an equal volume of MALDI matrix (20 mg/mL of 1:1 2,5-Dihydroxybenzoic acid and α-cyano-4-hydroxycinnamic acid in 1:3 (v/v) H2O:MeOH+0.2% formic acid) was added. Samples were printed onto a stainless steel blank MALDI plate using an ATS-100 acoustic transfer system (BioSera) with a sample deposition volume of 10nl. Samples were printed in clusters of four replicates, with the microarray spot pitch (center-to-center distance) set at 900 um). MS-based imaging was

performed using an ABI/Sciex 5800 MALDI TOF/TOF mass spectrometer with laser intensity of 3,500 over a mass range of 50–3,000 Da. Each position accumulated 20 laser shots. The instrument was controlled using the MALDI-MSI 4800 Imaging Tool. Surface rasterization was oversampled using a 75 um step size. Average ion intensity for all reported ions was determined using the OpenMSI Arrayed Analysis Toolkit (OMAAT) software package (66).

DNA extraction. Cervical brushes were resuspended in PBS. Two negative extraction controls using sterile PBS were prepared alongside the samples. Aliquots of 100-200 µl were added to lysing matrix E tubes pre-aliquoted with 500 of hexadecyltrimethylammonium bromide (CTAB) DNA extraction buffer and incubated at 65°C for 15 minutes. An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added to each tube and samples were homogenized in a Fast Prep-24 homogenizer at 5.5 m/s for 30 seconds. Tubes were centrifuged for 5 minutes at 16,000 x *g* and the aqueous phase was transferred to individual wells of a 2 ml 96-well plate. An additional 500 µl of CTAB buffer was added to the lysing matrix E tubes, the previous steps were repeated, and the aqueous phases from paired extractions were combined. An equal volume of chloroform was mixed with each sample, followed by centrifugation at 3000 x *g* for 10 minutes. The aqueous phase (600 µl) was transferred to a clean 2 ml 96-well plate, combined with 2 volume-equivalents of polyethylene glycol (PEG) and stored overnight at 4°C to precipitate DNA. Plates were centrifuged for 60 min at 3000 x *g*. DNA pellets were washed twice with 300 µl of 70% ethanol, air-dried for 10 minutes and re-suspended in 100 µl of sterile water. DNA was quantified using the Qubit dsDNA HS Assay Kit and diluted to 10 ng/µl, when possible. Although DNA was extracted from CVF, attempts to extract DNA from saliva were unsuccessful, potentially due to the storage swabs trapping the biomaterial.

Amplicon gene sequencing. To amplify the V4 region of the bacterial 16S rRNA gene, 10 ng of DNA template was combined with PCR master mix (0.2 mM dNTP mix, 0.56 mg/ml BSA, 0.4 uM Illumina adapter sequence-tagged forward primer (515F) (67), 0.025 U/µl Taq DNA polymerase) and 0.4 uM barcode-tagged reverse primers (806R) then amplified in triplicate 25 µl reactions, along with a no-template control, for 30 cycles (98°C for 2 min; 98°C for 20 sec, 50°C for 30 sec, 72°C for 45 sec; repeat steps 2-4 29 times; 72°C for 10 min). PCR conditions were identical for ITS2 amplification (primer pair fITS7 (5′-GTGARTCATCGAATCTTTG-3′) and ITS4 (5′-TCCTCCGCTTATTGATATGC-3′)) except for the annealing temperature, which was 52°C. Triplicate reactions were combined and purified using the SequalPrep Normalization Plate Kit (Invitrogen) according to manufacturer's specifications. Purified amplicons were quantified using the Qubit dsDNA HS Assay Kit and pooled at equimolar concentrations. The amplicon library was concentrated using the Agencourt AMPure XP system (Beckman-Coulter), quantified using the KAPA Library Quantification Kit (KAPA Biosystems) and diluted to 2nM. Equimolar PhiX was added at 40% final volume to the amplicon library; the 16S rRNA amplicon pool was sequenced on the Illumina NextSeq 500 Platform on a 153bp x 153bp sequencing run, and the ITS2 amplicon pool was sequenced on the Illumina MiSeq platform on a 290bp x 290bp run.

Shotgun metagenomics sequencing. Sequencing libraries were prepared using the Illumina Nextera kit and methods described in Baym et al. (68). Briefly, DNA from each sample was diluted to 0.5ng/µl and tagmented with the Nextera enzyme (Illumina) for 10 min at 55°C. Following tagmentation, each sample received 1 µl forward and 1 µl reverse barcodes, which were added via PCR using Phusion DNA polymerase (New England BioLabs). After PCR, the libraries were cleaned of smaller DNA fragments, using AMPure XP magnetic beads (Beckman-

Coulter), and pooled by concentration. Libraries were quantified using the Quanti-iT PicoGreen dsDNA kit (Thermo Fischer Scientific), and DNA was run on a gel to check fragment size. These libraries were loaded onto the Illumina Next-Seq 500 at 1.8 picomolar concentrations and Illumina's mid-output kit for 75 bp paired-end sequencing.

OTU Table Generation. Raw sequence data were converted from bcl to fastq format using bcl2fastq v2.16.0.10. Paired sequencing reads with a minimum overlap of 25bp were merged using FLASH v1.2.11. Index sequences were extracted from successfully merged reads and demultiplexed in the absence of quality filtering in QIIME (Quantitative Insights Into Microbial Ecology, v1.9.1), and reads with more than two expected errors were removed using USEARCH's fastq filter (v7.0.1001). Remaining reads were de-replicated, clustered into operational taxonomic units (OTUs) at 97% sequence identity, filtered to remove chimeric sequences, and mapped back to OTUs using USEARCH v8.0.1623. Taxonomy was assigned with the most current Greengenes database for bacteria (67) (May 2013), and UNITE vers. 6 for fungi (69). OTUs detected in Negative Extraction Controls (NECs) were considered potential contaminants and filtered by subtracting the maximum NEC read count from all samples; any remaining OTU with a total read count less than 0.001% of the total read count across all samples was removed. Sequencing reads were rarefied to an even depth (28,972 reads for 16S; 91,232 reads for ITS2). To maximize similarity between the raw and rarefied OTU tables, random subsampling was performed at predefined depths 100 times, and the most representative subsampled community was selected based on the minimum Euclidean distance to the other OTU vectors generated for each sample.

16S rRNA Gene Analysis. Alpha-diversity indices and Bray-Curtis dissimilarity matrices were generated in QIIME (70). Linear outcomes were assessed by linear mixed-effects (LME)

modeling to adjust for repeated measures using the nlme package (71) in the R environment (72). Variables of $p<0.05$ were considered statistically significant. Data was visualized using Tableau and Adobe Illustrator unless otherwise noted.

Metagenomic Analysis. Raw sequences (mean 2,977,881 paired-end reads per sample from 35/38 successfully sequenced samples) were filtered using Prinseq v0.20.4 (73) to filter out sequences that had a mean quality score of 30 or less. Human DNA was next filtered out by aligning the filtered reads to the human genome (hg38) using Bowtie2 v2.2.7 (74), and keeping the reads that failed to align (mean 220,355 paired-end + 33,744 singleton reads per sample or 10.9% of quality filtered reads per sample). To analyze functional potential, the reads were run through HUMAnN2 v0.1.9 (75) using default parameters and differences in pathway abundances were analyzed using LEfSe (76). These reads were also cross assembled using SPAdes v3.8.2 (77). Each sample was then mapped to this cross-assembly using Bowtie2, and samples from the same subject were merged together using Samtools v1.9 and the resulting bam files and the cross-assembly were imported into Anvio4 (78). Taxonomy was assigned to each gene call using Kaiju (79) which subsequently informed a more accurate metagenomics binning of the most abundant microbes present.

Statistical Analysis. Unless otherwise noted, statistics were done using the ecological statistics program Primer-e v7 (80). Metabolic data were normalized in Primer-e by dividing by sum total for each sample. The specific programs used in Primer-e were permutational multivariate analysis of variance (PERMANOVA) and distance-based linear models (DistLM), the former of which calculates the significance and variance explained by a given factor and the latter determining which environmental variables correlate with the biological (microbiome) data. RFPermute (81), an R package for permutated random forests, was also performed to determine

which annotated GC metabolites were indicative of microbial composition. PERMANOVA also partitions variance based on each factor, which is done in Primer-e by dividing the factor estimate by the sum total estimates of components of variation (ECoV). Traditional R^2 values were also calculated by dividing the sum of squares by total sum of squares. LMEs were carried out as described above; R^2 values for linear mixed models was calculated using the MuMIn package in R (82). Relate tests (analogous to Mantel tests), were used to compare GCMS and LCMS data. Bray Curtis distances were used for all distance-based analyses. To consider repeated measures, linear mixed effects modeling (nlme package in R) was used to analyze stability of the microbiome and metabolome through time.

Results

Description of cohort and data obtained from samples

Saliva, urine, and cervical vaginal fluid (CVF) were collected from eighteen women, at early, middle and late pregnancy with the gestational age range of the included women at each timepoint (Fig. 1.1). At the time of enrollment into the cohort, the average woman's age was 27.8 years old and average pre-pregnancy BMI was 24.8 (Table 1.1). The cohort was 39% white Hispanic and 61% non-Hispanic white; there were no significant differences in BMI (T-test, P=0.28) or age (T-test, P=0.89) between ethnic groups. Saliva and urine were collected at indicated intervals from each infant up until one year of age (Fig. 1.1). Saliva, urine, and CVF samples were subjected to metabolomics analysis whereas only CVF was used for sequence analysis. Sequence analysis included amplicon-based sequencing of the 16S rRNA gene (bacteria & archaea) and ITS2 (fungi) loci, and shotgun metagenomic sequencing of the entire microbial community.

Vaginal microbiota support high abundances of Lactobacillus *and* Bifidobacteriaceae *throughout pregnancy*

Sixteen individuals (42 total samples) produced sufficient sequence reads for taxonomic assignment using the 16S rRNA gene. Amplicon sequencing stratified cervical samples into those where the most abundant taxon was *Lactobacillus spp.* (34/42, 81%) or *Gardnerella spp.* (8/42, 19%) (Fig. 1.2A, Fig. S1.1A). The bacterial taxa in samples with abundant *Lactobacillus spp.* were significantly less evenly distributed (LME, P=0.001, Fig. 2C), with *Gardnerella vaginalis* being the most abundant in seven of eight samples (88% relative abundance) and a *Shuttleworthia* taxon being most abundant in one sample (at 23% relative abundance). In samples where *Lactobacillus spp.* had the highest abundance, a single taxon comprised 50% or more of the sequencing reads (27/34, 79%). *L. iners* was the most abundant taxon detected in 14 samples from 7 subjects, with a median relative-abundance of 79%. In twelve samples from 7 subjects, a *Lactobacillus* taxon, putatively identified as *L. crispatus* through metagenomic sequencing (Fig. S1.1B,C), had a median relative abundance of 96%, and persisted at a relative abundance greater than 90% in subjects 1088, 1120, and 1191. Altogether, the most abundant 3 taxa (*L. iners* (OTU_1), *L. crispatus* (OTU_2), and *G. vaginalis* (OTU_3) comprised 66% of the total bacterial sequencing reads.

Whole genome shotgun sequencing produced, on average, 2.9 million paired-end reads per sample, which decreased to an average of 220,355 paired-end reads per sample following removal of reads that aligned to the human genome. Thirteen individuals (35 total samples) produced sufficient sequence reads for taxonomic assignment, which was concordant with 16S rRNA gene sequence results (Fig. S1.1B,C). Most of the reads classified as *L. crispatus* or *L.*

iners mapped to a single metagenomic assembled genome, with completeness of 95.7% and 97.1% and redundancy of 0% and 1.4% respectively.

Twelve samples from seven subjects produced ITS2 sequences (Fig. 1.2A); we do not have quantitative data characterizing the abundance of bacterial or fungal biomass. Eleven samples from six subjects contained species of *Candida*, classified as *C. albicans* (Fig. S1.2A), the most abundant fungal taxon in these data. Shotgun metagenomics confirmed these results, and allowed for additional identification of reads mapping to taxa such as *Malassezia* spp. (Fig. S1.2C). Subject 1088 was the only participant to deviate from this trend, with a high relative abundance of *Aspergillus* during the first trimester of pregnancy (Fig. S1.2A).

Alpha diversity indices based on 16S rRNA, and ITS2 data when available were compared across trimesters. While some subjects exhibited qualitative evidence of compositional shifts in vaginal microbiota with advancing gestation (Fig. 1.2A), we did not observe a significant difference in bacterial richness (number of observed OTUs) (LME, $p = 0.17$), evenness (Pielou's evenness index) (83) (LME, $p = 0.46$) or phylogenetic diversity (LME, $P=0.21$)] across trimesters (Fig. 1.2D-F).

Highly abundant bacterial taxa were significantly associated with community composition

An nMDS plot of Bray-Curtis dissimilarities showed that vaginal communities clustered by their most abundant bacterium (Fig. 1.3A). This association between the most abundant bacterial taxa and microbial composition of the sample was significant and explained more than half of the variance using PERMANOVA ($R^2= 56\%$, $p= 0.0001$). To account for repeated measures from longitudinal samples from the same individual we also performed an LME, which required dimensional reduction (LME, $R^2=69\%$, $p < 0.0001$). Communities with abundant *L. crispatus* were more similar to each other, sharing more than 90% similarity, in comparison to

communities where a different bacterial species was most abundant. While some individuals exhibited a relatively stable microbial community over time, others (6/16) experienced shifts in composition, resulting in a statistically significant change in Bray-Curtis dissimilarity on PC axis-1 between trimesters (Fig. 1.3B).

The five subjects with *Candida* detected in at least one of two longitudinally-paired samples displayed a significant increase in inter-sample Bray-Curtis dissimilarity in their bacterial profiles over that interval (e.g. the intervals between trimester 1-2, 2-3, or 1-3), suggesting the presence of *Candida* may be associated with greater shifts in bacterial composition than those who had no *Candida* detected (Fig. S1.2B).

Metabolites have strong associations with vaginal microbial community structures

Using GC-TOF MS, we detected 330 metabolites from urine, saliva, and CVF with 133 identified compounds. In the same samples, 1946 metabolites were also detected by LC-QTOF MS/MS (lipidomics, Table S1.1), with an additional 353 identified compounds. The CVF metabolome as assessed by both mass spectrometry methods did not significantly differ across trimesters (LME, GC-TOF MS: $p = 0.6378$, LC-MS/MS: $p = 0.3942$). This stability was even true for the subset of individuals who exhibited shifts in microbiota composition over trimesters (LME, GC-TOF MS: $p = 0.6594$, LC-MS/MS: $p = 0.2482$). CVF samples dominated by distinct bacteria exhibited significant differences in metabolic profiles (PERMANOVA, $R^2 = 12\%$, $p = 0.0195$). A constrained, distance-based ordination plot recapitulated 67% of the community variation observed in the vaginal microbiota (Fig. 1.4). Superimposed on the ordination plot are GC-TOF predictor metabolites, calculated using the DistLM program in Primer-e. Indole-3-lactate (ILA) accounted for 27% of variation observed in the vaginal microbiota data, and was found to be more abundant in vaginal microbiota with abundant *L. crispatus* (Fig. 1.4B).

Mannitol was also more abundant in samples dominated *L. crispatus* (Fig. 1.4C). In parallel we found that a pathway for mannitol-1-phosphate production is also more abundant in shotgun metagenomic datasets of CVF samples dominated by *L. crispatus* (Fig. S1.3). This linear model identified the top ten annotated GC-TOF metabolites that were associated with variation in the microbial community composition are shown in Figure 1.4a; these ten metabolites together might explain almost 57% of the total variation in the microbial community composition. A permutated random forest recapitulated what we found in the DistLM, identifying mannitol and indole-3-lactate as two of the top variables of importance, specifically for distinguishing microbiomes with high abundances of *L. crispatus* (Fig. S1.4A). To explore the ability to analyze the metabolome in high-throughput, the same sample sets were analyzed using Matrix Assisted Laser Desorption Ionization (MALDI) Imaging Mass Spectrometry (MSI) (Table S1.1). Detected ions by MALDI were compared to those identified by GC-MS and LC-MS, and found that ~55% of the metabolites identified had corresponding ions in the MALDI analysis (Table S1.1).

Metagenomics and functional potential of communities

Distinct functions were associated with each of the vaginal microbial community clusters (PERMANOVA, R^2= 70%, p= 0.0001, Fig. 1.5). LEfSe identified several pathways that differed between *L. crispatus* and *G. vaginalis*, in particular, an enrichment of ammonia assimilation genes in *G. vaginalis* (Fig. S1.4B). Genes involved in mannitol metabolism were enriched in communities where *L. crispatus* was highly abundant (Fig. S1.3C,4B). Searching the PATRIC database of all sequenced *L. crispatus* (64 genomes), *L. iners* (22 genomes), and *G. vaginalis* (127 genomes) strains revealed annotated genes for mannitol usage and transport for *L. crispatus*, but not for *L. iners* or *G. vaginalis.*

Mothers and infants have significantly different saliva and urine metabolomes

Maternal and infant saliva and urine metabolomes were assessed with both GC-TOF and LC-MS/MS (lipidomics) in order to study the relationship between maternal and infant metabolomic compositions during early life (see Fig. 1). PERMANOVA analysis of lipidomics data from saliva samples showed the largest difference between mothers and offspring (PERMANOVA, R^2= 69%, p = 0.001, Fig. S1.5A). A subset of 50 lipidomics metabolites with high mean abundance, 70% of which were unannotated, showed distinct profiles between mother and offspring salivary metabolomes (Fig. S1.6A). Likewise, GC-TOF salivary metabolomes were also significantly different between mother and offspring, but far less variation was explained (PERMANOVA, R^2= 12%, p = 0.0001, Fig. S1.5B). Some metabolites, such as lactulose, were much more abundant in infants and largely absent in mothers (Fig. S1.6B). Maternal metabolomics profiles (both GC-TOF and LC-MS/MS lipidomics) have a strong individual signature, while infants do not (see PERMANOVAs, Supp, Table 1.2). The infant metabolome for saliva and urine had little variance attributed to which subject donated the sample, but GC-TOF was able to detect a significant change between the infant urine metabolome at birth versus 6 and 12 months of age (PERMANOVA, R^2= 34%, p = 0.0007, Table S1.2). Moreover, from lipidomics data, the infant metabolome profile seemed to converge on mothers' metabolomes as they aged, though more samples would be needed to confirm this finding (Fig. S1.5C). For both saliva and urine, GC-TOF and lipidomics detected metabolites were more similar for mother-child pairs than for unrelated individuals (Fig. S1.7). Mantel tests to determine if inter-sample relationships were similar between chromatography methods (including both GC-TOF vs lipidomics) showed a strong correlation between saliva samples, and weaker correlations between urine and CVF (Table S1.3).

Discussion

Exposure to the microbiome in early life is critical for immune and physiological development (34–37), yet the factors that set this trajectory remain poorly understood. In this study, we followed the vaginal microbiome through the trimesters of pregnancy for 18 women, tracking changes in the bacterial communities with longitudinal samples, and capturing their functional potential with metagenomic sequencing and multiple platforms to assess metabolomic profiles. The resolution provided by shotgun metagenomic sequencing allowed us to identify species and characterize functional gene content of CVF microbiomes. An additional strength of this work is the strict inclusion criteria defining healthy pregnancy (see *Subject Information*, in Methods). Moreover, as part of an existing sample cohort, we had the opportunity to measure saliva and urine metabolomes from mothers and children. We aim to establish how the metabolome develops in the first year of life, and how maternal-infant saliva and urine metabolomes relate. In our study, most healthy pregnant women exhibited a relatively stable vaginal microbiota throughout the trimesters of pregnancy, dominated by *Lactobacillus* or, in some cases, more diverse, *Bifidobacterium*-dominated microbiota. However, a subset of women exhibited compositional shifts in their CVF microbiota as pregnancy progressed, as has also been seen in other larger cohorts (44) . We found several strong correlations between particular vaginal communities and metabolites, which may help us understand the physiology underlying distinct vaginal microbiota structures that were evident in our study. Lastly, vaginal microbiota composition predicted which metabolites were present in the CVF samples, but not urine or saliva samples from the mothers or infants, suggesting that local microbial metabolism may represent the dominant contributor to the metabolic milieu of the vaginal tract during pregnancy.

Our study supports the results from several other studies that have indicated that the vaginal microbiome is stable during pregnancy (39, 84). Specifically, in a longitudinal study

including 90 women, most retain a microbial community with the same dominant member (in *L. crispatus* communities, 75% remain stable, in communities with high *L. iners* abundance, 71% remain stable, and in more diverse communities like those sometimes associated with BV, 58% do not shift) (44). Using metagenomic sequencing to probe microbial community variation, our findings indicate that few bacteria, particularly *Lactobacillus* species, are highly abundant in the vaginal environment. Indeed, for individuals with vaginal microbiomes numerically dominated by *L. crispatus* or *L. iners*, the vast majority of reads mapped to contigs from one strain of *L. crispatus* or *L. iners* (Fig. S1.1b). Of note, the microbiome of some individuals did differ considerably with advancing pregnancy. For instance, the vaginal microbiota of subjects 1180 and 1222 had higher abundances of *L. crispatus* during the first trimester, but *L. iners* was more abundant in the remaining trimesters. Brooks et al. (85) demonstrated that shifts in vaginal microbiota structures can be described probabilistically, where shifts from *L. crispatus* to *L. iners* are the most likely to occur. This is consistent with the observations made in two individuals from our study, however, due to the small number of samples exhibiting this phenomenon qualitative assessments were more appropriate than statistical analysis. Of note, vaginal microbiota instability throughout pregnancy was associated with the presence of *Candida,* a known opportunistic pathogen of the vaginal tract. Since inclusion criteria for this study stipulated no antibiotic treatment, it is unlikely that *Candida* detection was a result of antibiotic administration. The prevalence of *Candida* in our cohort is more likely to be reflective of the fact that pregnancy is a known risk factor for candidiasis (86) and to related differences in the vaginal environment, including microbiological colonization. Indeed *L. crispatus* has been shown to have anti-*Candida* activity (87), and 90% (10/11) of samples that were *Candida* positive came

from individuals whose vaginal microbiota were dominated by an organism other than *L. crispatus*.

A few metabolites were highly indicative of the bacterial community present in each subject and may be useful biomarkers for the type of vaginal microbiota present. The most indicative metabolite was indole-3-lactate (ILA), a tryptophan metabolite whose abundance was correlated with communities having abundant *L. crispatus* (Fig. 1.4B and Fig. S1.4A). One potential explanation is that *L. crispatus* produces ILA to competitively exclude the growth of other species (Fig. 1.6). At physiologically relevant concentrations, ILA has been shown to have anti-microbial properties against both Gram-positive and Gram-negative organisms (88, 89). Although the production of lactic acid is generally thought of as a strategy *Lactobacillus* spp. use to prevent other species from colonizing the vagina, perhaps these organisms also use ILA in a similar or supplementary capacity. Additionally, bacterial derived ILA (also referred to as indole-lactic acid) has been recently shown to directly move from maternal to infant tissue (8). It has been suggested that indoles may play an important role as a ligand for the human aryl hydrocarbon receptor (AhR), which have diverse functions from immune regulation to metabolism (reviewed in (90)). Zelante et. al further showed that some lactobacilli produce the related tryptophan catabolite, indole-3-aldehyde (IAld), which provides protection against candidiasis by increasing IL-22 production via AhR receptor binding (91). The study also demonstrated that vaginal specific bacteria, such as *L. acidophilus*, produce IAld in the vaginal environment, which protected against vaginal but not intestinal candidiasis. We measured indole-3-acetate (IAA), the direct precursor to IAld, in our study, but found no difference in its abundance between women dominated by different species of *Lactobacillus* (data not shown). Because indole-3-lactate can act as a ligand for AhR, we speculate that *L. crispatus* may regulate

the IL-22-AhR response in the vagina, reducing the risk of vaginal candidiasis in the same way IAld does, and potentially activating the AhR response in newborns to prevent early life candidiasis. Additionally, indole itself may play a role in structuring community composition by selecting for organisms that have adapted to a high abundance of this metabolite, repelling more transient microbes that have not been exposed to higher indole concentrations previously (92).

Whole genome shotgun metagenomics allowed us to begin to address the functional capacities of these microbiomes. The largest differences between functional capacity appeared to be between communities where *L. crispatus* or *G. vaginalis* were the most abundant bacterial taxon. One pathway particularly enriched in *Gardnerella* communities was the ammonia assimilation cycle (Fig. S1.4B). Studies have pointed out *Gardnerella's* preference for ammonia as a nitrogen source (93); moreover, this ability to assimilate ammonia and produce amino acids has been implicated in mutualistic interactions between species of *Prevotella*, in the context of BV (94). Together, these BV-associated organisms contribute to genital inflammation, which may play a role in the susceptibility of certain diseases, such as HIV (95).

Increased abundance of mannitol when *L. crispatus* was present is an important and unexpected finding (Fig. 1.4). Most likely, mannitol contributes to optimizing the tonicity of the vaginal environment, and has recently been considered for this use in developing effective therapeutics for altering the vaginal microbiota (96, 97). Even more, mannitol may assist *L. crispatus* in adhering to the epithelial layer, a strategy the organism may use to competitively inhibit other microbes from colonizing, potentially by drawing out excess water in the mucin layer and altering the mucin structure (98). Irrespective of the biochemistry, these genes, and mannitol in general, represent very specific markers of a community where *L. crispatus* was most abundant.

Although it is known that homofermentative lactic acid bacteria (LAB) such as *L. crispatus* (99) convert glucose primarily to lactic acid, it is unclear why mannitol accumulates in this niche. Interestingly, there was no difference in the glucose abundance between the four distinct vaginal communities. Further, metabolomic analysis of our CVF samples failed to capture significant levels of the mannitol precursor fructose; however, previous studies have indicated an appreciable amount of fructose within the cervical mucus of humans (100) and the capability of *L. crispatus* to utilize fructose as a carbon source (101, 102). We speculate that this high extracellular mannitol abundance phenotype may underlie the cell's need to regenerate NAD+ for use in glycolysis. When faced with a limiting amount of pyruvate (or perhaps an upstream glycolytic metabolite) to convert to lactate, homofermentative LAB may be unable to produce sufficient NAD+ to allow glycolysis to continue. To this end, reducing fructose-6-phosphate to mannitol-1-phosphate may be an alternative and vital way *L. crispatus* regenerates NAD+ for glycolysis (103, 104). We did find the gene mannitol-1-phosphate dehydrogenase, responsible for converting fructose-6-phosphate to mannitol-1-phosphate, was highly correlated ($R^2 = 0.9$) with the relative abundance of *L. crispatus*. The genes for the conversion of mannitol-1-phosephate to mannitol (presumably a M1P phosphatase), and its subsequent export are currently unknown (104). This may imply that mannitol accumulation is a marker of a cellular switch to NAD+ regeneration by fructose reduction rather than converting pyruvate to lactic acid. Consequently, the decrease in lactic acid production may contribute to community dysbiosis due to a rise in pH (Fig. 1.6). Future experiments using culturing to elucidate whether these *in vivo* community data are recapitulated with axenic cultures *in vitro* are needed.

Furthermore, this study enabled comparison of two metabolomic methods (GC-TOF and LC-MS/MS lipidomics) for analysis of the pregnancy and early life metabolomes. Our data

showed that metabolite intensities obtained by GC-TOF were more tightly correlated with microbial community composition than those obtained by lipidomics, perhaps indicating that GC-TOF is more effective at detecting microbial metabolites than lipidomics, especially during pregnancy. We also show that both the saliva and urine metabolomes of children are more similar to their own mother than to unrelated individuals (Fig. 1.6). Strikingly, the ability of lipidomics to differentiate mothers from children via saliva was the strongest signal in our metabolomics data (Table S1.1). The oral microbiome may play a role in this, as there is a well-established community succession in children during early life (reviewed in (105)), where children begin life with oral microbes that differ from those in adults. Lactulose, detected by GC-TOF, was a specific metabolite with increased abundance in infant saliva, which may reflect its use as a treatment for constipation (106) or perhaps even its presence in heated milk (107). Finally, urine metabolomes had a distinct age profile, especially with the lipidomics data. Our data suggests that, over the first year of life, the urine metabolome rapidly converges on the adult metabolome. This is likely a result of the development of renal system in children (108), along with the development of the gut microbiome and the related metabolites which are processed through the liver and kidneys. Other reasons for this age-related shift include a change in diet and a weaning off of breastmilk or formula (108). Expectedly, we did not see a strong influence of the vaginal microbiome during pregnancy on the infant saliva and urine metabolome. We suspect that if differences in the vaginal microbiome were to affect the early life saliva and urine metabolome, those effects would be subtle. The lack of stool samples from the mothers and infants is a limitation of this study, as they may contain a stronger signal of shared metabolomes across mother-infant dyads. Additionally, this study explored an approach to characterize the metabolomes in high-throughput. By using acoustic deposition, in combination with MALDI-

MSI, a throughput of ~1 seconds per sample was reached using only 2 microliter of sample. Of the metabolites identified by GC-MS and LC-MS, ~55% had corresponding ions in the MALDI-MSI analysis (Table S1.1). Future work will focus on confirming these metabolite identifications, but the initial results are promising and indicate that rapid analysis of microbial metabolites using MALDI, an analysis platform routinely used in clinical microbiology laboratories, is feasible (109).

Overall, we provide a broad look at the metabolome during pregnancy and early life, detailing the utility of GC-TOF, lipidomics and MALDI-MSI for saliva, urine, and CVF.

Conclusion

Here we share a high-resolution characterization of the vaginal microbiome, longitudinally sampled throughout healthy pregnancy. We show that, despite the generally accepted view that lactobacilli are indicative of healthy vaginal communities, many women in our healthy pregnancy cohort had non-*Lactobacillus* dominated communities. The vaginal communities were characterized by a high abundance of one or a few acid-tolerant species, which dictated the physiologic potential and the metabolic profiles of the vaginal microbiome. Many of the metabolites that were specific to these different organisms warrant further investigation, especially considering the recent development of VMT as a treatment for BV (59). The metabolites we found to be associated with *L. crispatus* may be useful as microbiome cultivation approaches are developed to intentionally direct the composition of the vaginal microbiome. For example, indole-3-lactate may support *L. crispatus* colonization, while mannitol may indicate a shift in metabolism away from fermentation and the production of acid, relaxing the low-pH selection pressure which normally gives *L. crispatus* an advantage.

Acknowledgements

The pilot grant awarded to start a UC Center for Pediatric Microbiome Research awarded by the UCI Institute for Clinical and Translational Science (UL1 TR001414) gave rise to this project. A.O. was supported by a NIH T32 training grant (1T32AI14134601A1) from UC Irvine's Training Program in Microbiology and Infectious Diseases. We would like to thank Prof. Dan Cooper for enthusiasm and strategic advice. We would like to thank the members of the Whiteson Lab, especially Dr. Whitney England for insight into metagenomics analysis, Joann Phan and Tara Gallagher, for thoughtful comments during the preparation of this manuscript. Dr. Heather Maughan made insightful and clarifying edits that we are grateful for.

Data Availability

Sequence data for 16S, ITS2, and shotgun metagenomes were deposited on the National Center for Biotechnology Information (NCBI) sequence read archive (SRA) under the BioProject PRJNA612083. Metabolomics data for all samples can be found in Supplementary Table 1.1. R-scripts for statistical analysis are published on GitHub: https://github.com/aoliver44/Cervicovaginal-Paper.

Ethics approval and consent to participate

This study utilized a subset of samples from a larger, longitudinal prospective cohort study designed to analyze the relationship between maternal stress and infant development, conducted at the University of California, Irvine (UCI) (61). The University of California's Institutional Review Board approved the protocol and written, informed consent was obtained from each participant. Research on human subjects was performed in accordance with the Declaration of Helsinki.

Table 1.1: Demographics of the eighteen mothers who participated in the study.

Maternal ID	Race-Ethnicity	Age	BMI (pre-pregnancy)
1018	White Hispanic	35	27.4
1062	White Hispanic	23	25.3
1088	White non-Hispanic	26	25.8
1089	White non-Hispanic	22	21.8
1103	White non-Hispanic	27	24.5
1111	White Hispanic	38	27.9
1120	White non-Hispanic	34	26.9
1126	White Hispanic	19	27.8
1137	White Hispanic	31	23.5
1146	White non-Hispanic	29	23.5
1151	White non-Hispanic	30	22.4
1157	White non-Hispanic	28	18.9
1180	White non-Hispanic	29	24.9
1191	White Hispanic	31	22.7
1198	White non-Hispanic	26	24.8
1201	White non-Hispanic	30	29.9
1202	White Hispanic	20	24.7
1222	White non-Hispanic	24	24.0

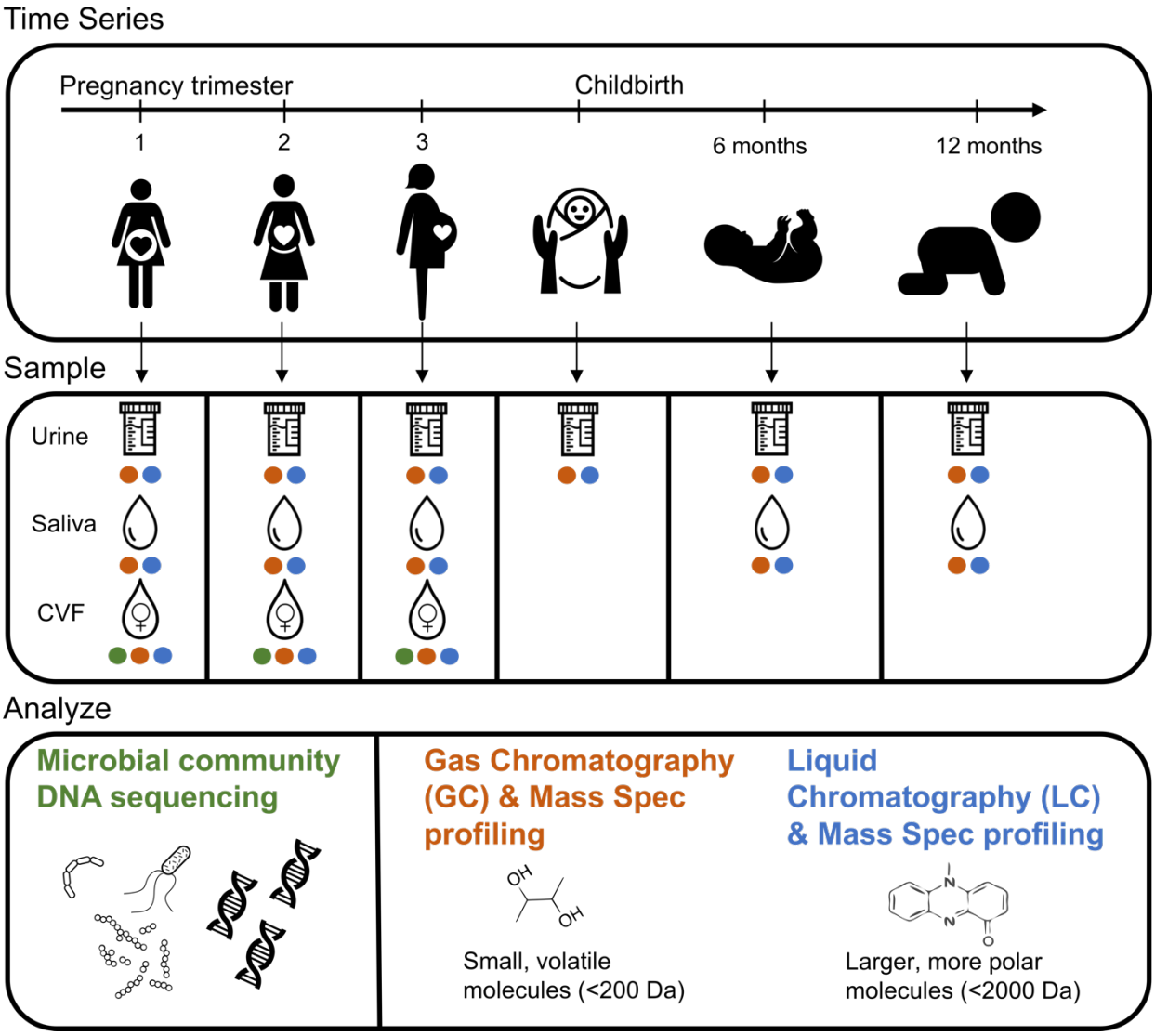
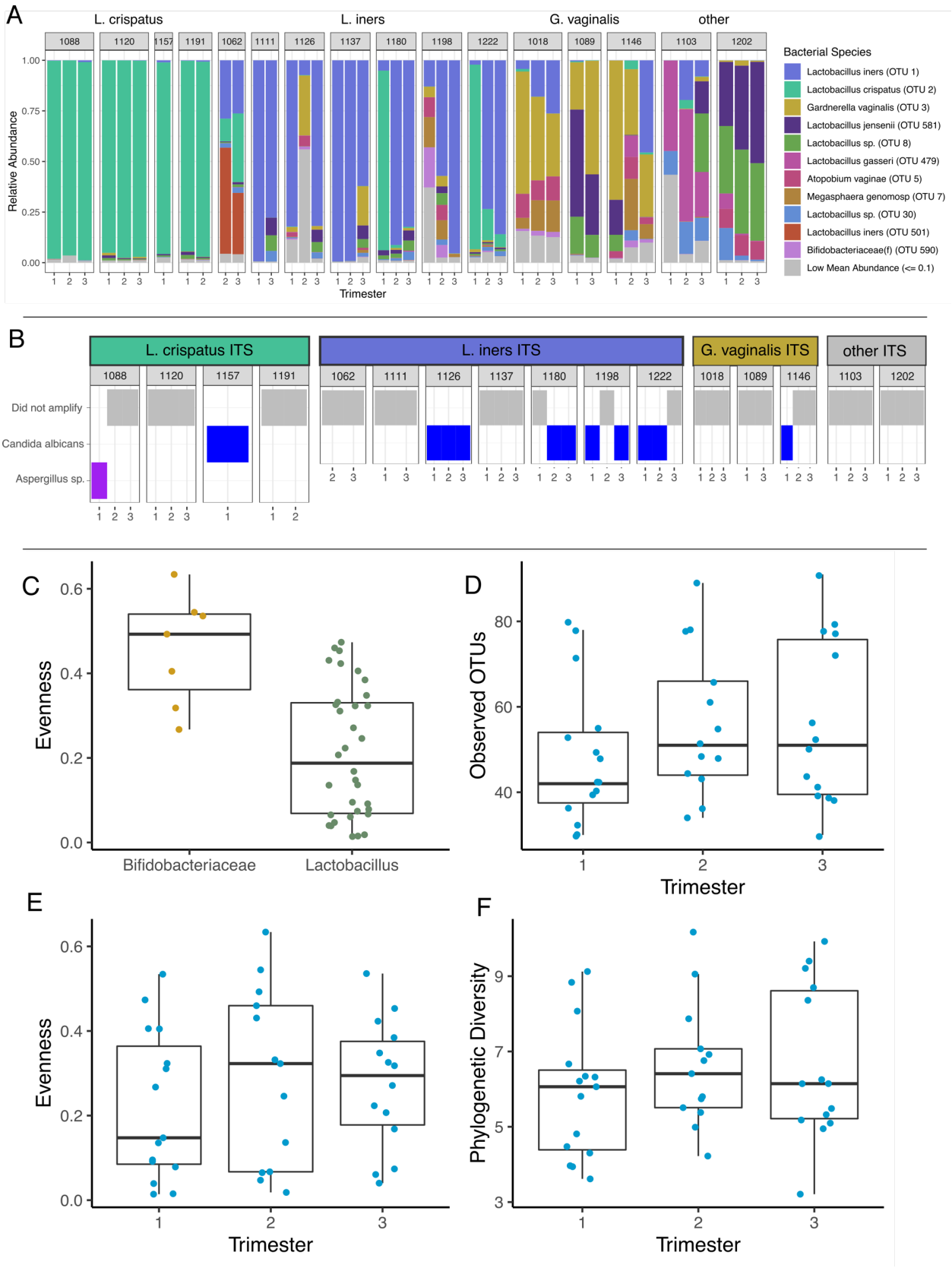


Figure 1.1: Study outline. Eighteen women were sampled throughout pregnancy and their offspring were sampled at birth, six, and 12 months of age. Samples collected were urine, saliva, and cervical vaginal fluid (CVF) for the mothers, and urine and saliva for the children. CVF was sequenced using shotgun metagenomics and amplicon sequencing. All samples were analyzed using GC-TOF and lipidomics.

Figure 1.2: Taxonomy and alpha diversity of vaginal microbiomes during pregnancy. A) Relative abundance plot of operational taxonomic units, from 16S amplicon data, grouped together by individual. Each individual is clustered into a larger category defined by the dominating microbe. B) Presence or absence of fungi, at the genus-level, per sample. Linear mixed-effects models (LME) were done on the alpha diversity metrics to account for repeated measures in the data C) Evenness between samples dominated by *Lactobacillus* (n = 34 samples) is significantly lower than samples dominated by *Bifidobacteriaceae* (n = 7 samples). D) No significant change in the observed OTUs between the trimesters (n = 15, 13, 14 samples respectively) of pregnancy and likewise E) there was not change in evenness or F) phylogenetic diversity throughout pregnancy.


A
L. crispatus
L. iners
G. vaginalis
other
1088
1120
1157
1191
1062
1111
1126
1137
1180
1198
1222
1018
1089
1146
1103
1202
Relative Abundance
1.00
0.75
0.50
0.25
0.00
Trimester
Bacterial Species
Lactobacillus iners (OTU 1)
Lactobacillus crispatus (OTU 2)
Gardnerella vaginalis (OTU 3)
Lactobacillus jensenii (OTU 581)
Lactobacillus sp. (OTU 8)
Lactobacillus gasseri (OTU 479)
Atopobium vaginae (OTU 5)
Megasphaera genomosp (OTU 7)
Lactobacillus sp. (OTU 30)
Lactobacillus iners (OTU 501)
Bifidobacteriaceae(f) (OTU 590)
Low Mean Abundance (<= 0.1)
B
L. crispatus ITS
L. iners ITS
G. vaginalis ITS
other ITS
Did not amplify
Candida albicans
Aspergillus sp.
C
Evenness
Bifidobacteriaceae
Lactobacillus
D
Observed OTUs
Trimester
E
Evenness
Trimester
F
Phylogenetic Diversity
Trimester

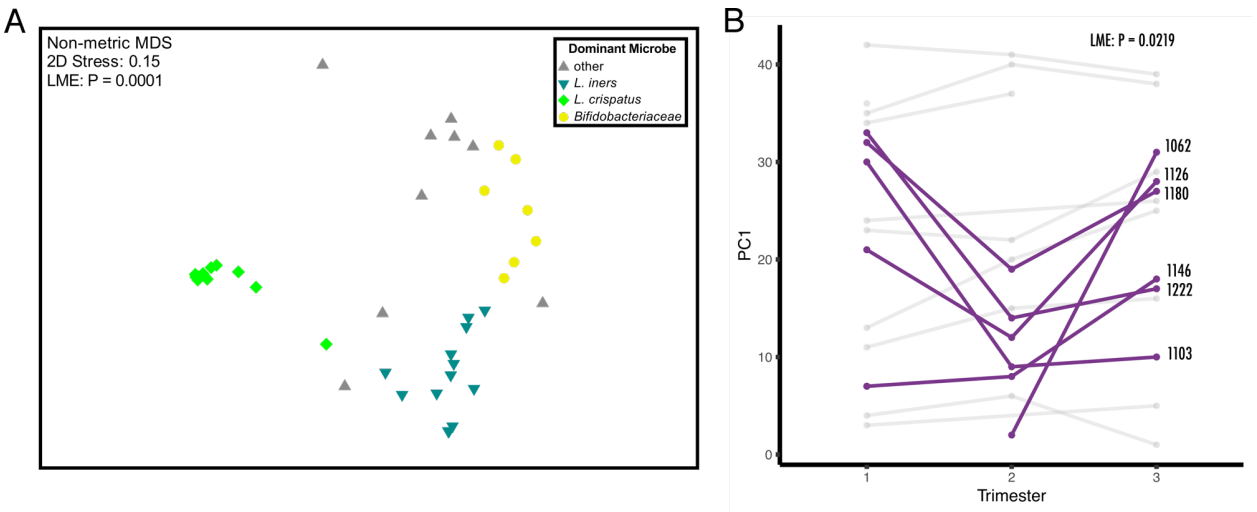


Figure 1.3: Ordination of vaginal microbiomes during pregnancy. A) Non-metric multidimensional scaling (nMDS) of Bray Curtis dissimilarity between vaginal microbiomes (n = 42 samples) of mothers. Color indicates the most abundant microbe within the microbial community. The most abundant microbe in the community plays a statistically significant role in the composition of the community (LME, R^2=69%, $p < 0.0001$) B) Some participants (6/16 individuals) experienced large, significant (LME: P = 0.0219) shifts in their microbiomes throughout the trimesters of pregnancy.

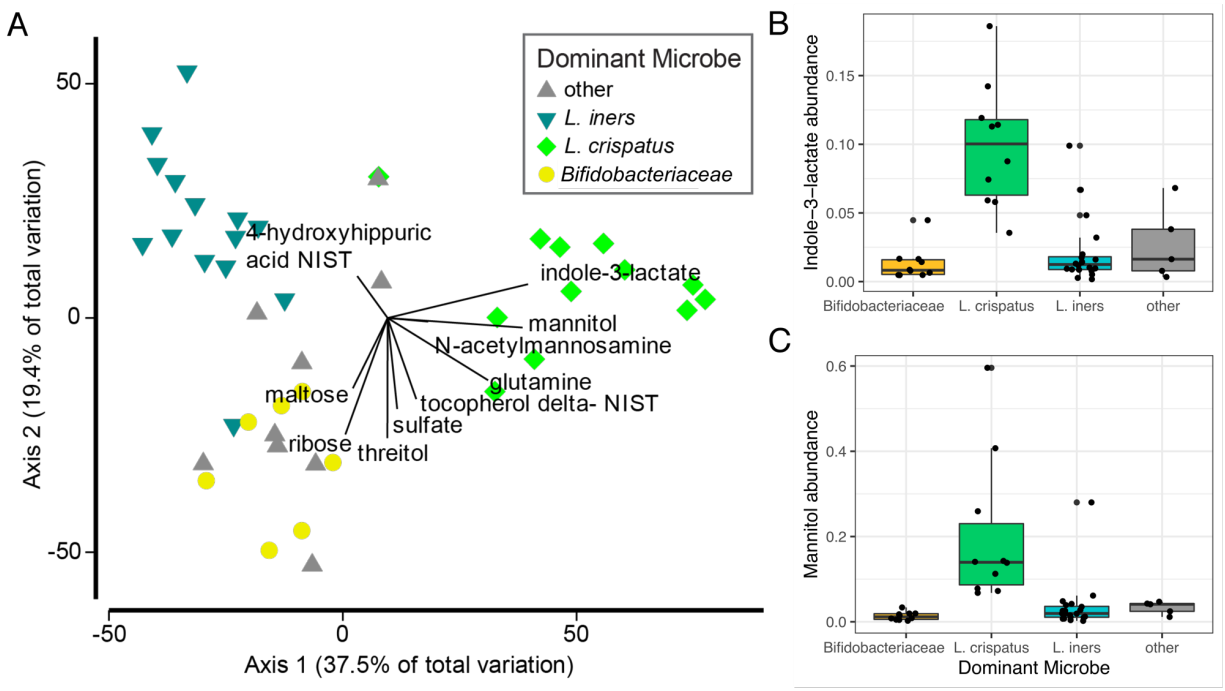


Figure 1.4: Relationship between vaginal microbes and metabolites. A) Distance-based linear model recapitulates the relationship between the vaginal microbiomes of these subjects (n = 42 samples). Superimposed are vectors showing which annotated GC-TOF molecules are best correlated with these microbial communities. Length and direction of vectors correspond to the strength of the association between the metabolite and the microbial communities. Boxplots show the raw abundance (n = 45 samples) of B) indole-3-lactate and C) mannitol.

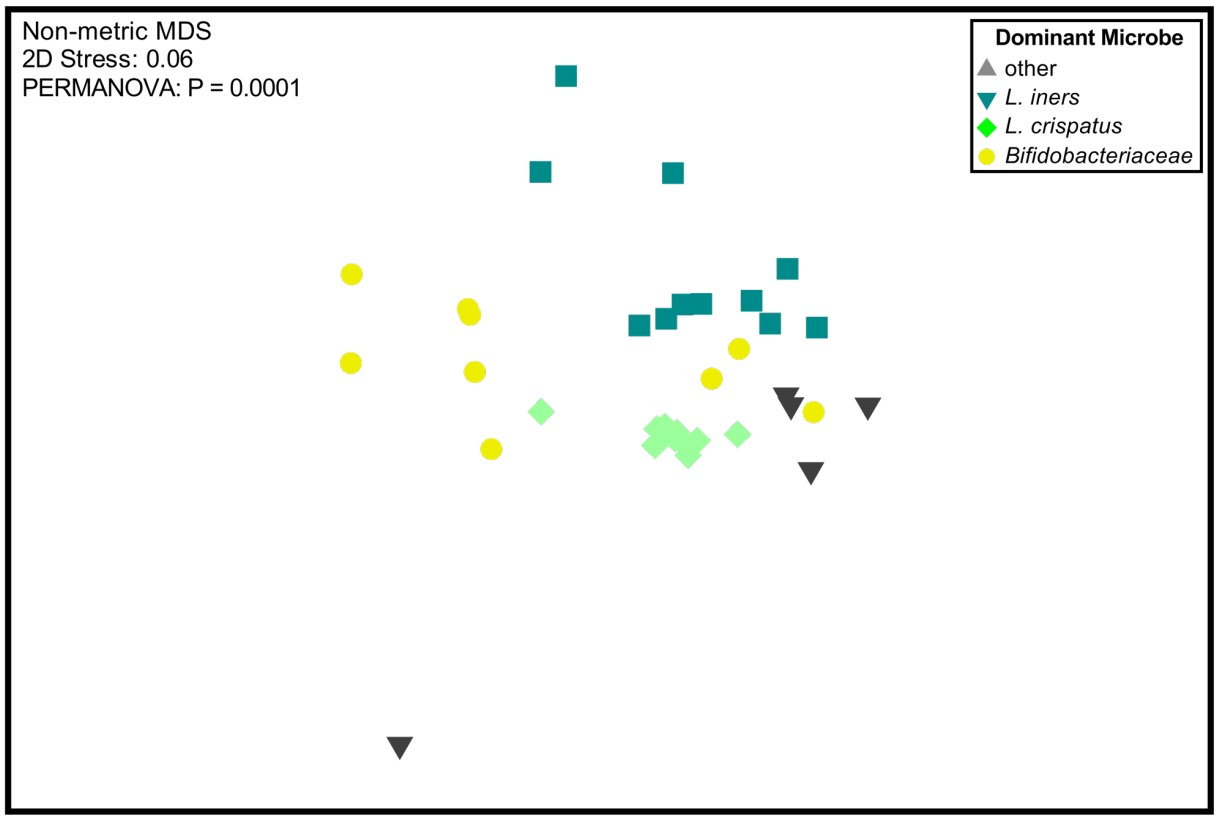


Figure 1.5: Ordination of functional pathways within the vaginal microbiome. An nMDS of HUMAnN2 analysis, examining the abundance of pathways in each microbiome (n = 35 samples). Vaginal microbiomes have functions that are indicative of the most abundant microbe present in the samples.

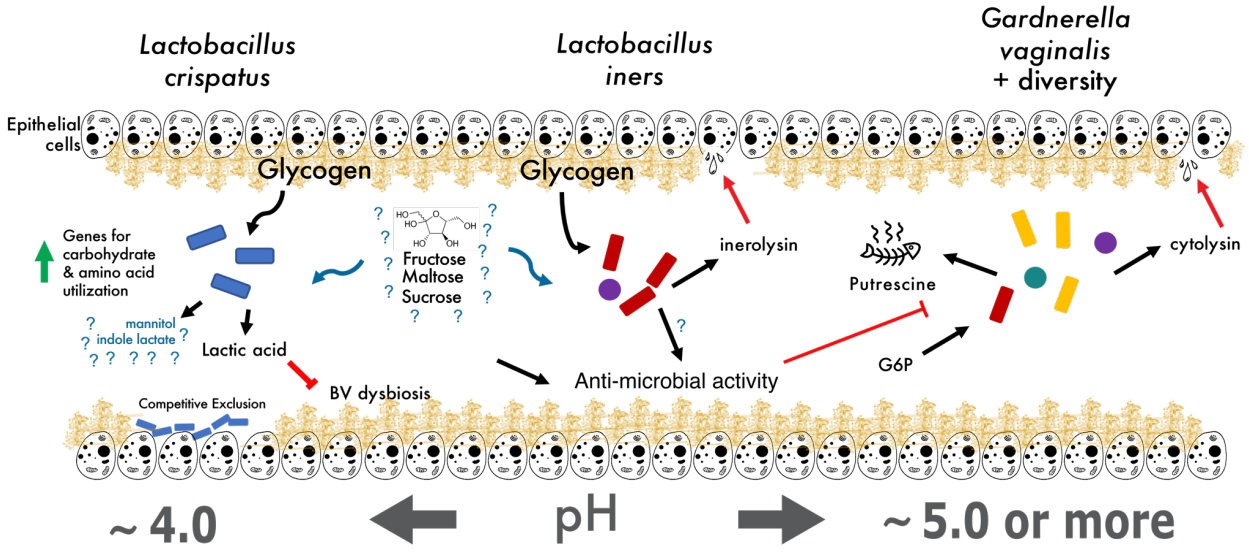


Figure 1.6: Proposed vaginal microbial community model. Current hypothesized model of vaginal microbial community physiology, with gaps in understanding (denoted by question marks) where future work is needed. Our study indicates that mannitol production is associated with a high relative abundance of *L. crispatus*.

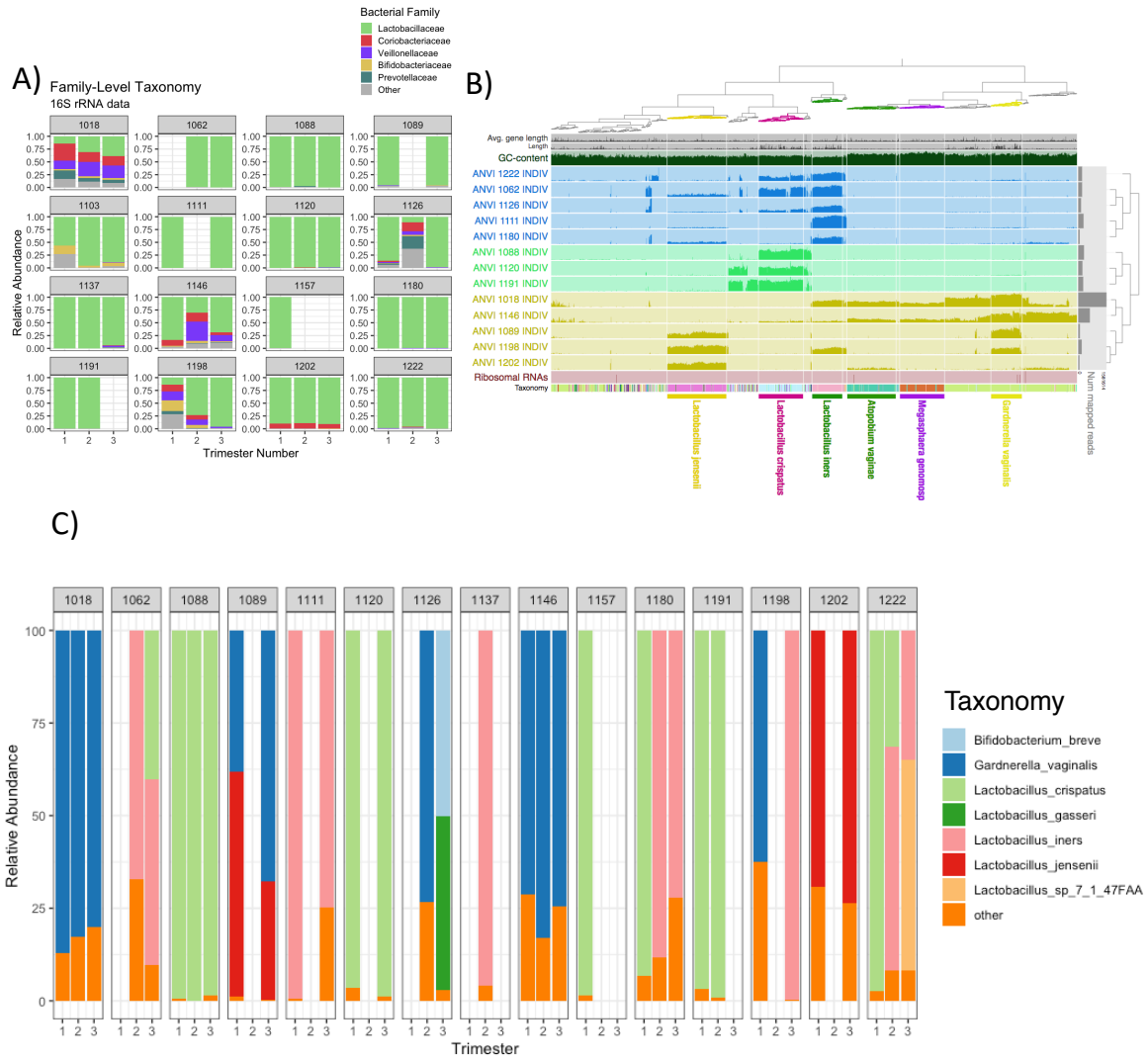
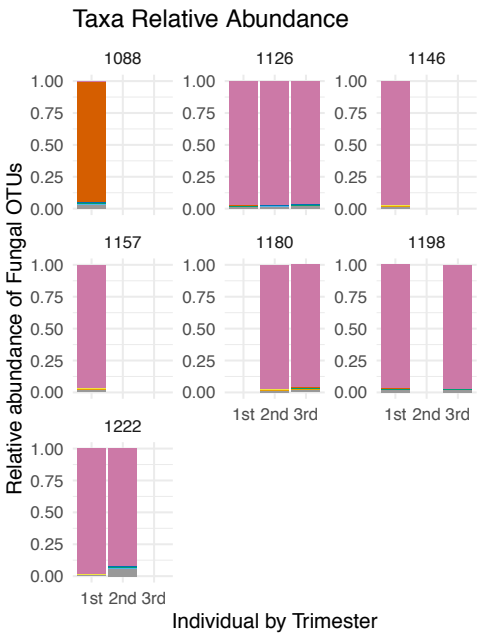


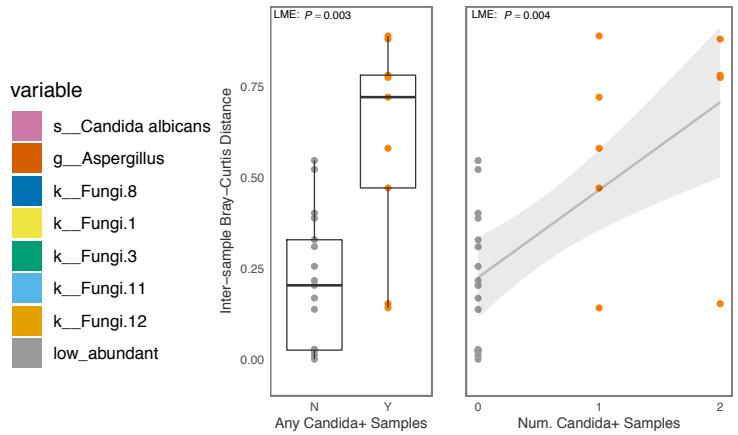
Fig. S1.1: A) Stacked bar plots of 16S rRNA gene taxonomy, collapsed down to the family level. ”Other” category contains taxa at a mean relative abundance of 1% or less. B) Anvi’o plot illustrating the abundance of contigs (>= 2,000bp) for each subject (trimesters combined). The participants are colored according to the most abundant microbe present (*L. iners* (blue), *L. crispatus* (green), *G. vaginalis* (yellow)). Short read taxonomic assigner Kaiju was used to assign taxonomy to the contigs. C) Species level taxonomy of 35 shotgun metagenomic samples, using the MiDAS taxonomic classifier.

Fig. S1.2: A) Bar plot showing the relative abundance of fungal OTUs within subjects. The most abundant OTUs by far were all classified as *Candida albicans* (pink). All other OTUs made up very little of the fungal presence in these samples, save for subject 1088, who had a species of *Aspergillus* as its most abundant fungal taxon. B) Sample where *Candida* was detected had greater shifts in bacterial composition than samples without. C) Shotgun metagenomics was able to pick up reads aligning to database containing human-associated fungal species and confirmed ITS results and classified additional taxa.

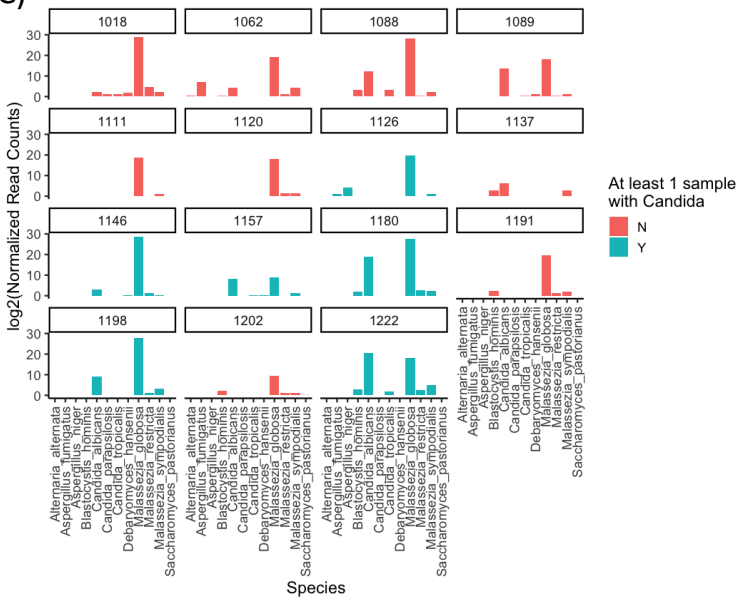
A)


Taxa Relative Abundance
1088
1126
1146
1157
1180
1198
1222
1.00
0.75
0.50
0.25
0.00
1st 2nd 3rd
Relative abundance of Fungal OTUs
Individual by Trimester

B)


variable
s__Candida albicans
g__Aspergillus
k__Fungi.8
k__Fungi.1
k__Fungi.3
k__Fungi.11
k__Fungi.12
low_abundant
LME: P = 0.003
Inter−sample Bray−Curtis Distance
0.75
0.50
0.25
0.00
N
Y
Any Candida+ Samples
LME: P = 0.004
0
1
2
Num. Candida+ Samples

C)


1018
1062
1088
1089
1111
1120
1126
1137
1146
1157
1180
1191
1198
1202
1222
log2(Normalized Read Counts)
30
20
10
0
Alternaria_alternata
Aspergillus_fumigatus
Aspergillus_niger
Blastocystis_hominis
Candida_albicans
Candida_parapsilosis
Candida_tropicalis
Debaryomyces_hansenii
Malassezia_globosa
Malassezia_restricta
Malassezia_sympodialis
Saccharomyces_pastorianus
Species
At least 1 sample with Candida
N
Y

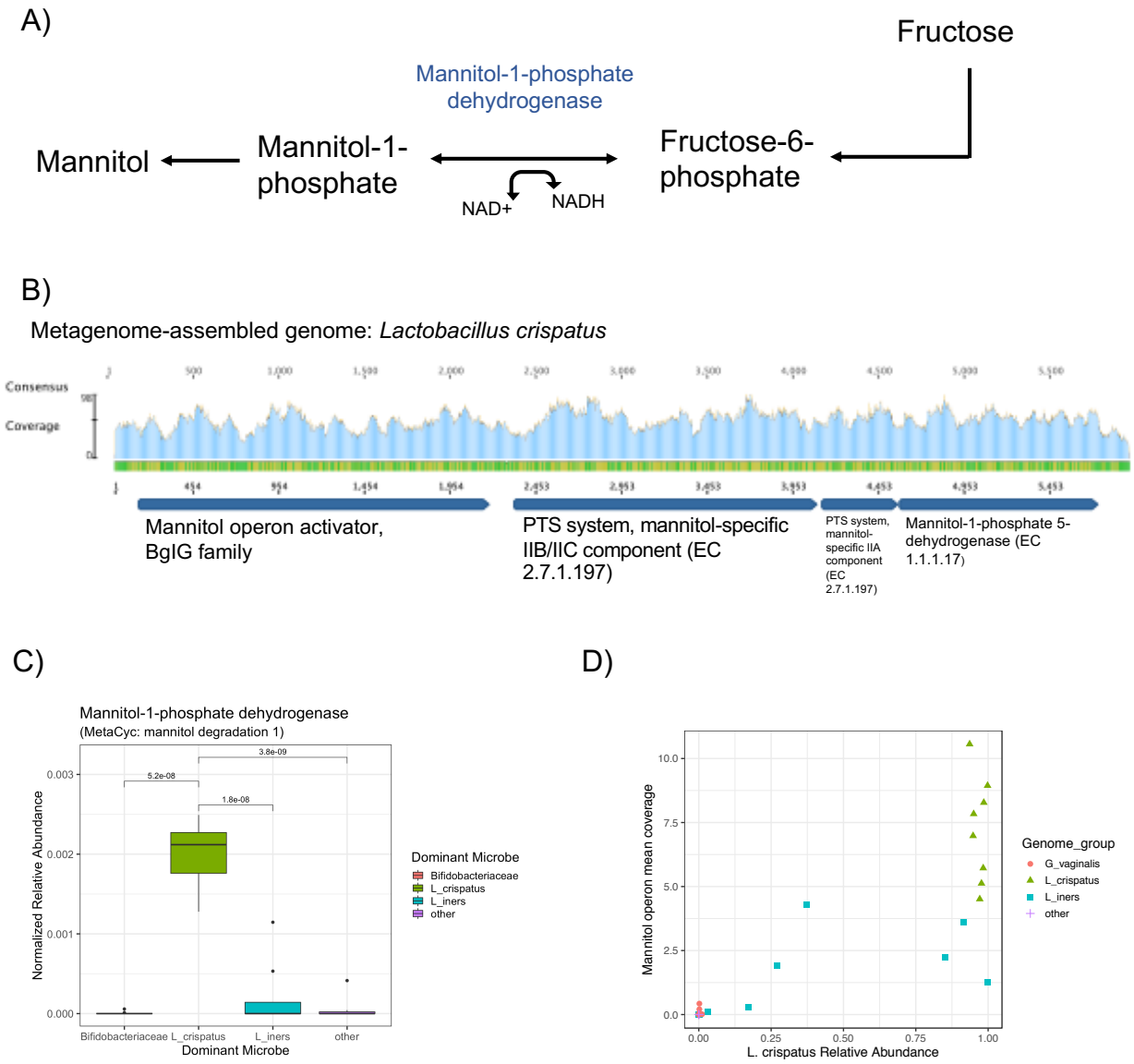


Fig. S1.3: A) Schematic of mannitol production. B) Coverage the mannitol operon from the metagenomes of individuals that had high abundances of *L. crispatus* in their samples. The operon was assembled using taxonomically assigned contigs in Anvio v5 and annotated using Patric. C) Analysis of the HUMAnN2 output, specifically the mannitol degradation pathway. Pairwise T-tests of *L. crispatus* against other dominant microbes. D) A Relationship between mean coverage of that operon with relative abundance of *L. crispatus*.

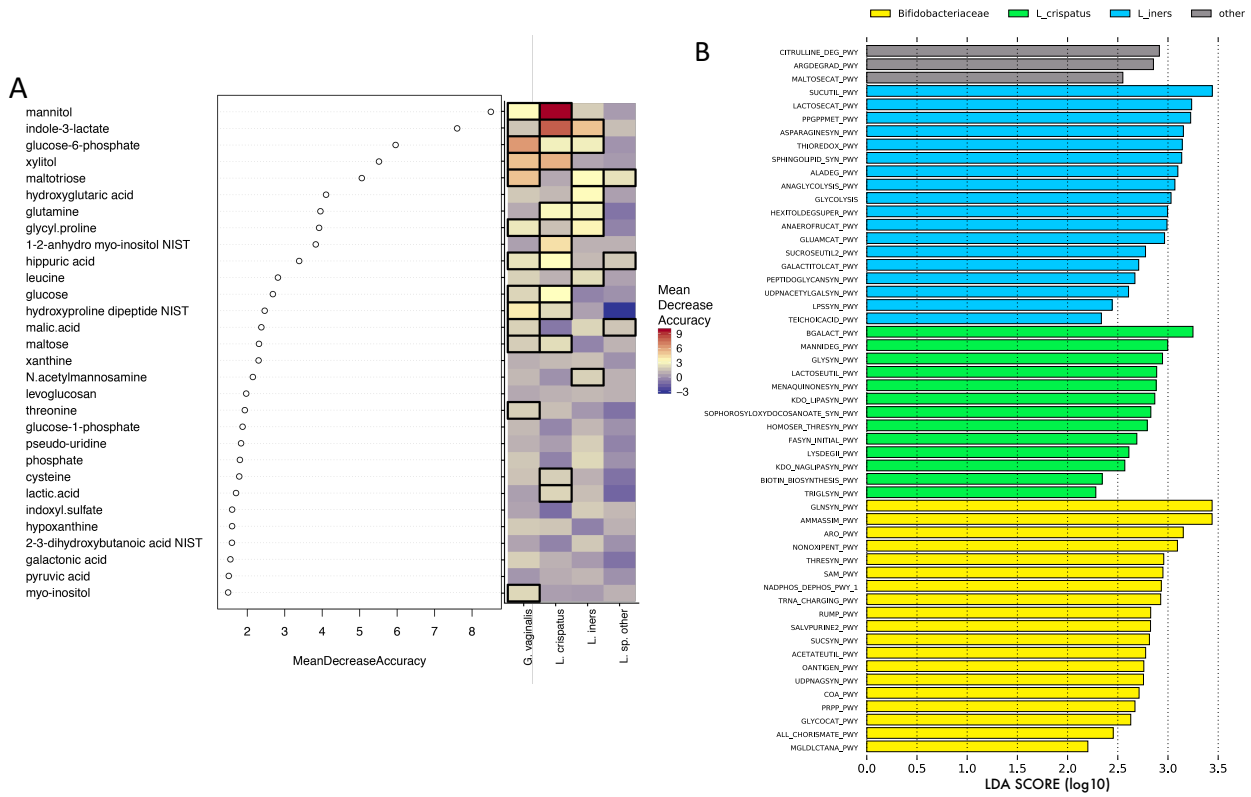


Fig. S1.4: A) Permutated random forest recapitulates the metabolites that drive differences between the vaginal communities. The heatmap shows the mean decrease in accuracy associated with the specific microbiomes. Bold boxes around the heatmap cell indicates statistical significance of that feature at $p < 0.05$. B) Lefse analysis of annotated HUMAnN2 output comparing the enrichment of functional pathways between vaginal microbiomes with abundant *Bifidobacteriaceae (G. vaginalis), L. crispatus, L. iners,* or other. This analysis was done using the -no_stratify flag for humann2, which analyzes the data without the taxa specific information.

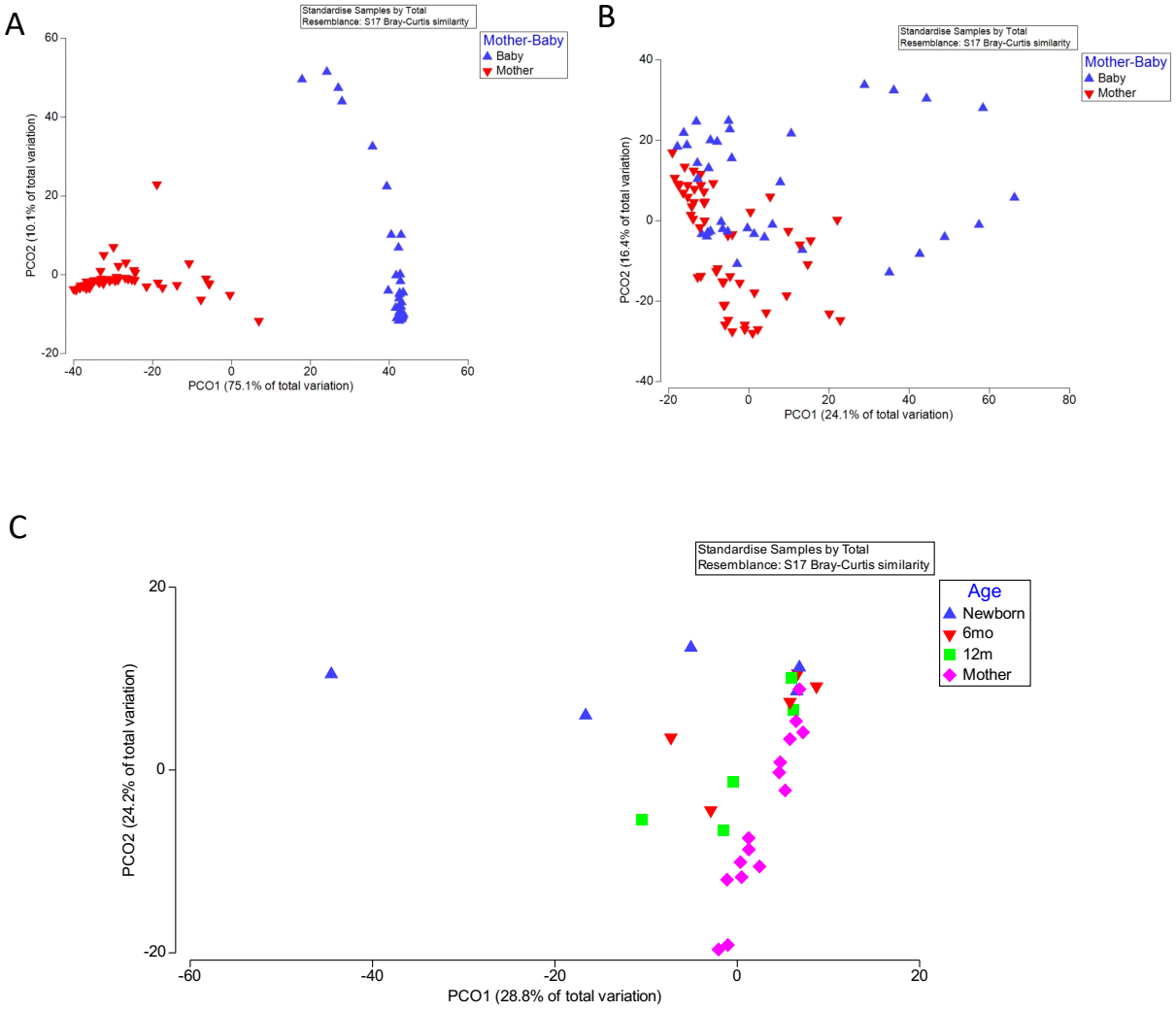


Fig. S1.5: A) Principle coordinates ordination of saliva LC-MS/MS lipidomics, colored by whether the sample originated from mother or child. B) Principle coordinates ordination of saliva metabolomes by GC-TOF, colored by whether the sample originated from mother or child. C) Principle coordinates ordination of urine lipidomes, colored by age. All timepoints for mothers (n=15 samples) were used (i.e. Trimesters 1-3).

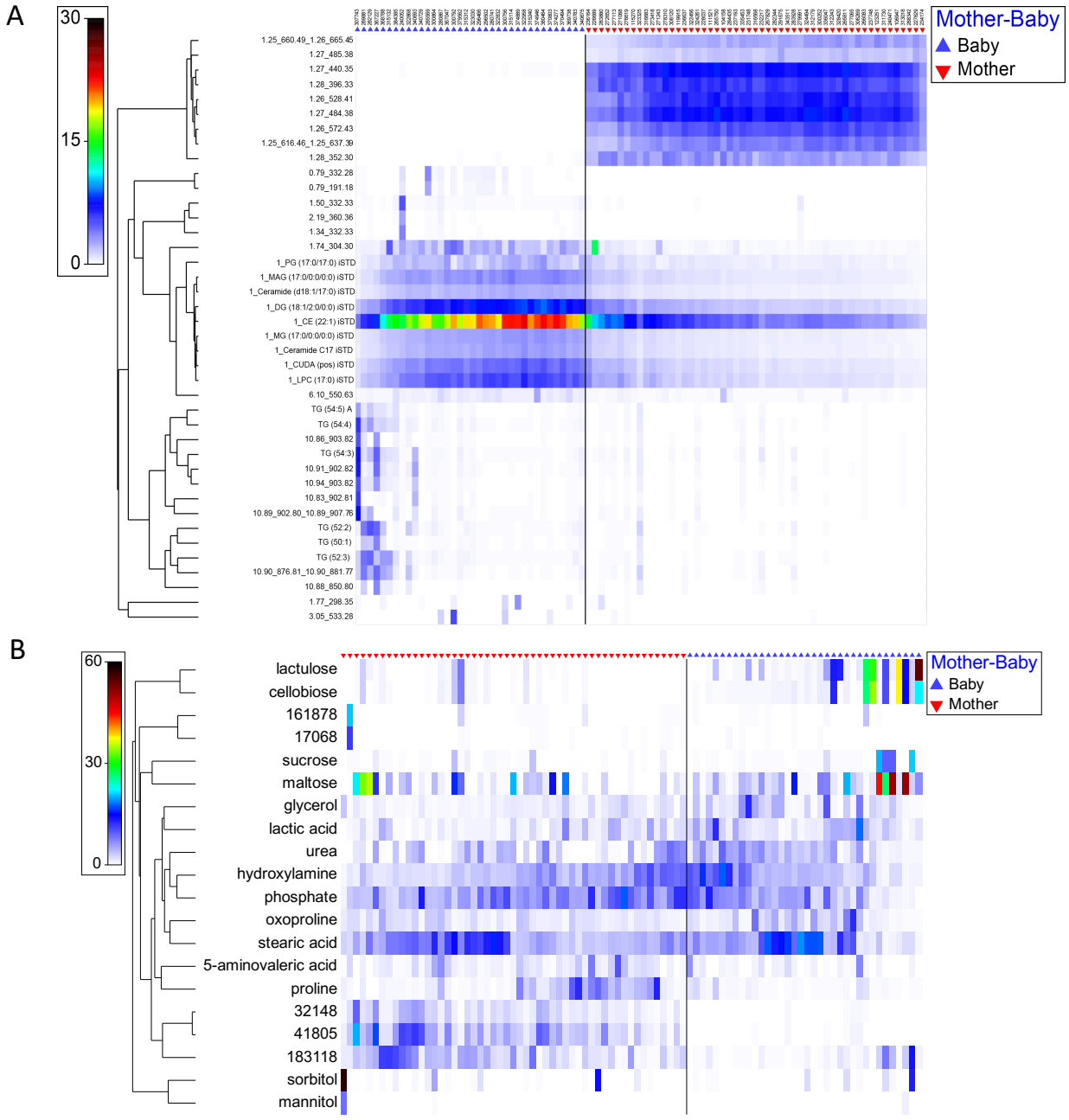


Fig. S1.6: A) Analysis of the top 50 lipidomic metabolites and B) 20 top GC-TOF metabolites from saliva that had the highest mean abundance across all samples using Primer-e software. Metabolites were initially standardized within each sample, and then across metabolites. Color indicates this normalized abundance of the metabolite. When present, annotations were used for data.

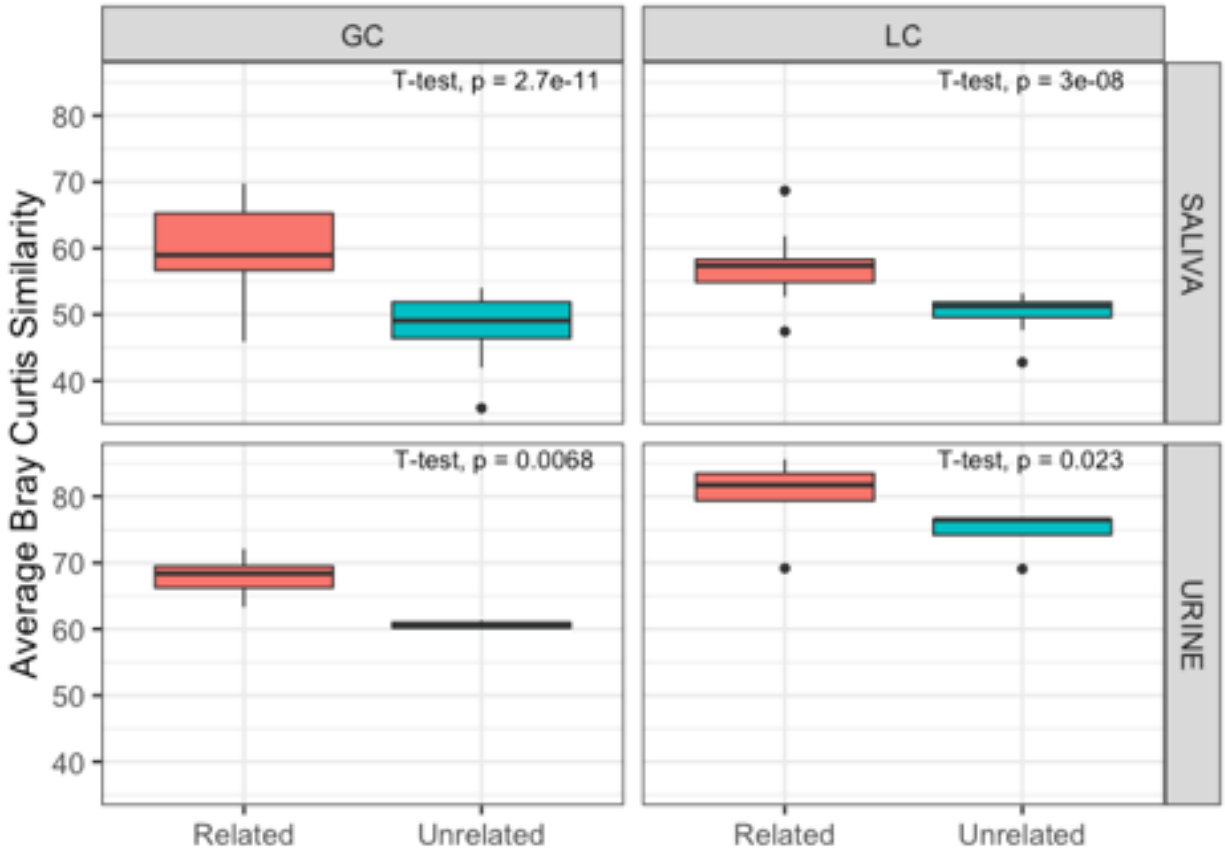


Fig. S1.7: Similarity of urine and saliva metabolomes between related and unrelated individuals. Related mothers and children have significantly more similar saliva and urine metabolomes than unrelated individuals. Graph shows average bray Curtis similarity between related and unrelated individuals for GC-TOF and lipidome metabolites. Paired T-tests were done to calculate significance. 53 mother samples and 36 infant saliva samples and 15 samples from both mother and infant urine samples were used in this analysis.

CHAPTER 2:

High fiber, whole foods dietary intervention alters the human gut microbiome but not fecal short-chain fatty acids

Authors: Andrew Oliver, Alexander B. Chase Ph.D., Claudia Weihe, Stephanie B. Orchanian, Stefan F. Riedel, Clark Hendrickson, Mi Lay, Julia Massimelli Sewall Ph.D., Jennifer B. H. Martiny , Katrine Whiteson Ph.D.

DOI: *10.1128/mSystems.00115-21*

ABSTRACT

Dietary shifts can have a direct impact on the gut microbiome by preferentially selecting for microbes capable of utilizing the various dietary nutrients. Intake of dietary fiber has decreased precipitously in the last century, while consumption of processed foods has increased. Fiber, or microbiota-accessible carbohydrates (MACs), persist in the digestive tract and can be metabolized by specific bacteria encoding fiber degrading enzymes. Digestion of MACs results in the accumulation of short-chain fatty acids (SCFAs) and other metabolic byproducts that are critical to human health. Here, we implemented a two-week dietary fiber intervention aiming for 40-50 grams of fiber per day within the context of a course-based undergraduate research experience (CURE) (n = 20). By coupling shotgun metagenomic sequencing and targeted gas-chromatography mass spectrometry (GC/MS), we found that the dietary intervention significantly altered the composition of individual gut microbiomes, accounting for 8.3% of the longitudinal variability within subjects. Notably, microbial taxa that increased in relative abundance as a result of the diet change included known MAC degraders (i.e., *Bifidobacterium* and *Lactobacillus*). We further assessed the genetic diversity within *Bifidobacterium*, assayed by amplification of the *groEL* gene. Concomitant with microbial composition changes, we show an

increase in the abundance of genes involved in inositol degradation. Despite these changes in gut microbiome composition, we did not detect a consistent shift in SCFA abundance. Collectively, our results demonstrate that on a short-term timescale of two weeks, increased fiber intake can induce compositional changes of the gut microbiome, including an increase in MAC degrading bacteria.

IMPORTANCE

A profound decrease in the consumption of dietary fiber in many parts of the world in the last century may be associated with the increasing prevalence of Type II diabetes, colon cancer, and other health problems. A typical U.S. diet includes about ~15 grams of fiber per day, far less fiber than daily recommended allowance. Changes in dietary fiber intake affect human health not only through the uptake of nutrients directly, but also indirectly through changes in the microbial community and their associated metabolism. Here we conducted a two-week diet intervention in healthy young adults to investigate the impact of fiber consumption on the gut microbiome. Participants increased their average fiber consumption by 25 grams/day on average for two weeks. The high fiber diet intervention altered the gut microbiome of the study participants, including increases in known fiber degrading microbes such as *Bifidobacterium* and *Lactobacillus*.

INTRODUCTION

Consumption of dietary fiber has declined dramatically in the last century as processed foods have become a larger part of diets in the industrialized world. Pre-industrial and modern-day rural societies consume between 60-120 grams (g)/day of fiber, while individuals in the United States consume about half of the daily recommended allowance of 38 g/day for men and 25 g/day for women (110, 111). Declines in fiber intake over the past century have contributed to

complications for human health. For example, chronic low fiber intake has been associated with Type 2 diabetes mellitus, heart disease, and colon cancer (112–114). Indeed, a reciprocal diet intervention exchanging African Americans low-fiber western diet with rural Africans high-fiber diet (increasing on average 40g per day) led to significant decreases in pre-cancerous biomarkers, further providing a link between fiber and human health (115). Furthermore, dietary fiber has been shown to protect against influenza infection (116), and may influence vaccine efficacy (117).

Dietary fiber is a mixture of polysaccharides that resist rapid digestion in the small intestine by endogenous enzymes and persists through the digestive tract into the colon. Once in the colon, fiber can be digested by the resident microbes (110, 118). This is due, in part, to the human genome encoding only 17 enzymes (i.e., glycoside hydrolases) that are capable of digesting carbohydrates (119). Conversely, the resident gut microbial communities collectively encode thousands of diverse enzymes from 152 gene families that can break down dietary fiber (120). In the colon, specialized microbes metabolize recalcitrant carbohydrates and produce fermented byproducts, including short chain fatty acids (SCFAs) such as acetate, propionate, and butyrate (11). SCFAs are capable of being absorbed across the human intestinal epithelial cells, and have direct impacts on human health (reviewed in (121)) such as stimulating and maintaining the mucus layer for the gut epithelium (15) and providing an energy source for butyrate-consuming colonocytes (122). SCFAs have also been shown to have immunomodulatory effects, including increased viral protection through altered T-cell metabolism (116), and inhibitory effects on pathogenic bacteria (e.g. *Clostridioides difficile*) (123).

Understanding the role of dietary fiber in structuring the gut microbiota could provide insights into managing chronic diseases associated with the gut microbiome. Typical diet intervention studies assessing the impact of fiber on gut microbial communities and the production of SCFAs have relied on single fiber supplements (124–126). Fiber supplements such as psyllium husks, inulin, wheat bran, resistant potato starch, and resistant corn starch vary in their efficacy for each individual (124, 127). Individuals might be more or less susceptible to the intervention depending on their initial resident microbial community and its ability to digest a particular fiber supplement. For example, one group investigating the impact of three fermentable fibers on gut microbiome composition and SCFA abundance found no significant effect when study participants consumed 20-24g resistant maize starch per day for two weeks (124). However, in addition to the quantity, the variety of dietary fibers may be important. Studies that have increased dietary fiber have previously observed changes in microbiome composition (112, 124, 125), yet results remain mixed on SCFA production (17, 115, 128). Further, the American Gut Project found that individuals who eat more than 30 types of plants in a week have a more diverse gut microbiome (129). Thus, the consumption of a diversity of fiber sources through whole foods may provide more opportunities for an individual's gut microbiome to respond to the dietary changes and result in more dramatic changes in fiber degrader abundance and activity in the gut microbiome. The increase of fiber from a diverse set of dietary foods, rather than single fiber supplements, may also contribute to increased consumption of other micronutrients and vitamins that affect the microbiome as well (130).

In this study, we sought to answer three questions: 1) does a diet rich in fiber from whole foods alter the overall microbiome? 2) does the intervention alter the abundance and diversity of known fiber-degraders (e.g., *Bifidobacterium*)? and 3) if we observe compositional shifts in the

microbiome, do these correspond with metabolic changes in the production of short-chain fatty acids? To address these questions, we developed and employed a course-based undergraduate research experience (CURE) at UC Irvine to assess individual responses to a high-fiber diet (131). Integrating authentic research experiences within lab courses in order to facilitate a deeper understanding of academic and industrial research continues to be a priority for both national education reform and the American Society for Microbiology (131–134). During the intervention, participants were given ten meals each week from a food service that specializes in providing high fiber, unprocessed meals. Individuals tracked dietary information of macronutrients for every meal for three weeks, with the goal of increasing dietary fiber intake to 50 grams/day during a two-week intervention period. We then compared overall bacterial composition using metagenomic sequencing and assessed the production of volatile SCFAs using mass spectrometry. In addition to the shotgun metagenomic sequencing, we targeted a known-fiber degrader, *Bifidobacterium*, by analyzing its diversity using amplicon sequencing of the *groEL* marker gene, enabling a unique high-resolution view of the impact of a dietary fiber intervention on a key taxon.

METHODS

Study Design

Twenty-six UC Irvine students and instructors volunteered for a three-week high fiber diet intervention study (**Figure 2.1A**); only 22 individuals elected to provide stool samples for microbiome analyses (20 of whom we recovered enough sequence data for analysis, see **Supplemental Table 1**). The dietary intervention was approved by UC Irvine IRB # 2018-4297. For the first week of the study, all participants consumed their normal diets, tracking all nutritional information using the smartphone application MyFitnessPal (MyFitnessPal, Inc.).

Prior to the end of week one, each subject provided three fecal samples from three days within the first week. The intervention commenced in week two, when participants were instructed to raise their dietary fiber intake to approximately 40 grams per day. To assist with the dietary shifts, we provided 10 meals per week with ~15 grams of fiber ~5.8 unique fruits or vegetables per meal from the food delivery service Thistle (https://www.thistle.co/ San Francisco, California, USA). During week three, subjects were encouraged to further increase fiber intake to ~50 grams of fiber per day. Subjects provided three fecal samples from three days during week three, concluding the intervention period. As part of the CURE course, students were educated on human health, dietary information on high-fiber meals, the human gut microbiome, and the quantitative methods for microbiome analyses (from DNA extraction and library preparation to metagenome and statistical analyses), as previously described (131).

Sample collection

Subjects were given materials to collect fecal samples at home. Each stool sample was split into three 2ml tubes by the individual and immediately stored in the freezer. When convenient, students transported their anonymized and coded samples using cold packs and insulated boxes to a common lab freezer. Upon the conclusion of the intervention period (week 1 or 3), all samples were transported to a -20 °C freezer.

DNA extraction and metagenomic library preparation

To characterize the bacterial community composition of the samples, DNA was extracted with the ZymoBIOMICS 96 DNA Kit (Product D4309) from Zymo Research using the manufacturer's suggested protocol. Sequencing libraries were prepared using the Illumina Nextera kit and methods described in Baym et al. (68). Briefly, DNA was diluted to 0.5ng/µl and added to 0.25µl of Nextera enzyme and 1.25 µl of Tagmentation Buffer. This mixture was

incubated at 55 °C for 10 minutes and then placed on ice for the remainder of the protocol. Barcodes were added using the Phusion polymerase (New England Biolabs) and excess adaptors were cleaned using AMPure XP (Beckman Coulter Life Sciences) magnetic beads. Quality and concentration were assessed using a Picogreen assay (ThermoFisher) and the distribution of fragment sizes was determined using a Bioanalyzer. These libraries were loaded onto the Illumina Next-Seq 500 at 1.8 picomolar concentrations and sequenced using Illumina's mid-output kit for 75 bp paired-end sequencing, resulting in a total of 144,023,583 reads and an average of 1,425,976 reads / sample (max: 5,902,966; min: 7) (**Supplemental Table 1**).

Amplicon library preparation

To characterize the genetic diversity of *Bifidobacterium* at a finer-genetic scale than could be assayed by metagenomics, we used genus-specific primers to target this group for sequencing (135). Sequencing libraries were prepared by setting up an initial 25 µl PCR reactions with AccuStart II PCR ToughMix (2x), the *groEL* forward primer (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCCGATTACGAYCGYGAGAAGCT-3', 20 µM), and the *groEL* reverse primer (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCSGCYTCGGTSGTCAGGAACAG-3', 20 µM). The initial PCR ran for 28 cycles 95°C for 30 sec, 60 °C 30 sec, 72 °C 50 sec followed by the addition of 0.5 µl of dual Nextera XT index (Illumina) to each sample proceeding with an additional 8 cycles 95 °C 30 sec, 60 °C 30 sec and 72 °C 50 sec. Amplicons were pooled based on visual quantification of the bands on an agarose gel and purified using magnetic Speed Beads The pool was run on a MiSeq PE 300 at University of California Irvine's Genetic High Throughput Facility resulting in a total of 20,052,935 reads and an average of 185,675 reads/sample (max: 6,815,601; min: 155).

SCFA extraction and measurements

SCFA extractions were done following the methods by Zhao et al. (2005) (136). One-hundred mg of fecal material was added to 1ml of HPLC grade water and vortexed for two minutes. Ten microliters of 6N HCl was added to the fecal slurry and vortexed briefly. This mixture was incubated at room temperature for 10 minutes with occasional shaking. Afterwards the mixture was centrifuged at 14,000g for 1 minute, and 400 µl of the supernatant was transferred to a new tube, which was then filtered through a 0.22 µm filter. An aliquot (200 µl) of this suspension was then transferred to a glass vial with a 0.2 ml vial insert and stored at -20 °C. When running the sample, 10 µl of an internal standard of 10mM ethyl butyrate was added to the extraction prior to the run. Before running each sample, the instrument was calibrated using a standard comprising 100 mg/l of acetate, propionate, isobutyrate, butyrate, isovalerate, valerate, and ethyl butyrate. Six samples were run on an Agilent 7890A gas chromatograph with dual column FID detectors. Two microliters per extracted sample were hand-injected on a stainless-steel column (2 meters x 3.2 mm) containing 10% SP-1000 and 1% H3PO4 on 100/120 Chromosorb W AW (Supelco, Inc., Bellefonte, PA, USA). The flow rate of the N2 carrier gas was 26.14 ml/min. Between sets of six samples the instrument was washed using water and phosphoric acid. Peaks were auto integrated using ChemStation v1.0 on a PC running Windows 2000 (Microsoft). A subset of samples (n = 44 from 8 individuals) were run in duplicate to examine technical variation (see coefficient of variation (CV) in **Table 2.1**), and the average CV was 55%.

Metagenomic sequence analysis

Raw shotgun metagenome sequences were filtered using Prinseq v0.20.4 (73) to remove sequences that had a mean quality score of 30 or less. Reads from human DNA were also removed by aligning the filtered reads to the human genome (hg38), using Bowtie2 v2.2.7 (74),

and keeping the reads that failed to align. A total of 130,755,383 paired-end reads (average 1,294,607 non-human reads/sample) were retained and passed through MIDAS, which assigns taxonomy to short read data using a marker gene approach (137). Species counts per sample represent the average of 100 subsamples, rarefied to 900 sequences per sample using the EcolUtils (v0.1) package in R. Taxonomy was also assessed using IGGsearch (138). To analyze functional differences related to SCFA metabolism between high and low fiber treatment groups, HUMAnN3 (75) was used with default parameters. All pathways within the MetaCyc pathway class "Fermentation to Short-Chain Fatty Acids" were searched for within the HUMANnN pathway output, which resulted in nine pathways used for analysis (139). For genes related to carbohydrate breakdown, we translated reads using Prodigal (140) to predict open reading frames (ORFs) and searched all ORFs against the Pfam database (141) with hmmer/3.1b2 (142). Resulting PFAM annotations were then screened against the CAZyDB.07202017 (143) with Blast/2.8.1 (144) using alignments >70% amino acid identity and 30% coverage. Alpha diversity and PERMANOVA analyses were performed using the Vegan v2.5-6 (145) package in R (146). Non-metric multidimensional analysis was done using the metaMDS function in Vegan on Bray-Curtis distances. StrainPhlAn (147), under default parameters, was used to analyze strain-level variation within the metagenomes. To root the phylogenetic tree, *Prosthecochloris aestuarii* (accession: GCA_000020625) was used, and two reference genomes of *Eubacterium rectale* (accession: GCA_000209935 and GCA_001404855).

GroEL amplicon analysis

We downloaded 780 genomes from the genus *Bifidobacterium* on the PATRIC database (148). All genomes were screened for completeness by searching for 21 single-copy ribosomal marker genes using Prodigal (140) and HMMer v3.1b2 (142) with an E value of 1×10^{-10}. The

remaining 578 genomes were used to create a multi-locus, concatenated phylogeny of the ribosomal marker genes with ClustalO v1.2.0 (149) to produce a 4272 amino acid alignment for phylogenetic analysis using RAxML v8.0.0 (150) under the PROTGAMMABLOSUM62 model for 100 replicates. Next, we parsed the filtered genomes for the *groEL* gene sequences by using 260 non-redundant gene sequences to build a *groEL* phylogeny under identical parameters to the whole-genome analysis. The *groEL* amino acid sequences, alignment, and phylogeny were used to construct BLASTp, HMMer, and pplacer reference databases for metagenomic analyses.

For each *groEL* amplicon library, sequences were quality trimmed and adapters were removed with BBDuk (151) (qtrim=rl trimq=10 ktrim=r k=25). Paired end sequences were merged together with BBMerge (151) and, if paired reads did not overlap, only the forward read was retained. The reads were then searched against the *groEL* reference databases using BLAT (152) and hmmsearch, respectively. Passed reads were aligned with ClustalO to the pplacer reference package and placed onto the *groEL* reference phylogeny using pplacer v1.1.alpha17 (153). Relative abundance was calculated from the single branch assignments and aggregated at the species level to be normalized by the total number of extracted *groEL* gene sequences. We show that the phylogenetic relationship between species of *Bifidobacterium* based on the *groEL* gene closely reflects a phylogeny based on 21 single copy marker genes from 578 *Bifidobacterium* genomes (**Figure S4)**.

Statistical analysis

Permutational analysis of variance (PERMANOVA) was conducted on Bray-Curtis dissimilarities at the genus level with 999 permutations using the Adonis test in the Vegan package in R (see Data availability and GitHub). We tested the effect of the intervention (pre- versus post-fiber increase), the effect of the individual, and the interaction between these two

factors. Genus contributions to significant results from the PERMANOVA model were determined by passing the resulting PERMANOVA object through the coefficients function found in the base Stats package R. A similar procedure was used to analyze compositional differences between CAZy enzymes and HUMAnN gene predictions in the metagenomes, with permutations on Euclidean distances. Linear mixed effects models, using the nLME package (44) in R, were also conducted for comparison because they take repeated measures into account. Specifically, to support the PERMANOVA analysis of beta diversity, an LME was performed on the rank-transformed first principal coordinate of a principal coordinates analysis on the Bray Curtis community dissimilarity matrix. Individual was used as the random effect and the model used the default autoregressive (Lag 1) structure (AR1) for regression across a time-series. For the functional analyses, reads analyzed using HUMAnN3 were normalized by copies per million; CAZy were normalized to the total number of reads per metagenome and compared using Wilcoxon rank sum test. Gene features for HUMAnN were reduced by analyzing only unstratified data, for which 70% of samples had non-zero reads mapping to each feature. HUMAnN pathway abundances were analyzed in their entirety with stratification and without feature reduction. Lefse (76) was used to determine pathways which may differentiate pre- vs post-intervention samples. Wilcoxon rank sum tests were also used to compare nutritional and gene differences between intervention periods when residuals were not normally distributed and reads or macronutrients were averaged out within individuals (by treatment) to account for repeated measures. When normality assumptions of residuals were met (tested using the Shapiro-Wilk test) ANOVAs were used. To assess which taxa were correlated with changing amounts of fiber, all species within each sample (the rarefied species abundance matrix) and fiber were correlated using the Corrr package v0.4.2 (154) in R. To analyze which genera co-correlate with

the genus *Bifidobacterium*, Spearman correlations were used and, where appropriate, p-values were corrected (q-value) for multiple comparisons using a false discovery rate cutoff of 0.05. To assess significance of strains between individuals, cophenetic distances were calculated on the RAxML tree output from StrainPhlAn and passed into the above PERMANOVA model.

Data availability

All scripts are stored on GitHub (https://github.com/aoliver44/Fiber-Analysis). All metagenomic and amplicon sequences are available on NCBI under the Bioproject PRJNA647720. Metadata linking the shotgun metagenomes and *groEL* sequences with the appropriate sample ID and intervention can be found in Supplementary Table 1.

RESULTS

Dietary intervention within the CURE course

Twenty-six individuals participated in a CURE course at UC Irvine, designed to tandemly investigate pedagogical methods (131) and the role of fiber on the microbiome. We collected nutritional data from all 26 individuals who initially began the intervention, over three weeks (one week prior to, and two during, the dietary intervention) (**Figure 2.1A**). We collated the total amount of macronutrients consumed per day, including fiber, protein, carbohydrates, fats, as well as overall calories (**Figure 2.1B-F**). Additionally, we informally surveyed food items the study participants frequently used to supplement their meal plans, beyond the meals supplied from Thistle, and found that items such as fiber fortified cereals, lentils or beans, and berries were common (131). For the intervention, subjects increased their average fiber consumption from 21.0 g/day (± 14.2 g/day) before the intervention to 46.4 g/day (± 12.5 g/day) during the intervention (**Figure 2.1B**; Wilcoxon rank sum test, $p < 0.0001$). While these dietary shifts increased carbohydrate intake by an average of 84% (36 g) during the intervention ($p = 0.013$),

other macronutrients measured, such as calories, fat, and proteins, did not significantly change ($p > 0.05$) (**Figure 2.1C-F**).

Diet intervention altered gut microbial community composition within individuals

To evaluate whether increased fiber consumption contributed to shifts in the gut microbiome, we characterized the microbial communities from 20 individuals using 86 shotgun metagenomic libraries collected before and after the fiber intervention (**Figure 2.2A**). Alpha-diversity of microbial taxa decreased during the high fiber diet intervention as measured by the Shannon diversity index (**Figure 2.2B**)(Wilcoxon rank sum test, $p < 0.05$). Using alternative approaches to assess taxonomy and diversity (see methods) showed either no change or supported the decreasing trend of diversity during the intervention period (**Figure S2.1).**

Despite little difference in alpha-diversity, beta-diversity changed significantly in response to a high fiber diet. Multivariate analysis of marker gene abundances showed that most of the variation in microbiome composition could be explained by the individual (PERMANOVA: main individual effect: $R^2 = 0.78$, $p < 0.001$, **Supplemental Table 2**). The diet intervention shifted the microbial composition of the entire study cohort significantly (main intervention effect: $R^2 = 0.014$, $p < 0.001$). Within samples from each individual, the pre- and post- diet intervention samples explain significant variation in the community composition (intervention-by-individual effect: $R^2 = 0.083$, $p < 0.001$). A linear mixed-effects (LME) model confirmed these results, which identified diet as a significant determinant of an individual's microbiome composition (LME, $p < 0.01$). Individual gut microbiome samples grouped together in nonmetric multidimensional space (nMDS; **Figure 2.2C),** further providing support that each individual is associated with a unique microbiome. Some individuals (i.e., Individual 13) gut microbiomes were more distinct from others (**Figure 2.2C inset)**. Additionally, we used

Eubacterium rectale (due to its high coverage in our data) to ask whether the diet intervention had an impact at the strain level. Strains were highly individual specific (PERMANOVA: main individual effect: $R^2 = 0.99$, $p < 0.001$) and did not change in response to increased fiber intake ($p > 0.05$; Figure 2D).

We next parsed the taxonomic data to assess which microbial taxa increased or decreased in response to the diet intervention. One species in the family *Lachnospiraceae* was significantly negatively associated with increasing fiber intake (Spearman, r = -0.43, q = 0.01) (**Figure S2.2A, B**). *Coprococcus sp.* and *Anaerostipes hadrus* were both positively associated with increasing fiber intake, but this association was not significant when p-values were FDR-corrected for multiple comparisons (r = 0.32, q = 0.33 both species) (**Figure 2.2A**). Furthermore, positive linear coefficients of a PERMANOVA model, which detect differences between community composition due to the diet intervention, included genera such as *Bifidobacterium*, *Bacteroides*, and *Prevotella* (**Figure 2.3A**). Conversely, *Blautia* and *Ruminococcus* contributed negative linear coefficients to the PERMANOVA model (**Figure 2.3A**).

Bifidobacterium *species were enriched by the diet intervention*

Of the 105 microbial genera detected in this study, *Bifidobacterium* was the strongest predictor genus for the post-intervention microbiomes (**Figure 2.3A).** Indeed, taxonomic analysis of the metagenomic samples identified *Bifidobacterium* abundances increasing, on average, 1.4-fold between the pre- and post-intervention periods (**Figure 2.3A)**. Further, we identified several species of *Bifidobacterium* present within and across individuals, with *B. adolescentis* being the most abundant species on average (**Figure 2.3B).** When we investigated the taxonomic profiles at the species level, we found that *B. adolescentis*, *B. biavatii*, *B. breve*, *B.*

longum, and *B. ruminantium* all increased in mean abundance on a high fiber diet whereas the other, lesser abundant species exhibited no change or decreased in abundance (**Figure S2.3B**).

Given that *Bifidobacterium* was the strongest predictor genus in the post-fiber gut microbiomes, we employed a targeted analysis into the diversity within *Bifidobacterium* to examine species-level patterns. Specifically, we applied targeted amplicon approaches to amplify the *groEL* gene, a conserved phylogenetic marker gene to track *Bifidobacterium* diversity (**Figure S2.4)**. Using phylogenetic inference of the *groEL* gene, we compared the observed *Bifidobacterium* diversity observed at the community level to our targeted analysis of the *groEL* gene. Similar to the metagenomic analysis, we found that individuals were largely comprised of *B. adolescentis* and *B. longum*, with six other abundant species of *Bifidobacterium* (**Figure 2.3C).** This analysis also revealed extensive *Bifidobacterium* diversity within the human gut, detecting 22 species across all individuals.

Since *Bifidobacterium* species are known to participate in cross-feeding with other gut microbes (reviewed in (155)), we next assessed the co-occurrence of *Bifidobacterium* with other genera. *Bifidobacterium* was positively correlated ($r = 0.43$, $q = 0.001$) with an increasing abundance of *Lactobacillus* and negatively correlated with *Roseburia* ($r = -0.49$, $q = 0.0002$) and *Ruminococcus* ($r = -0.38$, $q = 0.007$) (**Figure S2.5)** suggesting possible species interactions between these taxa.

Genes involved in inositol degradation increase on high fiber diet

Our results demonstrate that a shift in dietary fiber consumption influenced compositional changes in the gut microbial community. As such, we sought to correlate the observed taxonomic shifts to functional shifts, particularly the enrichment of genes related to carbohydrate degradation. Despite taxonomic shifts at the individual level, we observed no

changes in in the overall abundance (average number of normalized reads) mapping to gene families for glycoside hydrolases (GH) (Wilcoxon, $p = 0.42$), carbohydrate esterases ($p = 0.58$), glycoside transferases ($p = 0.73$), and polysaccharide lyases ($p = 0.77$) as a result of the intervention (**Figure S2.6A**). No individual families GH and polysaccharide lyase CAZy classes changed in abundance during the intervention when corrected for multiple comparisons (Wilcoxon, $p > 0.05$, **Figure S2.6B**). Further, the diversity (**Figure 2.4A,** n = 106 families, ANOVA, $p > 0.05$) and composition (PERMANOVA, $p > 0.05$) of GHs detected were indistinguishable between the pre- and post-intervention samples. Compositional analysis all genes identified by HUMAnN revealed no individual signature (PERMANOVA: main individual effect: $R^2 = 0.017$, $p > 0.05$, **Figure 2.4B**), and no shifts in response to a high fiber diet (intervention-by-individual effect: $R^2 = 0.015$, $p > 0.05$). We performed a linear discriminant analysis to determine if there were pathways that were differentially abundant due to the diet intervention, and found inositol degradation (in addition to several unintegrated pathways) to be increased in abundance on a high fiber diet (**Figure 4C, Figure S2.7**). For the pathways involved in SCFA metabolism, we found no significant (Wilcoxon, $p > 0.05$) changes as a result of the high fiber diet (**Figure 2.4D**).

Fecal short-chain fatty acid concentrations were unaltered by the diet intervention

While the presence of genes related to SCFA production provide insights into the functional changes of the microbiome, these genes only reflect the genomic potential to process these pathways. Therefore, we applied a targeted GC/MS analysis on 149 samples from 18 individuals for the presence of SCFA molecules. Across the intervention period, the average abundance of acetate, propionate, butyrate, and valerate increased (**Table 2.1**); however, these increases were not statistically significant (LME, $p > 0.05$) (**Figure 2.5)**. For the eight individuals with samples

run in duplicate, three had biological differences that were greater than the technical variation seen in the duplicates (**Figure S2.8**). Acetate had the least technical variation (mean CV = 45%), followed by propionate (mean CV = 51%) (**Table 2.1, Figure S2.8**).

DISCUSSION

We examined the impact of dietary foods, rich in their diversity of fiber, on the human gut microbiome. We expected that an increase in fiber consumption through whole foods consumed would lead to a more generalizable shift of the microbiome in contrast to previous studies that utilize a single fiber supplement. For instance, a recent meta-analysis (156) found mixed results in how fiber may impact the gut microbiome richness and composition. Among papers published prior to October 2017, only 18% (12 out of 64) studies (157, 158, 167, 168, 159–166) contained food-based fiber interventions, and most of these studies only modified one aspect of diet (e.g., addition of whole grain breakfast cereal). One study in particular increased dietary fiber by 40g from a diverse set of foods during a five day period (158). The authors similarly found microbiome composition changes within individuals when they accounted for differences in the subjects' starting microbiomes. Despite the variation in implementing a fiber intervention, it is becoming increasingly clear that fiber alters the composition of the gut microbiome (124) and the associated microbial changes affect human health (i.e., type 2 diabetes mellitus (112)). A common observation in fiber intervention studies (156) is the specific involvement of the genus *Bifidobacterium* in response to fiber interventions. However, to our knowledge, no study has documented how fiber impacts the genus at the strain-level in the human gut.

Does a diet intervention rich in fiber alter the microbiome?

Past studies have shown that an increase in the diversity of dietary foods could lead to an increase in microbial diversity (130). Moreover, individuals living in rural societies often harbor far greater gut microbial diversity than individuals from western societies (20, 169, 170), which may in part be linked to a greater proportion of plant-based polysaccharide intake. However, we did not measure an increase in species diversity (alpha diversity) after subjects consumed >40g of fiber from a diverse set of foods (**Figure 2.2B**). These results could be attributed to the brevity of the intervention as the rapid change in dietary composition may result in the loss of microbes poorly adapted to recalcitrant carbohydrates. Similarly, other studies have reported finding no increases in alpha diversity as a result of a fiber intake (158, 171–174), which may indicate a trade-off where fiber-degraders increased while other taxa decreased. Although alpha diversity was unaffected, we did observe a significant impact of the high-fiber diet on microbial community composition (beta diversity) (**Figure 2.2**). The composition of microbial communities within individuals shifted ~8% during the intervention period. We found changes in communities to be at broader taxonomic levels than strain-level. We were able to examine strains of *E. rectale* due to its high coverage in our data, and showed these strains stayed constant and individual specific during the intervention (**Figure 2.2D**). Future work should determine if this pattern holds up for other species. While we suspect the high fiber diet treatment played an instrumental role in shifting the microbial composition, we cannot rule out other factors such as host genetics or non-dietary behaviors. As discussed, many food-based fiber interventions have shown mixed results on changing the microbial communities (115, 175). The drastic increase in fiber from a variety of foods may lead to rapid shifts in community composition over the two-week period. Changes in community composition pre- and post- intervention were largely driven

by shifts in known-fiber degraders, such as *Bifidobacterium*, *Bacteroides*, and *Prevotella* (**Figure 2.3A**).

We expected the taxonomic shifts in the microbiota would be associated with changes in the functional potential of the microbial communities (**Figure 2.4**). While we initially hypothesized that a high fiber diet would increase the abundance or diversity of carbohydrate active enzymes, we did not detect changes associated with the intervention (**Figure S2.6)**. Our findings support a similar result showing no difference in CAZy abundance due to increased fiber intake (174). We acknowledge that sequencing depth is an important consideration in the detection of genes; increasing reads beyond our ~1.3 million paired-end reads (avg per sample, **Supplemental table 1**) may allow for greater detection. However, we did find a notable increase in the abundance of genes mapping to the inositol pathway (**Figure 2.4C**). We suspect that the increased consumption of fiber-fortified cereals and legumes, which contain higher levels of inositol, during the diet intervention allowed for an expansion in organisms capable of breaking down this sugar. There is substantial interest in the role of inositol (specifically phytic acid) in its protective role against colon cancer and other metabolic disorders (176, 177). Next, we assessed whether genes involved in SCFA metabolism changed in abundance during the intervention. Although appreciable cross-feeding between lactate-producing *Bifidobacterium* spp. and butyrogenic bacteria has been shown (178) we did not find significant increases in genes involved in various SCFA metabolic pathways (**Figure 2.4D**). This further supports our results showing no clear correlations between *Bifidobacterium* spp. and butyrate-producers within our diet intervention (**Figure S2.5)**. Indeed, we would not be the first to suggest that perhaps these complex trophic interactions require more time to establish (124). Rather, our results suggest that

while broad taxonomic shifts occur, these do not correspond to changes in functional potential and fine-scale (intraspecies) shifts are less susceptible to dietary shifts on short-term timescales.

Does the intervention alter the abundance and diversity of Bifidobacterium*, a known fiber-degrader?*

Many studies have indicated that bifidobacteria (often identified as the genus *Bifidobacterium* by FISH probes, PCR, or DNA-sequencing) are highly abundant in the gut following increased fiber intake (meta-analysis of 51 studies (156)). Increased abundance of *Bifidobacterium* is somewhat unsurprising, as they harbor numerous genetic components, such as carbohydrate active enzymes, that make them especially adapted to a fiber-rich diet (179). In one study, both resistant potato starch and inulin increased the relative abundance of *Bifidobacterium* spp.; however, the 16S amplicon sequencing in this study did not have the resolving power to identify which species of *Bifidobacterium* were increasing (124). Using a targeted amplicon approach, the *groEL* gene, has been shown to delineate species of *Bifidobacterium* that otherwise share >99% sequence identity in the 16S rRNA gene, making it a robust marker gene for analyzing within-genus species diversity (180). In our study, the most abundant species of *Bifidobacterium* were *B. adolescentis* and *B. longum*, both of which are efficient degraders of plant-based fructo-oligosaccharides (FOS) and produce acetate and lactate in the process (181). Mirroring our results, other studies have found selective increases in certain species of *Bifidobacterium* as a result of carbohydrate intake; for example, in one study, intake of inulin resulted in a greater increase of *B. adolescentis* (182). We speculate that on a high fiber diet, bifidobacteria are the initial members of the community accessing fiber substrates, easily adapted to utilize various FOS, and pivotal to the creation of the initial metabolic cross-feeding networks.

Future studies should extend the intervention period to examine the dynamics of longer-term trophic interactions in response to increased dietary fiber intake.

Can we detect diet-induced changes in the abundance of fecal short-chain fatty acids?

While SCFAs did generally increase during the diet intervention, trending toward their naturally occurring gut ratio of 3:1:1 (acetate:propionate:butyrate) (11, 183, 184), we did not observe a statistically significant increase in SCFAs post-intervention. Static fecal concentrations of SCFAs may not reflect the total pool of molecules fluxing through a given individual, as the molecules are preferred substrates of the cells lining the gut epithelia (122). It is also possible that the intervention period was too short to observe increases in SCFA abundances.

It should be noted that accurate SCFA measurements are notoriously difficult. Our examination of technical variability within 44 samples from eight individuals showed that technical variation between pre-intervention replicates or post-intervention replicates was greater than the average difference between pre- and post-intervention for any given SCFA. One study reported high intra-fecal variability of butyrate quantification (coefficient of variation = 38%), prior to optimizing a freeze-drying method (185). Numerous studies have indicated the benefit of SCFAs to human health (114, 186); yet the heterogeneity in reported acetate, propionate, and butyrate abundances remains high. In one meta-analysis of fiber studies, only butyrate was generally found to increase with fiber intake, yet the heterogeneity of reported results was 70% (I^2), similar to other SCFAs analyzed (156). Outside of technical limitations, shifts in microbial community structure are not predictive of changes in static measurements of fecal SCFA abundances (187). The difficulty of finding meaningful correlations between microbiome composition and SCFA abundances likely reflects a failure to measure both circulating and fecal SCFAs across time in conjunction with microbial abundances. Indeed, it has been observed that

fecal levels of acetate are inversely related to the rate of its absorption (188). Future studies are needed to confirm whether correlation analysis between fecal SCFAs and microbiome composition is a useful tool to understand the interplay between microbiome, SCFAs, and health.

In sum, our results indicate that gut microbial communities are malleable to an influx in recalcitrant carbohydrates, contributing to significant community and functional shifts in certain metabolic pathways. However, these compositional changes did not correspond to broad functional changes, at least over the short-term timescales for this intervention. Further studies exploring the impact of timing and composition of dietary fiber interventions, particularly while taking into account the starting composition of the gut microbiomes of study participants, are critical for understanding the generalizability of fiber interventions for engineering microbiomes. Increasing fiber intake could have the most impact in contexts where low gut microbial diversity increases risk of *C. difficile* infection, such as for nursing home residents, cancer patients or after antibiotic treatment.

ACKNOWLEDGMENTS

We would like to acknowledge the T32 training grant which supported Andrew Oliver (1T32AI14134601A1) from UC Irvine's training program in microbiology and infectious diseases. We would also like to acknowledge the UCI Microbiome Initiative for supporting the study, Thistle, for supporting our provision of high fiber meals, and Heather Maughan for thoughtful edits to the manuscript. We would like to give an enormous thank you to the students from the course M130L at UC Irvine Spring in 2018.

Table 2.1: Abundances of four SCFAs in samples pre- and post- intervention

	PRE (MEAN)	PRE (SD)	PRE CV	POST (MEAN)	POST (SD)	POST CV
ACETATE	480 mg/L	208 mg/L	43 %	525 mg/L	244 mg/L	47 %
PROPIONATE	277 mg/L	112 mg/L	41 %	288 mg/L	175 mg/L	61 %
BUTYRATE	244 mg/L	107 mg/L	44 %	257 mg/L	176 mg/L	68 %
VALERATE	49 mg/L	31 mg/L	65 %	43 mg/L	31 mg/L	72 %

Figure 2.1: Intervention timeline and sample collection. The study subjects began eating their normal diet for one week, tracking all their food intake using the MyFitnessPal app. At the end of the week, each individual provided a daily fecal sample on three different days. At the start of week two, subjects started a wholesome high fiber diet, getting at least 40 grams of fiber per day. During week three, subjects were encouraged to get 50 or more grams of fiber per day. At the end of week three, each subject provided a fecal sample on three different days. B-F) Self-reported macronutrients from individuals using the MyFitnessPal app. Change in macronutrients across the 3-week diet intervention for B) fiber, C) carbohydrates, D) protein, E) fat, and F) calories. Fiber changed the most in magnitude between pre-intervention intake and during the diet intervention (linear mixed-effects model, $p < 0.001$). There were modest, but significant, changes in carbohydrate, protein, and caloric intake, but not fat intake, across the same time interval.

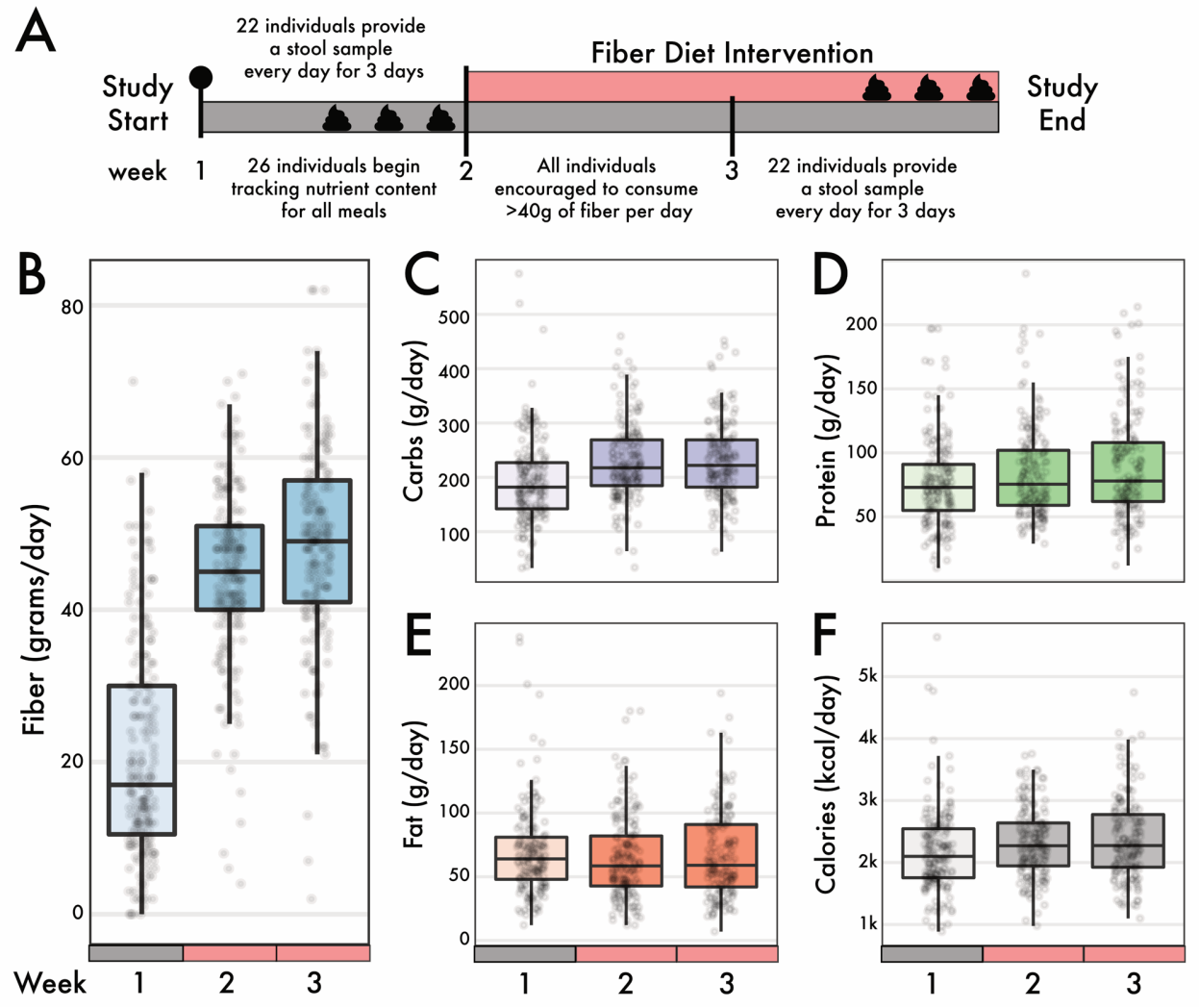
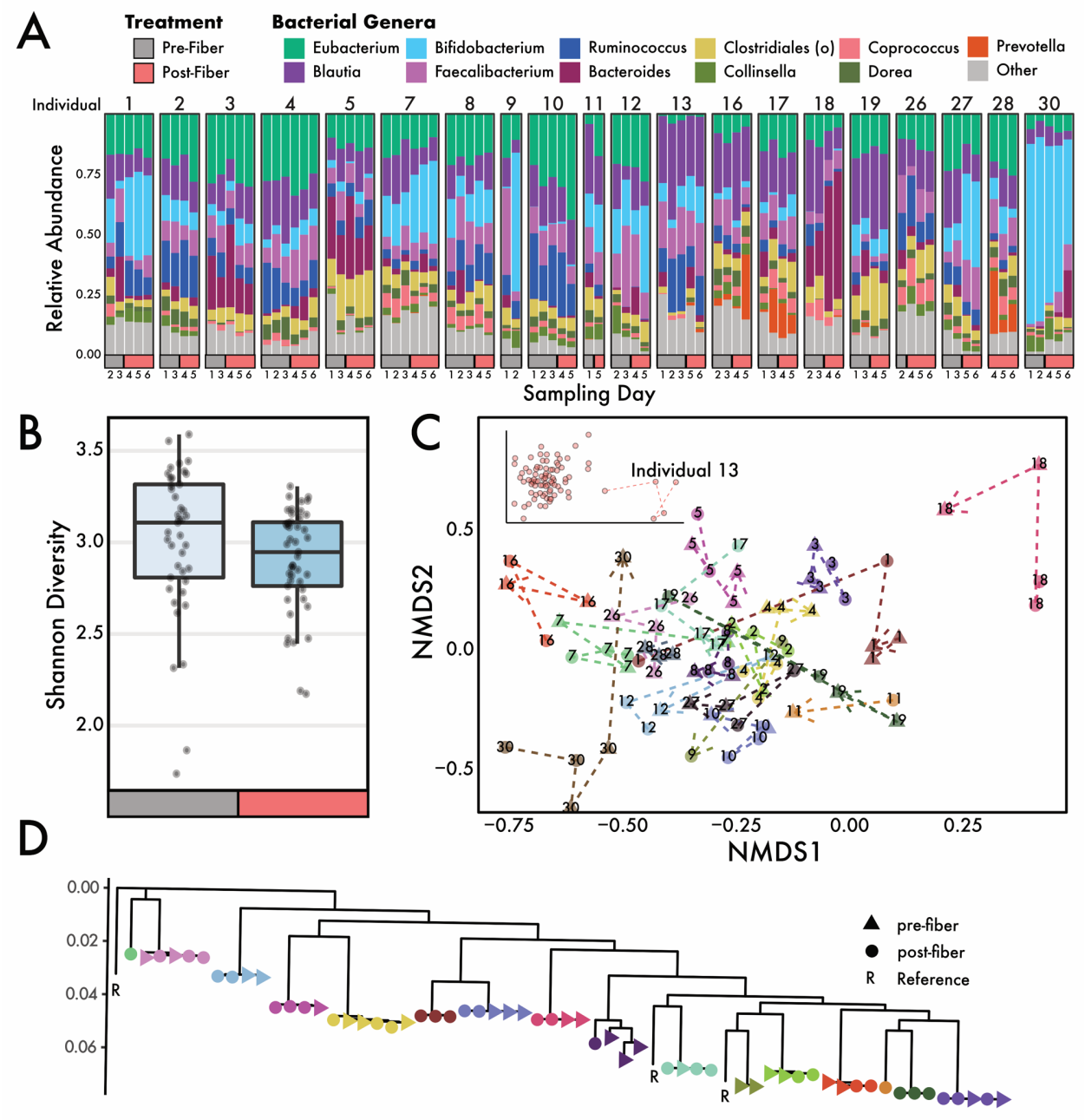

A
22 individuals provide
a stool sample
every day for 3 days
Fiber Diet Intervention
Study
Start
Study
End
week 1
26 individuals begin
tracking nutrient content
for all meals
2
All individuals
encouraged to consume
>40g of fiber per day
3
22 individuals provide
a stool sample
every day for 3 days
B
Fiber (grams/day)
0
20
40
60
80
Week 1 2 3
C
Carbs (g/day)
100
200
300
400
500
1 2 3
D
Protein (g/day)
50
100
150
200
1 2 3
E
Fat (g/day)
0
50
100
150
200
1 2 3
F
Calories (kcal/day)
1k
2k
3k
4k
5k
1 2 3

Figure 2.2: Microbiome community composition through a dietary fiber intervention. A) Relative abundances of genera detected in microbiomes from individuals throughout the diet intervention study. B) Alpha diversity, measured using the Shannon index, changed significantly during the intervention period (Wilcoxon, $p < 0.05$). C) NMDS ordination showed that samples from individuals mostly group together. Dotted lines connect the same individual and point towards the final post-fiber intervention sample. Samples in this study were highly personalized: the individual explained 78% of the variation in the data. The inset shows an extended version of the NMDS plot that includes Individual 13. D) A phylogeny of *Eubacterium rectale* strains found in individuals (denoted by color) during the intervention.


A
Treatment
Pre-Fiber
Post-Fiber
Bacterial Genera
Eubacterium
Blautia
Bifidobacterium
Faecalibacterium
Ruminococcus
Bacteroides
Clostridiales (o)
Collinsella
Coprococcus
Dorea
Prevotella
Other
Individual
Relative Abundance
Sampling Day
B
Shannon Diversity
C
Individual 13
NMDS1
NMDS2
D
pre-fiber
post-fiber
R Reference

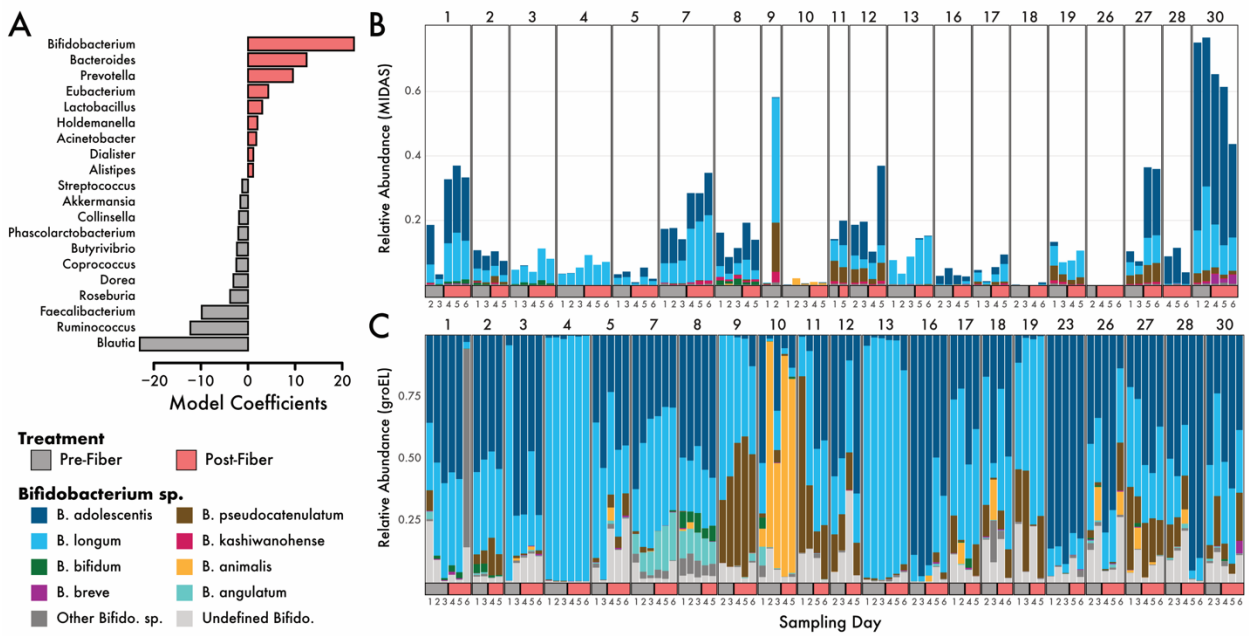
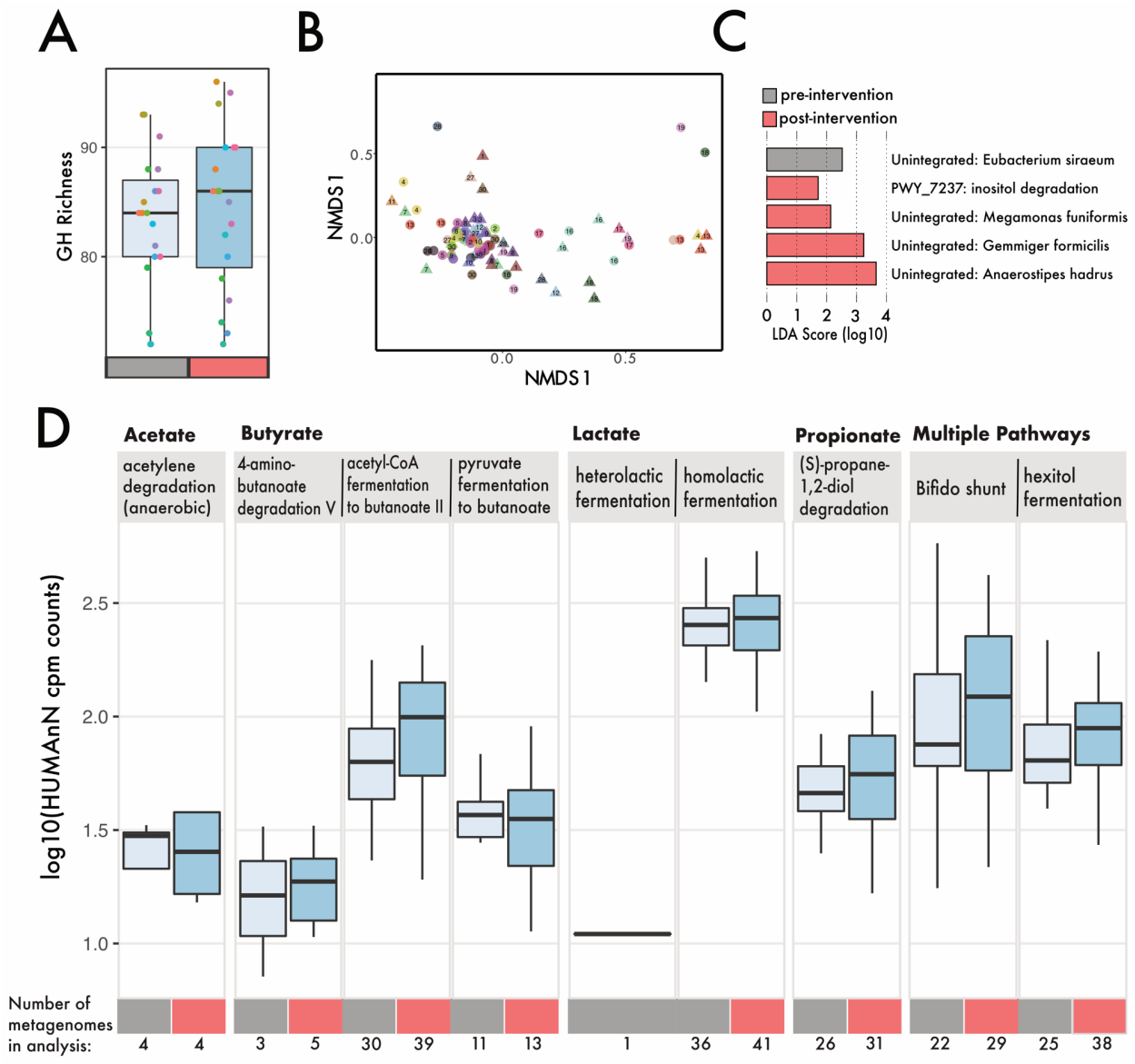


Figure 2.3: GroEL amplicon analysis of *Bifidobacterium* during the fiber intervention. A) Model coefficients of the PERMANOVA analysis (model: species ~ Individual*Intervention). Species with high coefficients (positive or negative) were best able to distinguish the pre vs post diet intervention groups. Only the top 20 genera are shown. The genus Bifidobacterium had the largest positive coefficient, indicating that it was important to the model for disguising

microbiomes before and after the diet intervention. Relative abundances of 12 detected species of Bifidobacterium from B) shotgun metagenomics and C) groEL amplicon sequencing.

Figure 2.4: Genes involved in carbohydrate degradation and SCFA metabolism within metagenomes. A) Number of distinct glycoside hydrolase families within individual metagenomes (different colored circles), separated by pre-intervention (mean = 83) (grey) and post intervention (mean = 84) (red). B) NMDS ordination of Euclidean distance matrix based on 19680 gene features, shape denotes intervention (triangle = pre-, circles = post-) and individuals are separated based on color. C) Lefse analysis of pathways that differentiate samples by intervention. D) Log abundance (copies per million) of pathways involved in SCFA production.


A
GH Richness
90
80
B
NMDS 1
NMDS 2
0.5
0.0
0.0
0.5
C
pre-intervention
post-intervention
Unintegrated: Eubacterium siraeum
PWY_7237: inositol degradation
Unintegrated: Megamonas funiformis
Unintegrated: Gemmiger formicilis
Unintegrated: Anaerostipes hadrus
0 1 2 3 4
LDA Score (log10)
D
Acetate
Butyrate
Lactate
Propionate
Multiple Pathways
acetylene degradation (anaerobic)
4-amino-butanoate degradation V
acetyl-CoA fermentation to butanoate II
pyruvate fermentation to butanoate
heterolactic fermentation
homolactic fermentation
(S)-propane-1,2-diol degradation
Bifido shunt
hexitol fermentation
log10(HUMAnN cpm counts)
2.5
2.0
1.5
1.0
Number of metagenomes in analysis:
4 4 3 5 30 39 11 13 1 36 41 26 31 22 29 25 38

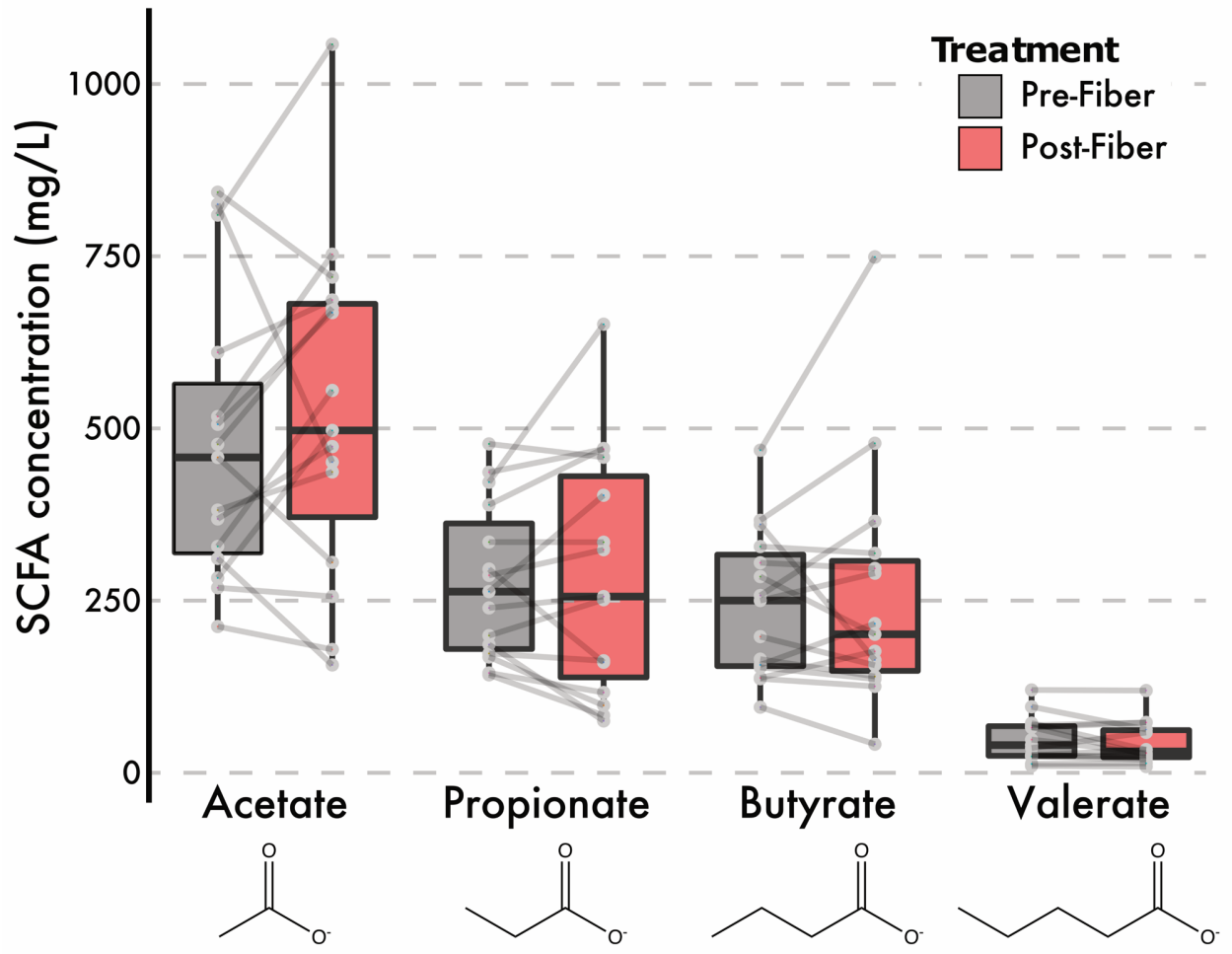


Figure 2.5: GC-FID measurements of fecal volatile SCFAs during intervention. Fecal SCFA abundances, averaged across replicates where applicable, before and after the intervention.

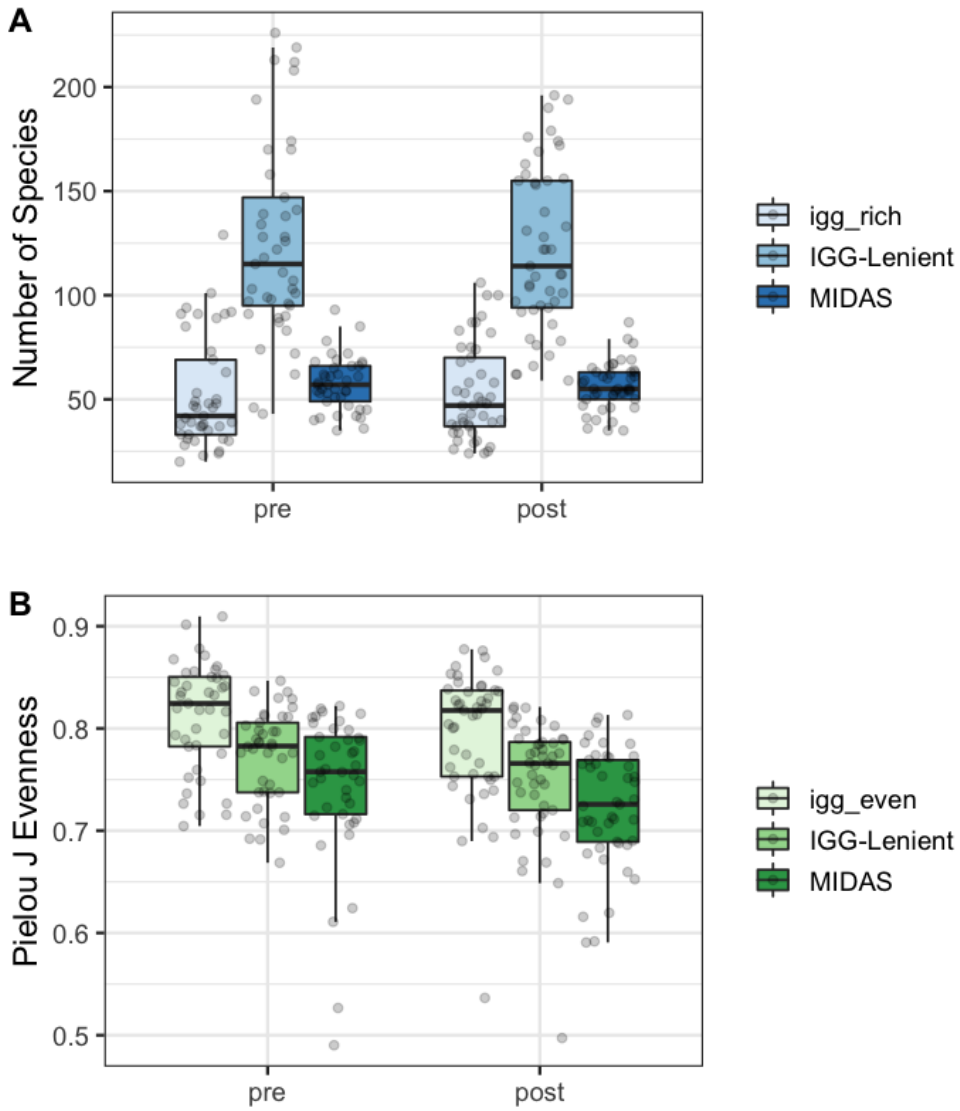


Figure S2.1: Comparisons of A) richness and B) evenness obtained using different databases for taxonomic assignments. IGG_rich and MIDAS were run using default parameters. IGG-Lenient was run at 25% species quality and 15% marker genes.

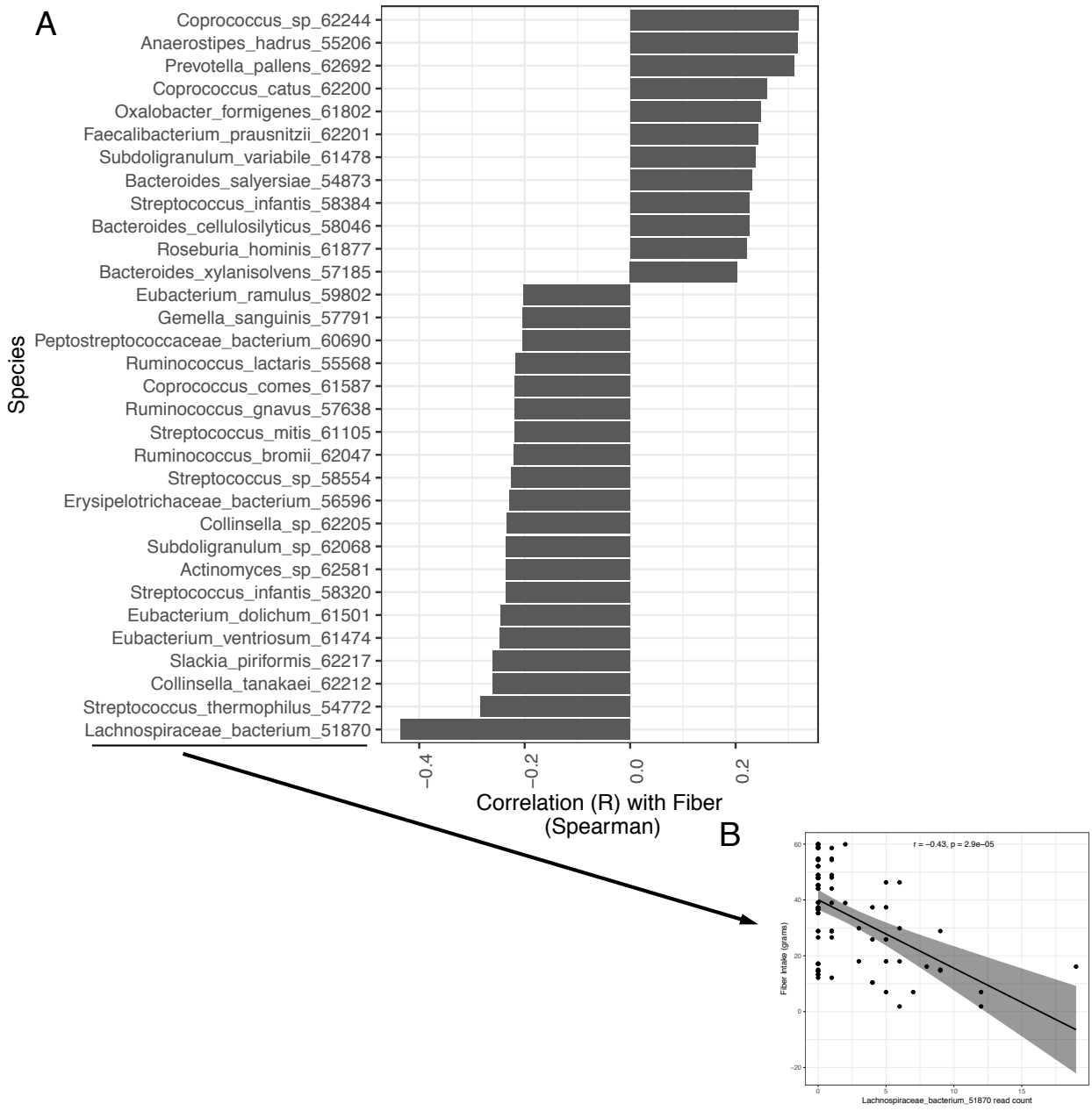


Figure S2.2: A) Correlations above a r > 0.2 cutoff of microbial abundance and fiber intake. Only *Lachnospiraceae* bacterium 51870 was significantly negatively correlated at an FDR cutoff of 0.05. B) Raw spearman correlation of a species of *Lachnospiraceae* with fiber intake.

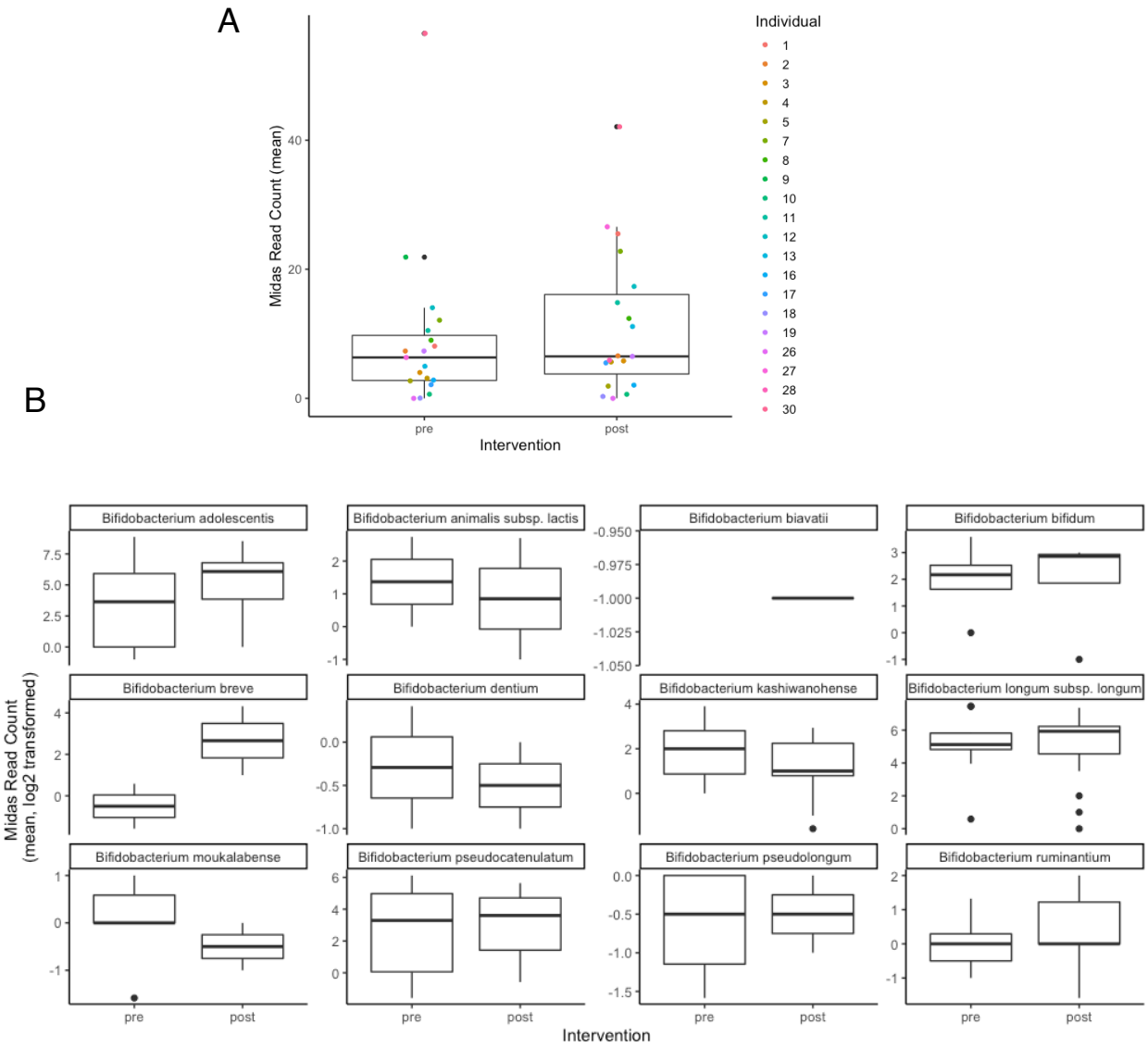


Figure S2.3: A) Mean abundance (MIDAS read counts) of the genus Bifidobacterium during the diet intervention. Points are colored by individual. B) Changes in mean abundance of each species of Bifidobacterium, detected by MIDAS, during the diet intervention period.

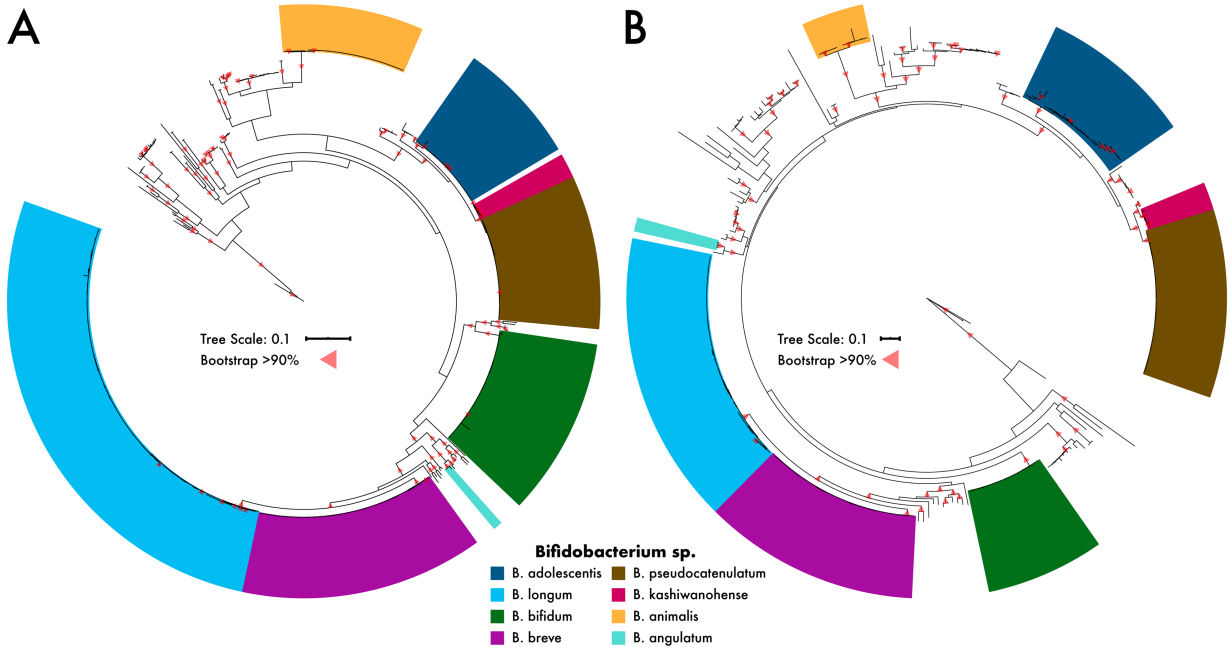


Figure S2.4: *Bifidobacterium* phylogenetic analyses. A) Multi-locus phylogenetic analysis of conserved ribosomal marker genes. B) Phylogenetic analysis of the *groEL* gene sequences used for amplicon analyses. The top 8 species observed are shown.

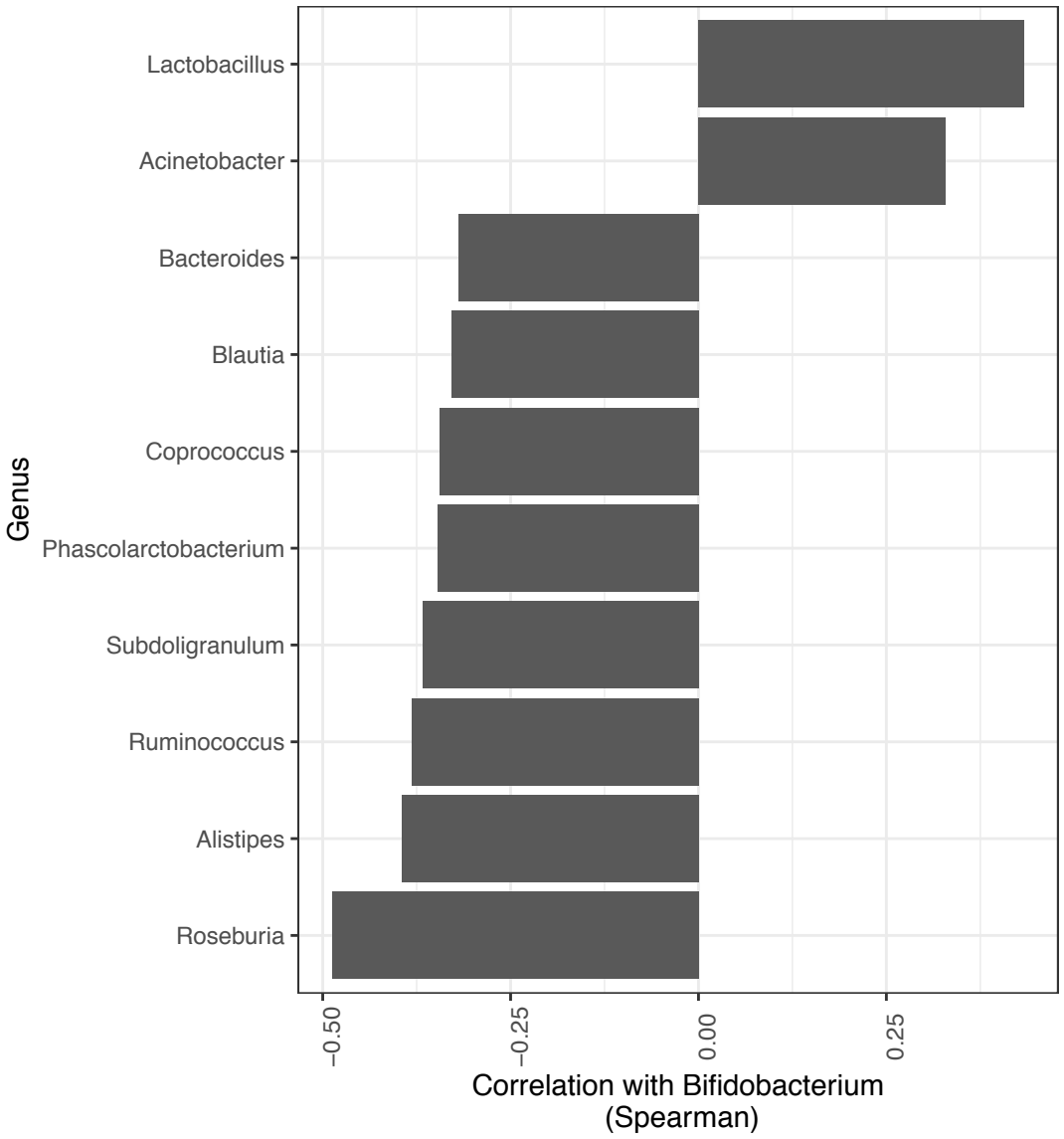
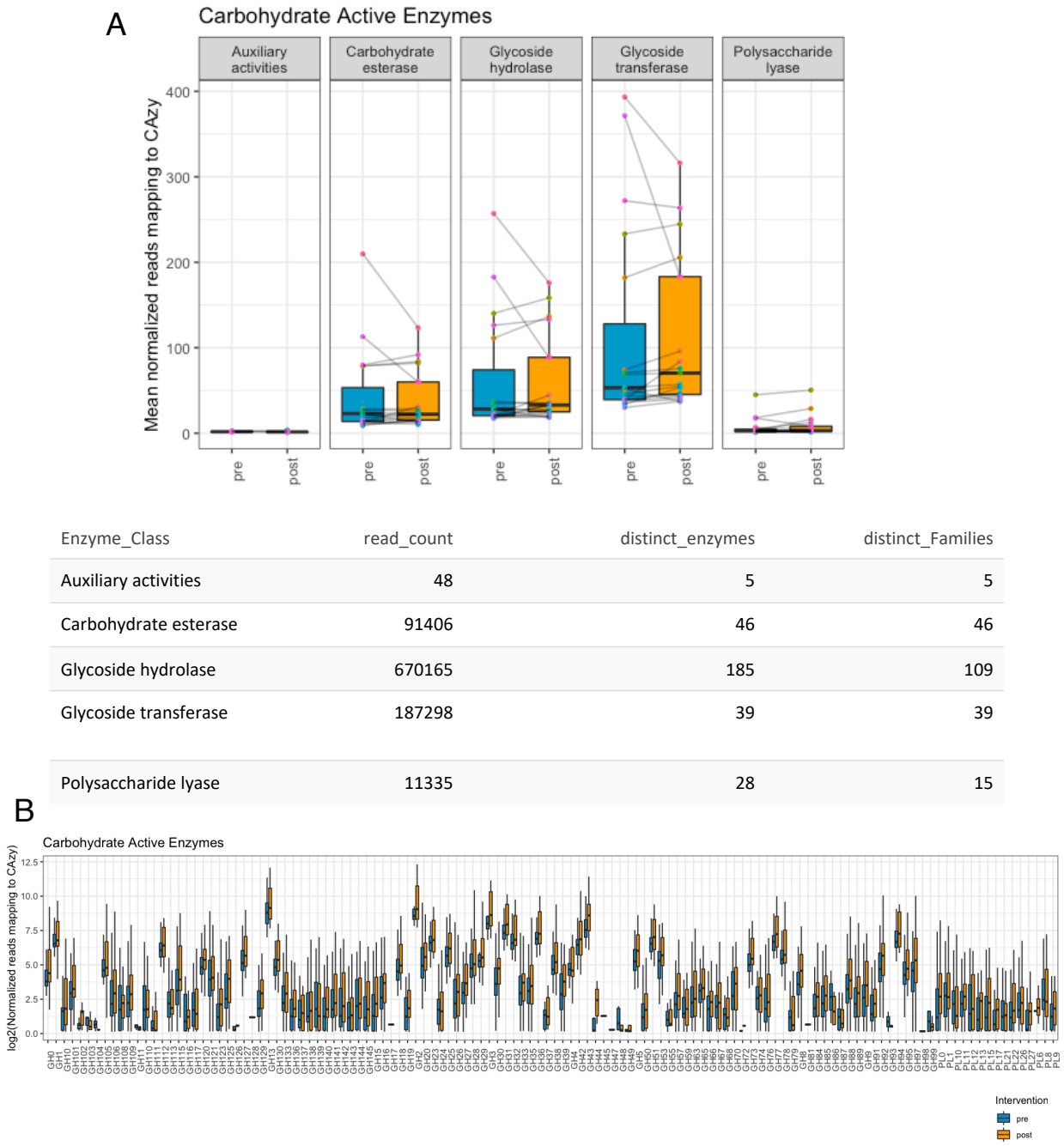


Figure S2.5: Significant (FDR = 0.05) correlations from comparing abundances of 99 different genera with *Bifidobacterium.*



Enzyme_Class	read_count	distinct_enzymes	distinct_Families
Auxiliary activities	48	5	5
Carbohydrate esterase	91406	46	46
Glycoside hydrolase	670165	185	109
Glycoside transferase	187298	39	39
Polysaccharide lyase	11335	28	15

Figure S2.6: A) Average abundances (normalized reads) of carbohydrate active enzymes within individuals during the diet intervention period. Each different color point represents an individual, and lines connect the same individual. B) Log2 transformed GH and polysaccharide lyase gene abundances during the intervention.

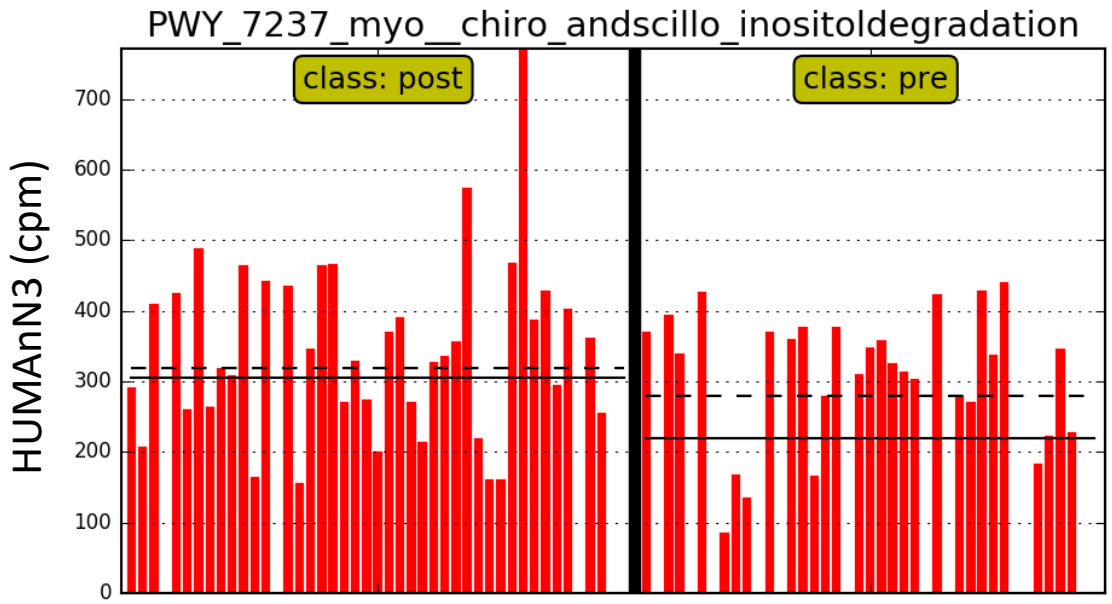
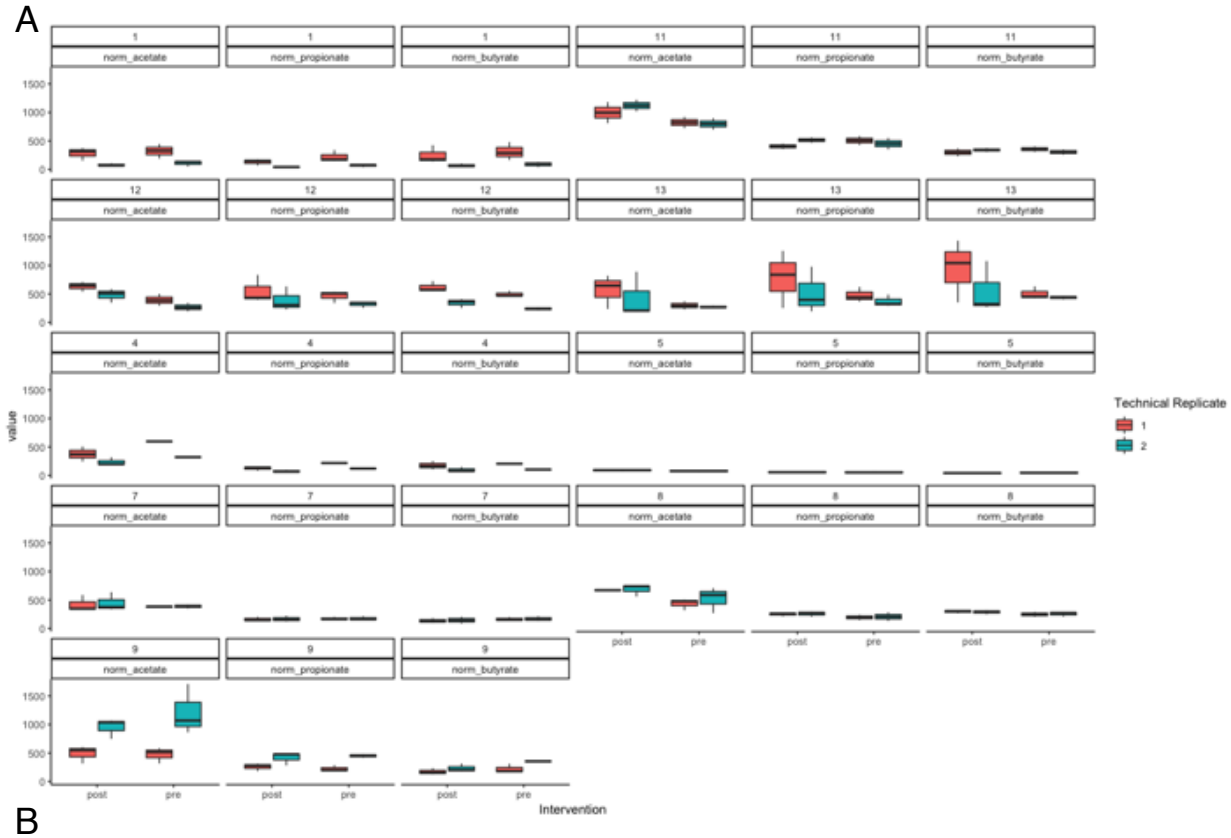


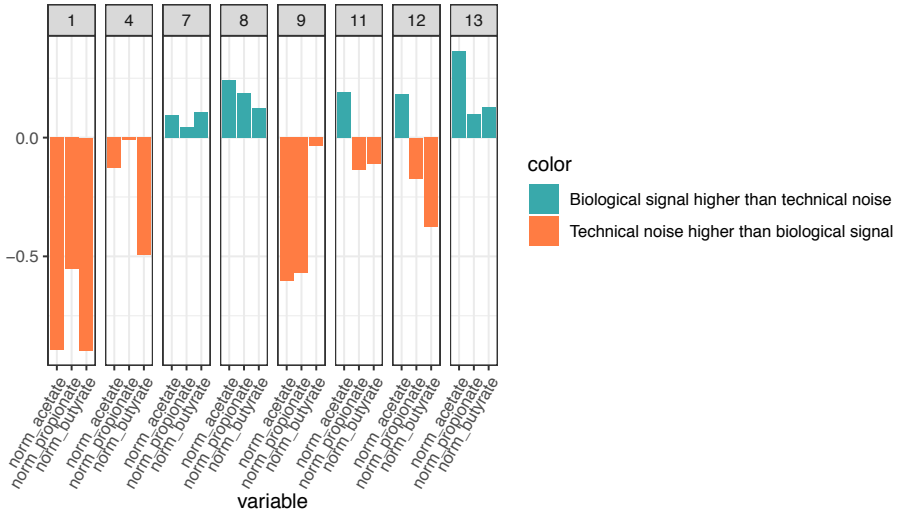
Figure S2.7: Inositol degradation abundance (normalized copies per million), for metagenomes before and after the intervention.

Figure S2.8: Technical variation of SCFAs seen in a subset of samples run in duplicate. A) Amount of SCFAs by individual, with color denoting if it was measured during replicate 1 or 2. B) Normalized (by mean) difference between absolute difference between treatment, subtracted by absolute difference between technical replicates. Larger negative values suggest differences between technical replicates were larger than the differences detected between pre- and post-intervention arms.

A


1 norm_acetate
1 norm_propionate
1 norm_butyrate
11 norm_acetate
11 norm_propionate
11 norm_butyrate
12 norm_acetate
12 norm_propionate
12 norm_butyrate
13 norm_acetate
13 norm_propionate
13 norm_butyrate
4 norm_acetate
4 norm_propionate
4 norm_butyrate
5 norm_acetate
5 norm_propionate
5 norm_butyrate
7 norm_acetate
7 norm_propionate
7 norm_butyrate
8 norm_acetate
8 norm_propionate
8 norm_butyrate
9 norm_acetate
9 norm_propionate
9 norm_butyrate
value
0
500
1000
1500
post
pre
Intervention
Technical Replicate
1
2

B


1
4
7
8
9
11
12
13
0.0
−0.5
norm_acetate
norm_propionate
norm_butyrate
variable
color
Biological signal higher than technical noise
Technical noise higher than biological signal

CHAPTER 3:

Comparisons of gut, saliva, and chicha microbiomes in a remote Amazonian people

Authors: Andrew Oliver, Eric Adams, Alexandria Gille, Nadia Alaniz, Carolina Jamie, Brenda Bowser, John Patton, Katrine Whiteson

**All supplemental tables can be found with the forthcoming publication*

SUMMARY

The Conambo people live in a remote village along the Conambo River, isolated from modern industrialized society. They lack access to modern medicine and sanitation infrastructure and consume diets entirely devoid of processed foods. We sought to characterize the gut and saliva microbiomes of the women and children of the Conambo village, and found high abundances of microbes, such as *Prevotellaecae*, which are anti-correlated with industrialized societies. We also examined the microbiome of "chicha", a cultural variant of cassava beer made by mastication of boiled yuca root and show the community succession of microbes during the fermentation process. We found cohabitation to drive significantly greater similarity within microbiomes for fecal, saliva, and chicha samples. Finally, we contextualized microbiomes from the people of Conambo by comparing fecal microbial composition and functional potential with other industrialized and non-industrialized cohorts, showing that urban samples from multiple sources strongly cluster together.

INTRODUCTION

Human-associated microbiome changes have occurred in tandem with modernization and industrialization of society, which has brought clean water, antibiotics, and processed foods (189). While much work has been performed investigating the role of the microbiome in diseases of industrialized communities, significantly less focus has been put on the microbiomes of non-

industrialized communities. These non-industrialized communities provide critical context to the changes that industrialization have imprinted on the modern human gut microbiome (190). Moreover, as society continues to globalize, habitat loss and the dissolution of rural–urban boundaries may contribute to the loss of societies that contain microbes selected against by industrialization. In the present study we sought to characterize the microbiomes of a remote community of humans, existing far from the influences of modern industrialized society.

The village of Conambo is one of several small, isolated villages along the Conambo River, deep within the Ecuadorian Amazon (**Fig. 3.1A**). Besides the anthropologists who have extensively studied the Conambo people (26, 28, 191) and the occasional missionaries, the village has little outside contact. The small community of just under 200 people is centered around a community building, school, and small dirt airstrip that the people maintain. There are no roads to the village and access by boat is nearly impossible most of the year. As such, a small single propeller plane is the only way to travel to and from Conambo.

The normal diet of the Conambo people includes a stew of available wild game (mostly monkey, tapir, peccary and large rodents), fish and various birds, which is served alongside boiled yuca or plantains (**Fig. 3.1B**). The people of Conambo practice horticultural foraging, where the men hunt, and the women maintain household gardens. No stores exist to purchase processed foods. Distinctive of the Conambo diet is a beverage made from fermented yuca root, known as chicha, which serves as a major source of calories for these people (192). The chicha is made by the maternal head of the household, with the assistance of her daughters, by repeated mastication of the boiled yuca root in a large wooden trough (**Fig. 3.1C**). This unique production process inoculates it with fermentative bacteria from saliva (193). After the mash is prepared, it is covered in banana leaves and allowed to ferment for four days before it is mixed with water and served in

a traditional ceramic bowl (**Fig. 3.1D**). Interestingly, the people almost exclusively drink chicha as opposed to water or other liquids.

This study aimed to characterize the gut and oral microbiomes of women and children living in Conambo, and the role of microorganisms in chicha. Our research asked three questions:

1) What microbial taxa are present in the saliva, fecal and chicha samples collected from the Conambo people? To what degree are fecal and saliva microbiomes shared with chicha microbiomes?
2) Because of the limited contact between households and the household-specific chicha making process, how does cohabitation influence the microbiome?
3) How does the Conambo microbiome compare to human-associated microbiomes in other non-industrial and industrialized societies?

We hypothesized that we would find a high degree of sharing between chicha and oral microbiomes due to the unique pre-mastication process of making this beverage. We did not expect to see a similar degree of sharing between chicha and fecal microbiomes. Further, we anticipated seeing variation in microbiomes to be significantly influenced by cohabitation, due in part to close living quarters and the household-specific chicha production. Finally, we hypothesized that the absence of factors that separate industrialized from non-industrialized societies (e.g., processed foods and access to modern medicine) will structure the microbiome more similarly to other non-industrialized cohorts.

MATERIALS AND METHODS

Sample collection and genomic DNA extraction

The sets of saliva, chicha, and fecal samples were collected in the field by anthropologists as a part of a longer, ongoing ethnographic study. Saliva samples were collected at two intervals,

coinciding with collection of the first and last chicha samples. Chicha samples were collected daily across the four days of the fermentation process. Fecal samples were collected on two different days with collection being performed by the subject after verbal explanation of sampling method. A total of 113 saliva, 108 chicha, and 103 fecal samples were collected in June 2018 and stored using DNA/RNA Shield swab collection tubes (Zymo Research, Cat# R1107). The samples were gathered from a total of 71 individuals in 28 households, encompassing every adult female living in the community as well as their offspring, up to age seven, residing with them. Samples were not collected from adult males in the community. The sample collection tubes were transported to CSU Fullerton and kept in a freezer at -20°C until transfer to UC Irvine for extraction and processing.

Microbial genomic DNA was extracted from each of the samples using the ZymoBIOMICS 96 DNA Kit (Zymo Research, Cat# D4309) according to the manufacturer's instructions. This included 5 x 1-minute bed-beating steps using a FastPrep-24 (MP Biomedicals, Cat# SKU 116004500) at maximum speed. The quantity of DNA in each sample was fluorescently measured with the Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher, Cat# P11496) using a Synergy H1 Microplate reader (BioTek, Cat # BTH1M).

Library preparation and sequencing

Sequence libraries were prepared from the extracted DNA samples using the Nextera DNA Flex Library Prep Kit (Illumina, Cat. # 20018705) following a low volume variation of the standard protocol. Samples were prepared for PCR with Kapa HiFi HotStart ReadyMix (Roche, Cat # 07958935001). All the PCR steps were performed in an Eppendorf Mastercycler Nexus Gradient (Eppendorf, Cat # 2231000665) using the standard thermal cycles as described in the Nextera Flex protocol. The resulting sequence fragments were analyzed on an Agilent

Bioanalyzer to determine fragment length distribution (Agilent, Cat # G2939BA). Sequence libraries were pooled based on DNA concentration as determined by the Bioanalyzer. After pooling, the libraries were sent to Novogene Co., Ltd. for sequencing on an Illumina HiSeq4000.

Metagenomic sequence analysis

Sequence reads were first quality filtered (Q > 30) using Prinseq v0.20.4 (73). Quality filtered sequences were then mapped to references of the human genome (HG28) with Bowtie2 v2.2.7 (74) and matching sequences were removed from further analysis. The remaining sequences were mapped against marker genes for taxonomic assignment using MetaPhlAn v3.0.7 (194). Read counts were extrapolated by multiplying the relative abundance by the number of input reads. Species counts per sample represent the average of 150 subsamples, rarefied to 23,000 sequences per sample using the EcolUtils (v0.1) package in R (195). The rarefaction limit was increased to 100,000 for analysis comparing industrialized to non-industrialized cohorts. Alpha diversity was assessed using the Shannon diversity index in the vegan (145) package in R. Sequence diversity was also assessed using k-mers of size 21 on all metagenomes, subsampled to an even depth (62,597,500 bases/samples). 21-mers have been shown to be highly correlated to sequence diversity (196), and were detected in our data using the kmercountexact.sh script from BBMap. Taxonomy was also assessed using an assembly approach. Briefly, a cross assembly was created using Megahit v1.2.9 (197). Reads were then mapped back to the resulting contigs using Bowtie2 (74) and counted using idxstats from Samtools v1.3 (198). Taxonomy for each contig was assessed using CAT-BAT (199), and a taxa was assigned to the contig if the contig had at least three taxonomically identifiable proteins and 60% of the taxonomy calls per contig were in agreement. Counts per contig for each sample were normalized to the length of the contig (contig coverage) and rarefied to 23,000 counts as

described above. The presence of carbohydrate active enzymes (CAZy) was determined by first subsampling all microbiomes to an even depth (62,597,500 bases/samples) using the reformat.sh script from BBMap (151). Reads that could be merged were concatenated with unmerged reads using bbmerge.sh from BBMap (151). These reads were passed to HMMER (v3.3.2) (200) against CAZyDB (CAZyDB.07312019.fa.nr, (143)), and filtered for hits using the cutoff values coverage = 0.35 , E-value = 1e-15 using the hmmscan_parser.py script from run_dbcan (201). Strain level analysis was done using StrainPhlAn v3.0 using default parameters (194).

Statistical analysis

Comparing alpha diversity between sample types was performed using the Kruskal-Wallis test in R and post-hoc Dunn's test. Bray Curtis comparisons of fecal and salivary microbiomes to the chicha microbiome were done using a Kruskal-Wallis test, and a linear mixed-effects model (LME) (using household as a random effect) was used to test chicha centroid comparisons. Additionally, LMEs were used to test the effect of cohabitation on average bray distances. Permutated multivariate analysis of variance (PERMANOVA), from the adonis function in the Vegan package (145), was used to analyze the significance of group membership by detecting differences in group centroid location. We utilized this approach for both community and strain-level analysis. Ordination of microbiome communities was done using the metaMDS function in Vegan. Taxa that explained differences between industrial and non-industrial microbiomes were identified using a permutated random forest (81), and vectorized to NMDS space using the envfit() function in Vegan. Phylogenetic trees from StrainPhlAn analysis was visualized using the ggtree package in R (202).

Data availability

All DNA sequences from the Conambo cohort will be found under a BioProject # at the time of publication. Code used in the analysis can be found on GitHub. Cohorts used to compare the against the Conambo were obtained from previously published work (19, 22, 203–206), and accession IDs of utilized samples can be found in **Supplemental Table 3.1**.

RESULTS

Cohort and study design

The Conambo cohort consisted of 43 children and 28 mothers from 28 different households. Sequencing of the prepared libraries yielded 436,402,509 raw paired end reads. After quality filtering and removing reads that mapped to the human genome, 287 samples were retained and each sample contained an average of 1,354,451 ± 600,423 paired end reads (**Supplemental Table 3.2**).

Streptococcus was shared between saliva and chicha microbiomes, while *Prevotella* dominated Conambo gut microbiomes

To examine the microbiomes of the Conambo people, we analyzed 287 metagenomes from 71 individuals, from samples of feces (n = 101), saliva (n = 85) and chicha (n = 101). Pairwise comparisons using Dunn's test showed alpha diversity, measured using the Shannon diversity index, was significantly lower in chicha when compared to saliva ($p < 0.01$) or fecal samples ($p < 0.001$) (**Figure 3.2A**). Because samples from the Conambo people may contain organisms that are underrepresented in current databases and difficult to classify using reads alone, we also assessed diversity using an assembly-based method. We also measured alpha diversity using assemblies and were able to detect more distinct species compared to the marker gene approach (**Supplemental Figure 3.1**).

To test the hypothesis that the microbiomes of chicha would overlap more with saliva than gut samples, we measured the average Bray Curtis dissimilarities between these sample types. The average Bray Curtis dissimilarity was significantly smaller between chicha and saliva (0.916) when compared to chicha and gut samples (0.981) (Kruskal-Wallis test, $p < 0.001$) (**Figure 3.2B**). This was true even when using the first day of chicha production for the comparison, since chicha production at this time has the greatest opportunity to be inoculated with human microbes from the production process (data not shown).

We next examined the taxa identified within the Conambo samples **(Figure 3.2C-D, Supplemental Figure 3.2**). We assessed taxonomy separately using assemblies (**Supplemental Figure 3)** and short reads; the assemblies were more sensitive to rarer taxa due in part to a lower threshold for taxonomic assignment (a minimum of 3 proteins per contig; **Supplemental Figure 3.4**). Nearly 96% of our taxonomically assigned reads from gut samples were bacterial in origin. The remaining reads mapped mostly to viruses and, to a lesser extent, archaea and eukaryotes. The gut microbiomes of the Conambo people were dominated by the family *Prevotellaceae* (**Figure 3.2C**), which accounted for 36% (± 19) of the taxonomically assigned reads in the average fecal metagenome. The most abundant species detected in this family was *Prevotella copri* (mean relative abundance = 22 ± 15%). *P. copri* was ubiquitous as well, having the second lowest coefficient of variation among species found in fecal samples (70.2%), behind *Faecalibacterium prausnitzii* (54.8%). Taxa from the families *Ruminococcaceae, Lachnospiraceae,* and *Bifidobacteriaceae* were also highly abundant in the gut microbiomes of the Conambo people, together making up ~34% of the identifiable taxa in the gut. Non-bacterial, less abundant members of the gut microbiome included *Saccharomyces cerevisiae* and *Blastocystis* subtype 1. Methanogens such as *Methanobrevibacter smithii* and *Methanosphaera*

stadtmanae were found in the archaeal fraction of the microbiome. Of the classified viral fraction in the gut, most reads were assigned to the phage families Siphoviridae (0.02%) and Myoviridae (0.004%).

The oral microbiomes of the Conambo people had abundant *Streptococcaceae*, with the average saliva sample having 25% (± 21) of taxonomically assigned reads mapping to this family (**Figure 3.2D)**. Of the reads mapping to *Streptococcaceae,* 75% were assigned to either *S. mitis* or *S. salivarius*. Notably, the presence of *S. mitis* or *S. salivarius* dominated over the other *Streptococcus* sp., and rarely coexisted at similar abundance levels (**Supplemental Figure 3.5A**). We found that the ratio between *S. salivarius* to *S. mitis* was inverted between mother and children (**Supplemental Figure 3.5B**); moreover, *S. mitis* had significantly higher abundances in children than mothers (Wilcoxon, $p < 0.01$). Other abundant species found in saliva included *Neisseria flavescens* (mean abundance 7.5%) and *Rothia mucilaginosa* (mean abundance 7.4%). Like the gut microbiome, we also detected *P. copri* within the top five most abundant species in saliva. Viruses made up a large part of the detected taxa (8.9%), with the most abundant being an endogenous retrovirus and various herpes viruses.

The microbiomes of chicha fluctuated in a time-dependent manner during fermentation (**Figure 3.2E**). Early in fermentation, the typical chicha microbiome had a high abundance of *Streptococcus* and, to a lesser extent, *Lactobacillus*. Specifically, *S. salivarius* and *L. fermentum* were the majority members of the chicha microbiome, making up 20% (± 17) and 38% (± 22) relative abundance, respectively. Sometime between the first and second day of fermentation, *L. fermentum* increased in abundance and remained the majority member of the chicha microbiome for the remainder of fermentation. *S. cerevisiae* appeared to increase in abundance during the third day of fermentation. During the first three days of fermentation, the chicha microbiomes

significantly converged (LME, $p < 0.001$), adopting a similar microbiome irrespective of household origin (**Figure 3.2F**).

Cohabitation drives similarity between the Conambo microbiomes

The microbiomes from gut, saliva, and chicha samples all reflected significant effects of cohabitation on microbial community composition (**Figure 3.3A**). Furthermore, within each household, saliva and gut microbiomes appeared to be significantly personalized (PERMANOVA, nested effect of individual, $p < 0.001$). Cohabitation appeared to have the greatest effect on structuring the chicha microbiome, which was reflected in significant differences between average Bray Curtis distances between microbiomes from cohabiting and not cohabiting individuals (**Figure 3.3B**).

In addition to microbiome community differences, we also investigated whether there was significant strain-level variation between different households. One species of interest, *S. salivarius*, was an abundant member of the saliva and chicha microbiomes (**Figure 3.2**) and had enough coverage to examine strain dynamics between eight households in the community. We showed that *S. salivarius* strains differed between households but not between saliva and chicha samples (after accounting for household strain differences) (**Figure 3.3C**). This suggests that chicha may support the transmission of *S. salivarius* within households in the context of cohabitation in Conambo.

Similarly, we analyzed strain similarities at the household and individual levels for *P. copri*, which was abundant in the fecal and oral microbiomes. We found strains of *P. copri* were highly individual specific (PERMANOVA, nested effect of individual, R^2: 0.57, $p < 0.001$; **Figure 3.3D**). Moreover, cohabitating explained a large portion of the variation of the single nucleotide polymorphism profiles between *P. copri* strains (PERMANOVA, Household, R^2:

0.35, $p < 0.001$). These data suggest cohabitation structured not only the gut microbial community at large, but also a significant portion of the variability found at the strain level.

The human microbiomes from the Conambo cohort are compositionally similar to other non-industrial cohorts and comparatively reflect a high degree of diversity in carbohydrate break down genes

To put the Conambo gut microbiome in the context of other industrialized and non-industrialized communities, we compared microbial taxonomic composition and function between the Conambo cohort and seven other studies (19, 22, 203–206). Except when compared to the Hadza, the Conambo gut microbiomes contained the highest fraction of unknown reads among these eight cohorts (**Supplemental Figure 3.6**). The Conambo cohort clustered with the microbiomes from other non-industrialized cohorts, and together these microbiomes were distinct from more urbanized microbiomes (PERMANOVA, $R^2 = 0.15$, $p < 0.01$) (**Figure 3.4A**). A random forest analysis identified taxa that helped differentiate non-industrial samples from industrialized samples; these taxa included the genera *Prevotella, Catenibacterium* and *Treponema.* More industrial-associated taxa included *Anaerostipes, Alistipes,* and *Bacteroides*. Although the Conambo and Tanzania cohorts (22) were compositionally indistinguishable from one another (pairwise PERMANOVA $R^2 = 0.034$, $p = 0.18$), we found significant differences between the microbiome compositions of the Conombo and the Peruvian Lewis (19) cohorts (pairwise PERMANOVA $R^2 = 0.067$, p.adj (Bonferroni) < 0.05). Similarly, pairwise comparisons between the Conambo microbiomes and the industrial cohorts revealed significant differences, with the explained variation between industrialized cohorts and the Conambo microbiome ranging from 7 – 22% (pairwise PERMANOVA, p.adj (Bonferroni) < 0.05).

The number of distinct species identified within the cohorts varied considerably, with Conambo having the lowest richness on average. Because this might reflect poor mapping to the MetaPhlAn3 database, we also analyzed richness of k-mers (k=21bp) (**Figure 3.4B**). We observed the largest amount of k-mer diversity within the Conambo, suggesting these microbiomes contain a significant amount of diversity which goes undetected using database-dependent methods. We next asked whether microbiome diversity extended to differences in functional potential between the industrialized and non-industrialized cohorts. We focused on enzymes involved in the breakdown of carbohydrates (CAZymes), specifically glycoside hydrolases and polysaccharide lyases, since the diversity of these enzymes may be influenced by diet. After controlling for sequencing depth, we found extensive CAZyme richness within the Conambo cohort that was significantly higher than all other cohorts examined (Dunn test, $p < 0.05$, **Figure 3.4B**). The cohorts were also compositionally different in their carbohydrate breakdown genes (PERMANOVA, Lifestyle/cohort nested effect, $R^2 = 0.71$, $p < 0.01$, **Supplemental Figure 3.7**). Pairwise analysis of CAZyme composition between the cohorts showed that the Conambo cohort was significantly different from all other cohorts analyzed (pairwise PERMANOVA, p.adj (Bonferroni) < 0.05).

We also sought to contextualize strain-level variability within the Conambo cohort. Specifically, *P. copri* diversity is distributed amongst four clades, each of which are found worldwide (206). Analysis of single nucleotide polymorphisms (SNPs) between strains of *P. copri* found Conambo microbiomes with strains from the four proposed clades showed spatial clustering of Conambo with Clade C in ordination (**Figure 3.4C**). PERMANOVA analysis showed that clade membership accounted for most of the variation in SNP profiles between these

strains (PERMANOVA, $R^2 = 0.76$, $p < 0.01$). Notably, *P. copri* strains from Conambo showed very little dispersion in ordination space compared to strains from known clades.

Finally, we conducted a similar strain level analysis on strains of *S. salivarius* and compared these strains to those found in samples of chicha produced by anthropologists (samples were not included in previous analyses involving chicha) living among the Conambo (**Figure 3.4D**). Ordination of SNP profiles showed that the anthropologist-produced chicha harbored different strains of *S. salivarius* compared to all strains identified in the Conambo people. This is consistent with the ideas that *S. salivarius* found in chicha is oral-derived and these strains are distinct to the Conambo people.

DISCUSSION

In the present work, we sought to characterize the microbiomes from a remote community of individuals living in the Ecuadorian Amazon. The Conambo people have little to no contact with urban centers or individuals from industrialized locales. Moreover, substantial distance and dense jungle separate households from one another, such that contact between different households is not commonplace. These people have no regular access to modern medicine or clean water infrastructure. Their diets are structured by local game, gardening, and chicha. These factors likely drive unique microbiome compositions within Conambo. We investigated the structure and function of their microbiomes and contextualized these communities using microbiomes from previous characterized individuals from industrialized and non-industrialized cohorts. Finally, we characterized the microbiomes of chicha, a drink that is both culturally and calorically central to the people of Conambo.

The gut microbiomes of the Conambo were diverse (**Figure 3.2A, Supplemental Figure 3.1**) and dense with microbial DNA whose sequences were not homologous to sequences in

current databases (**Figure 3.2C, Supplemental Figure 3.6**). We were able to detect far more organisms using assembly-based taxonomic assignment compared to short-read based methods, which were similar to species richness found in another rural Ecuadorian cohort (207). Many of the taxa identified within the gut microbiomes of these people belonged to a group of organisms commonly referred to as VANISH taxa (Volatile and/or Associated Negatively with Industrialized Societies of Humans) (208). Known VANISH taxa detected in the Conambo cohort included *Treponema* and *Succinivibrionaceae. Elusimicrobia* sp., an enigmatic organism previously associated with insect guts and the guts of non-industrialized humans (209, 210), was also detected in the assemblies. While relatively little is known about this genus, we speculate that its presence might be explained by the Conambo people's high level of insect consumption.

Notably, we detected an abundance of reads mapping to the family *Prevotellaceae* (**Figure 3.2C**). Across many cohorts, taxa within the family *Prevotellaceae* have been shown to associate with the gut microbiomes of individuals living non-industrialized lifestyles (19, 23, 211–214). *Prevotellaceae* are particularly specialized at breaking down fiber to short-chain fatty acids (215), and the Conambo people's diet rich in chicha and garden vegetables provide highly desirable substrates for gut-living *Prevotella*. Specifically, *P. copri* made up a substantial portion (~26%) of the taxonomically assigned reads. A variety of beneficial and detrimental health outcomes have been linked to the abundance of *P. copri*; for example, improved glucose tolerance (216) and association to inflammatory disorders (217). While there is no clear consensus as to the role of *P. copri* in health, the species appears to be widespread and phylogenetically diverse, associating more strongly with the gut microbiomes of non-industrialized communities (206). Four distinct clades have been found to make up the *Prevotella copri* Complex, each with distinct functional profiles (206). Strains from the

Conambo cohort appeared to form a tight group that loosely clustered with Clade C in ordination (**Figure 3.4C**). The tight grouping of these strains from Conambo suggests that perhaps a factor specific to the Conambo (e.g., diet or environment) or geographical isolation causes them to cluster away from previously sequenced strains. Within Conambo, these strains are highly household and individual specific (**Figure 3.3C**). This is concordant with other research in a closed cohort (little to no outside contact) showing greater similarities of strains within households (24).

Analogous to *Prevotella* in the gut, we examined the microbiome composition of saliva in Conambo and found a high abundance of *Streptococcus* spp. present (**Figure 3.2D**). When we analyzed the species that contributed to this signal, an interesting pattern emerged: the two predominant species, *S. salivarius* and *S. mitis*, rarely co-occurred. Because members of the *S. mitis* group appeared to associate more with the oral microbiomes of infants (218), we investigated whether this signal could be attributed to mother–child origin. Our results indicated that the abundance of *S. mitis* is significantly higher in the oral microbiomes of Conambo children compared to mothers (**Supplemental Figure 3.5**). We suspect intra-household sharing of chicha, which has a high abundance of *S. salivarius* (**Figure 3.2E**), reduced large differences between mother and child. Indeed, Conambo mothers begin feeding pre-fermented chicha pulp during infancy, which may promote transmission of *S. salivarius* from mother to offspring; a similar route of transmission has been speculated before in a study of an indigenous chicha-consuming Bolivian cohort (219). This is further supported by a significant clustering of *S. salivarius* at the household level, but no differences were observed in strains between chicha and saliva within households (**Figure 3.3C**). This may suggest that dispersal of *S. salivarius* occurs largely within households. While our study was not designed to detect transmission of strains

between household, we suspect chicha may contribute to the sharing of oral *Streptococcus* between mothers and children.

A major source of calories and hydration for the Conambo people comes from chicha. Indeed, the Conambo will preferentially drink chicha over water, and the men of the community can drink up to four gallons of chicha per day (Dr. John Patton, personal correspondence). Much work has been done describing the microbial and physiochemical properties of chicha from indigenous populations (193, 219–223). A study investigating chewed cassava beer from Ecuador was able to culture *S. salivarius* and *L. fermentum* (193), concordant with our data showing a high amount of reads mapping to these organisms. Our results indicated compositional succession over time, whereby oral streptococci began the early fermentation and were quickly followed by *L. fermentum* and *S. cerevisiae,* which was likely environmentally sourced (193) (**Figure 3.2E**). The strengths of our study were threefold, specifically: 1) we obtained samples from four timepoints during the chicha fermentation process, 2) these chicha samples came from 27 households, along with paired saliva samples from the women who prepared the chicha, and 3) we utilized shotgun sequencing for a broader look at taxonomy present compared to amplicon based approaches.

Overall, the Conambo microbiome distinctly reflects a non-industrial lifestyle. The presence of the genus *Prevotella* was highly indicative of the Conambo and other non-industrialized cohorts measured (**Figure 3.4A**). A high abundance *Prevotella* has regularly been detected in non-industrial cohorts, especially those with hunter–gatherer and horticultural lifestyles (224). It remains difficult to accurately access alpha-diversity of non-industrial microbiomes due to poor representation of these microbiomes in taxonomic databases (**Supplemental Figure 3.6**). We show that a database-dependent method (specifically

MetaPhlAn3) was able to resolve beta-diversity sample relationships sensibly (see **Figure 3.3A, Figure 3.4A**), but suggested greater species richness for many of the industrialized cohorts when compared to the Conambo (**Figure 3.4B**). To overcome this, k-mers can be used to describe the diversity of the entire sequence space, taking into account reads that would not currently map to databases (196, 225). Analysis of 21 bp k-mers across all eight cohorts revealed a high degree of sequence diversity in the Conambo not reflected in the database-dependent approach (**Figure 3.4B**). These k-mer results mirrored the diversity of CAZymes we measured within the cohorts analyzed. A high diversity of carbohydrate active enzymes have been observed in both ancient and non-industrialized samples (226), which might reflect diets high in fiber-rich foods. This is further supported by the observation that microbiomes of rural agriculturists contain a high abundance of genes involved in short-chain fatty acid production (227), which are created as byproducts of dietary fiber fermentation.

Recent attention has been given to the archiving of microbial strains missing or underrepresented in current databases and repositories (228, 229). One of the largest efforts to catalog gut microbial biodiversity has been the Global Microbiome Conservancy (228), which has isolated and sequenced nearly 6,000 bacterial strains to date. A strong focus of their work and others has been the characterization of microbiomes associated with non-industrial living. These efforts are critical; in a 2018 report from the United Nations it was projected that the percent of the world's population living in urban areas would increase 13% (to 68%) by 2050 (230). With the impending expansion of industrial living, microbial biodiversity and human culture associated with non-industrial living will be lost. It has been suggested that efforts to archive should be secondary to efforts to preserve (228), and we agree. While an effort of

cataloging microbiome diversity should continue, to the extent that is possible, scientists should contribute to the preservation of societies still untouched by globalization.

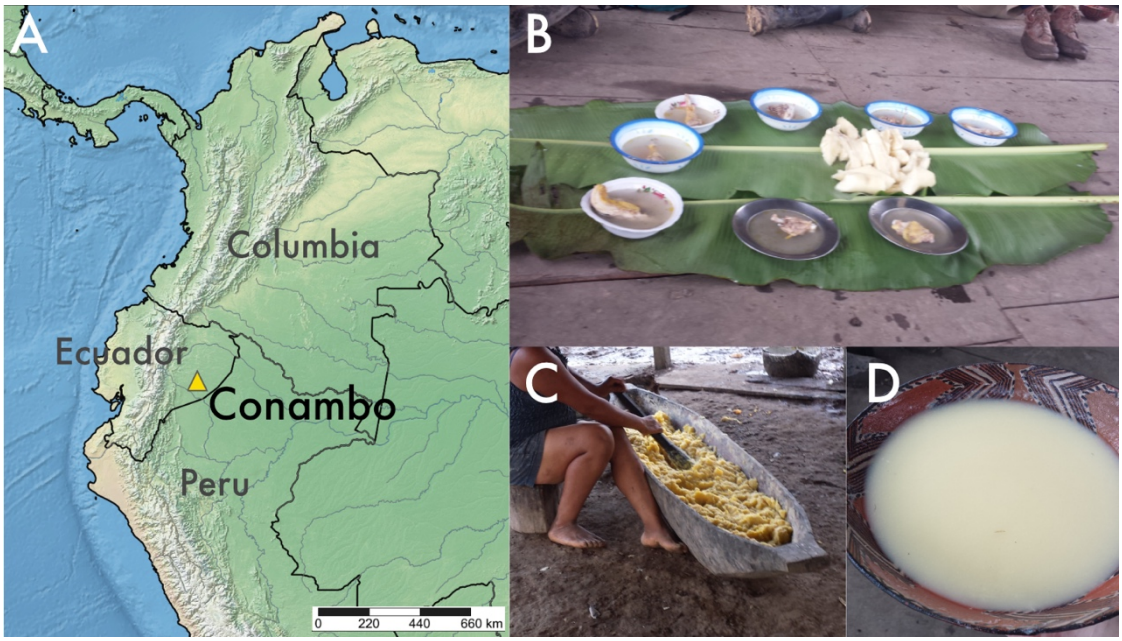


Figure 3.1: Description of the Conambo. (A) Location of Conambo (yellow triangle) within the borders of Ecuador (map generated using simplemappr.net). (B) A meal set out for visiting guests consisting of stewed meat and boiled plantains that is representative of the normal dietary intake. (C) Traditional preparation method for the yucca mash which is fermented to produce chicha. (D) A serving of chicha in a traditional clay bowl made in the village.

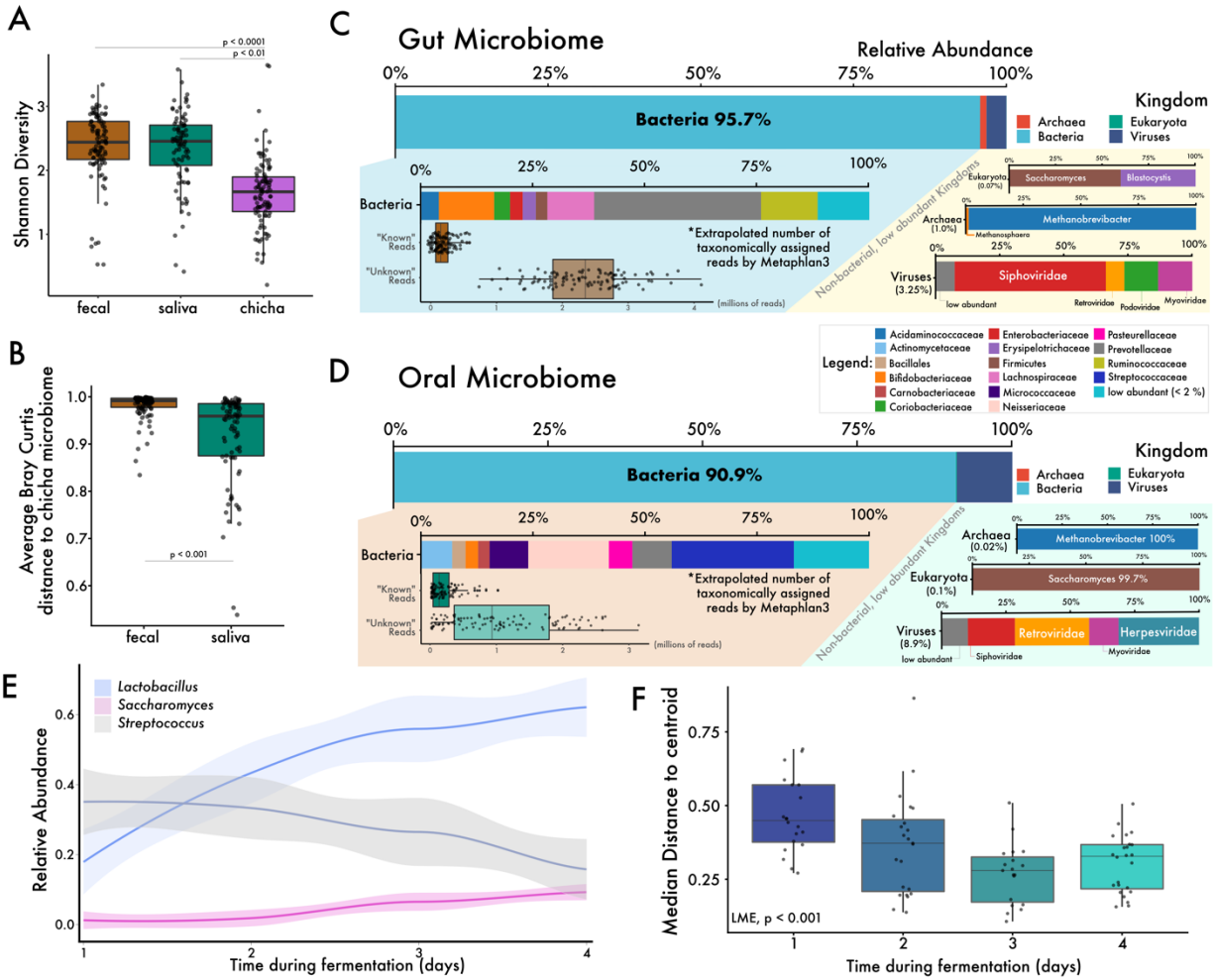


Figure 3.2: The microbiomes of the Conambo people. (A) The alpha diversity, measured using the Shannon diversity index, was significantly lower in chicha samples compared to oral and gut samples (Kruskal-Wallis, $p < 0.01$). (B) The average Bray Curtis distance to the chicha microbiome showed higher community similarity between oral and chicha microbiomes compared to gut and chicha microbiomes (Kruskal-Wallis, $p < 0.01$). (C & D) Average taxonomic composition found in gut (C) and oral (D) samples. (E) The chicha microbiome underwent a succession of community composition, where *Streptococcus* was initially the majority member of the community then decreased in abundance over time as *Lactobacillus* increased through timepoint four (standard error of the mean represented around the lines). (F) Over the course of fermentation, the chicha microbiomes significantly converged.

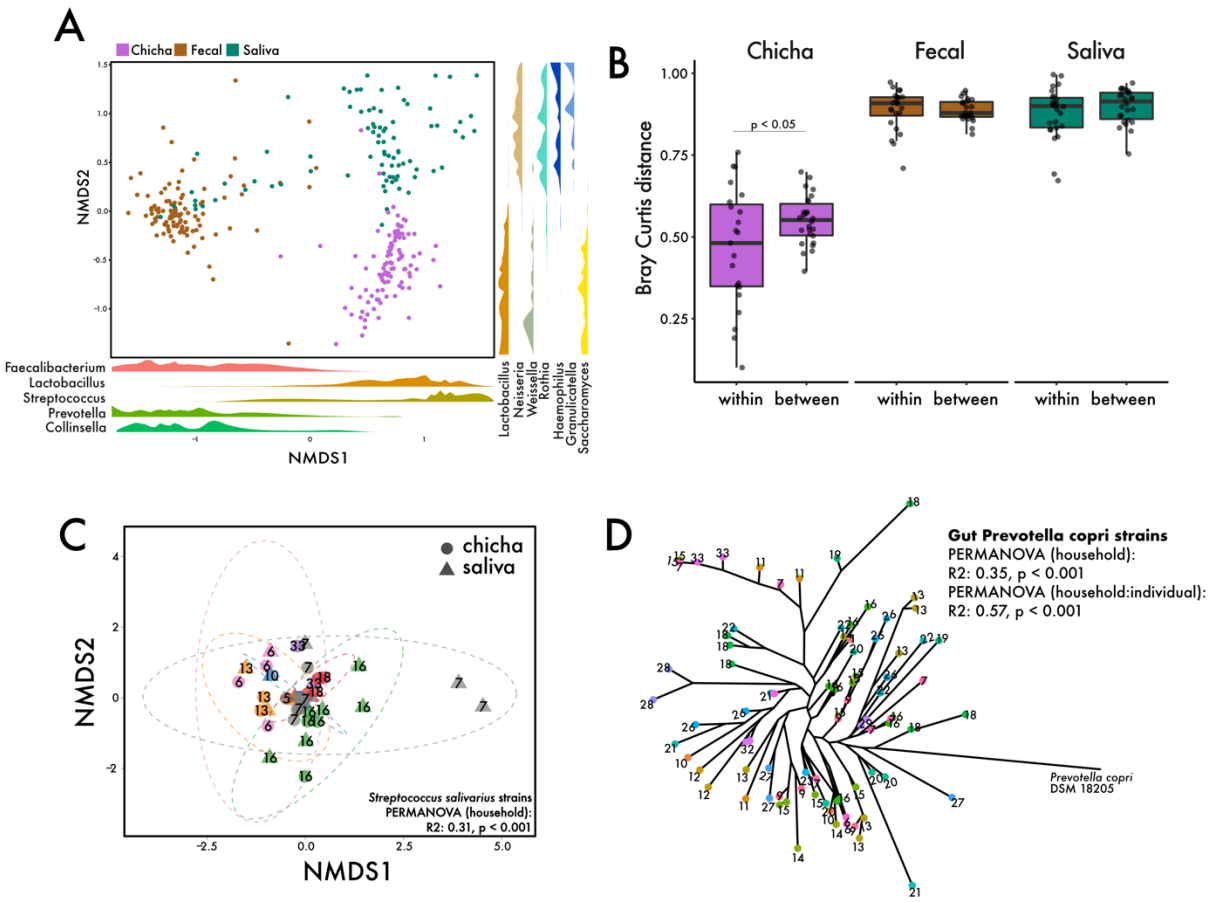


Figure 3.3: The influence of cohabitation on the Conambo microbiome. (A) Ordination of fecal, chicha, and saliva samples from the Conambo, with taxa distinguishing these samples plotted in the margins. (B) We found a significant difference in the Bray Curtis pairwise distances between samples of chicha (LME, $p < 0.05$) made in the same household compared to chicha from other households. (C) Strains of *Streptococcus salivarius* reflected a significant household signature; however, we detected no difference between strains derived from chicha or saliva within the households. (D) An unrooted phylogenetic tree of gut *Prevotella copri* strains revealed these strains were highly individual specific and reflected a significant household signature.

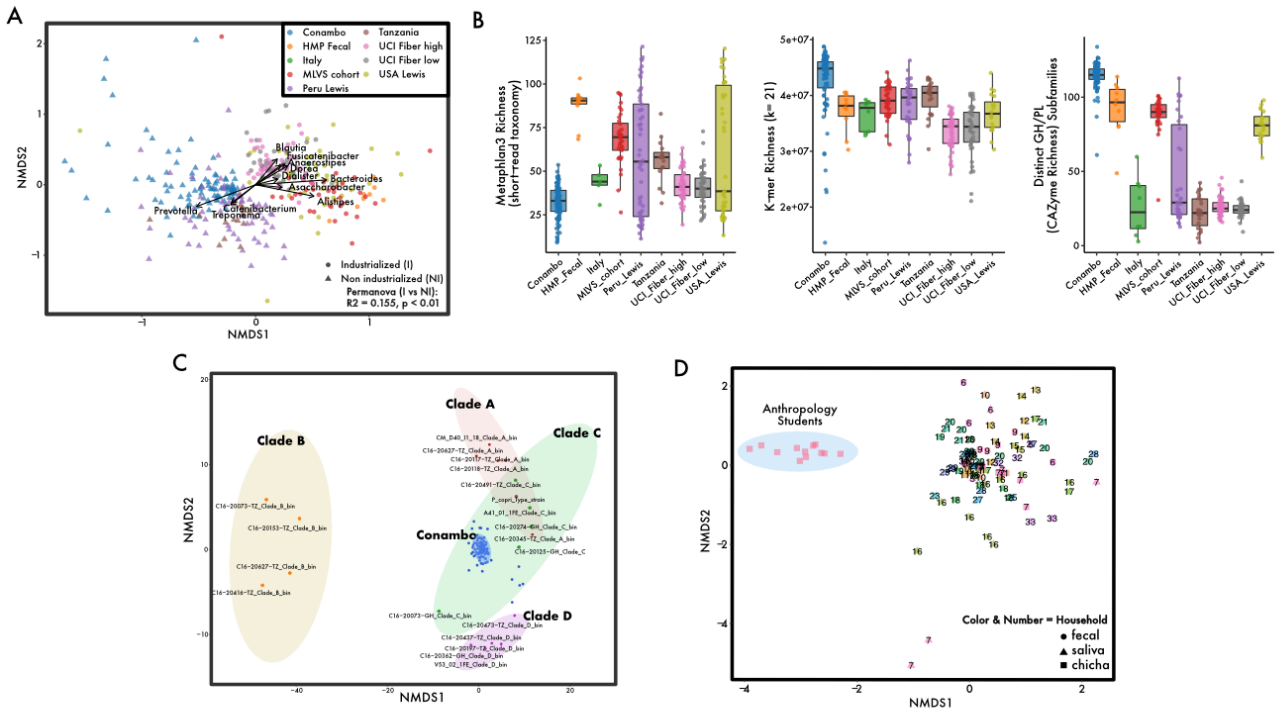
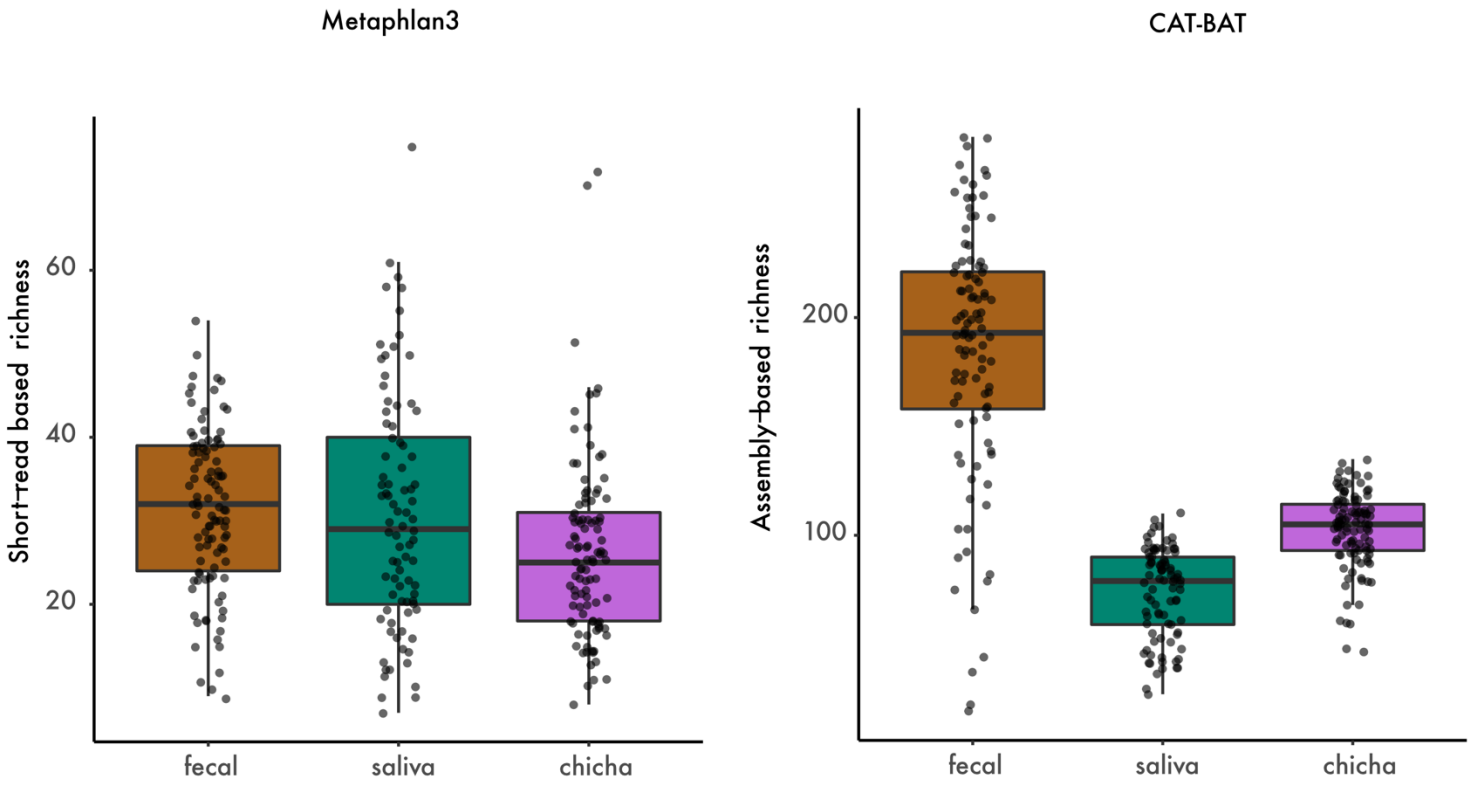
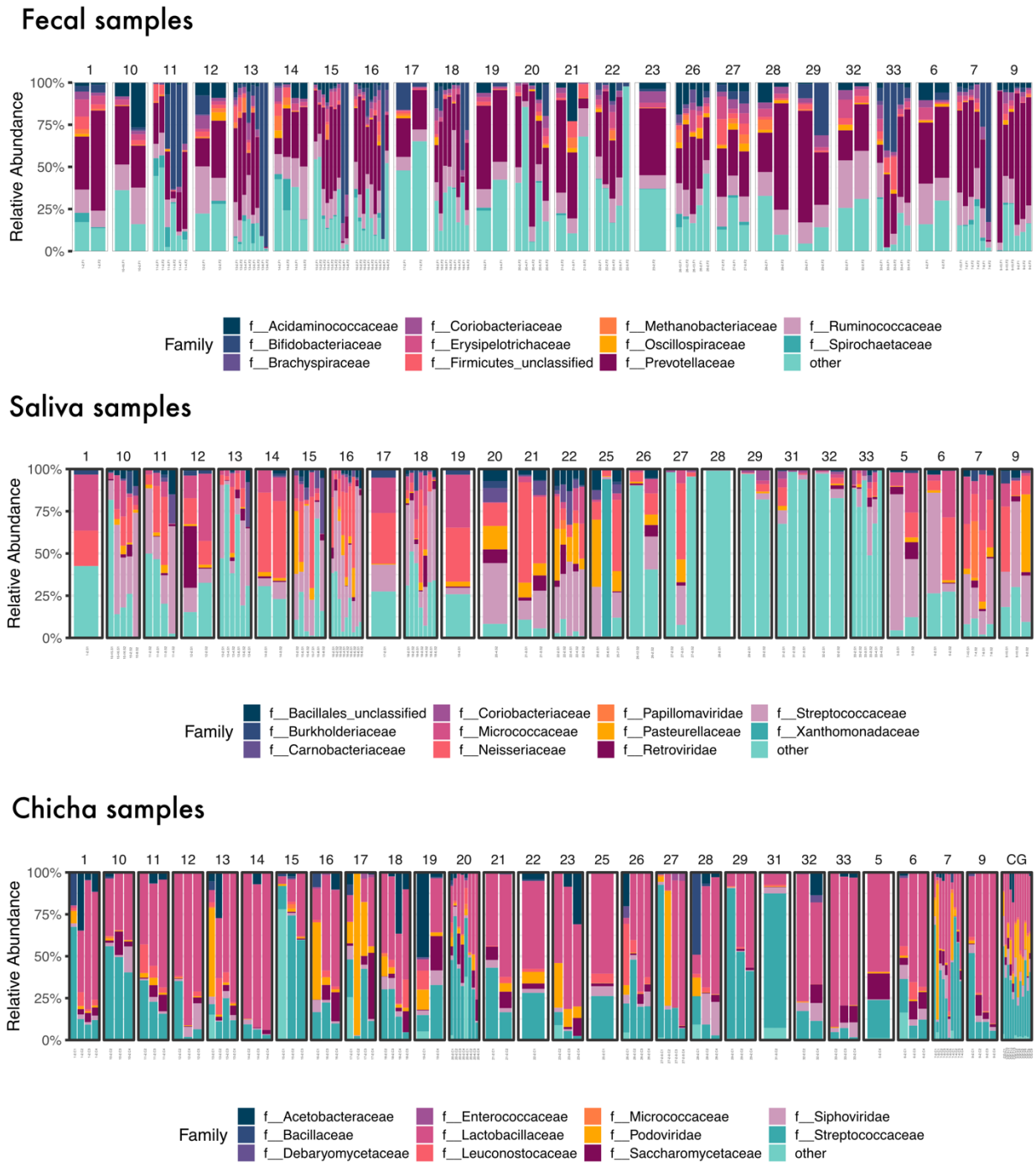


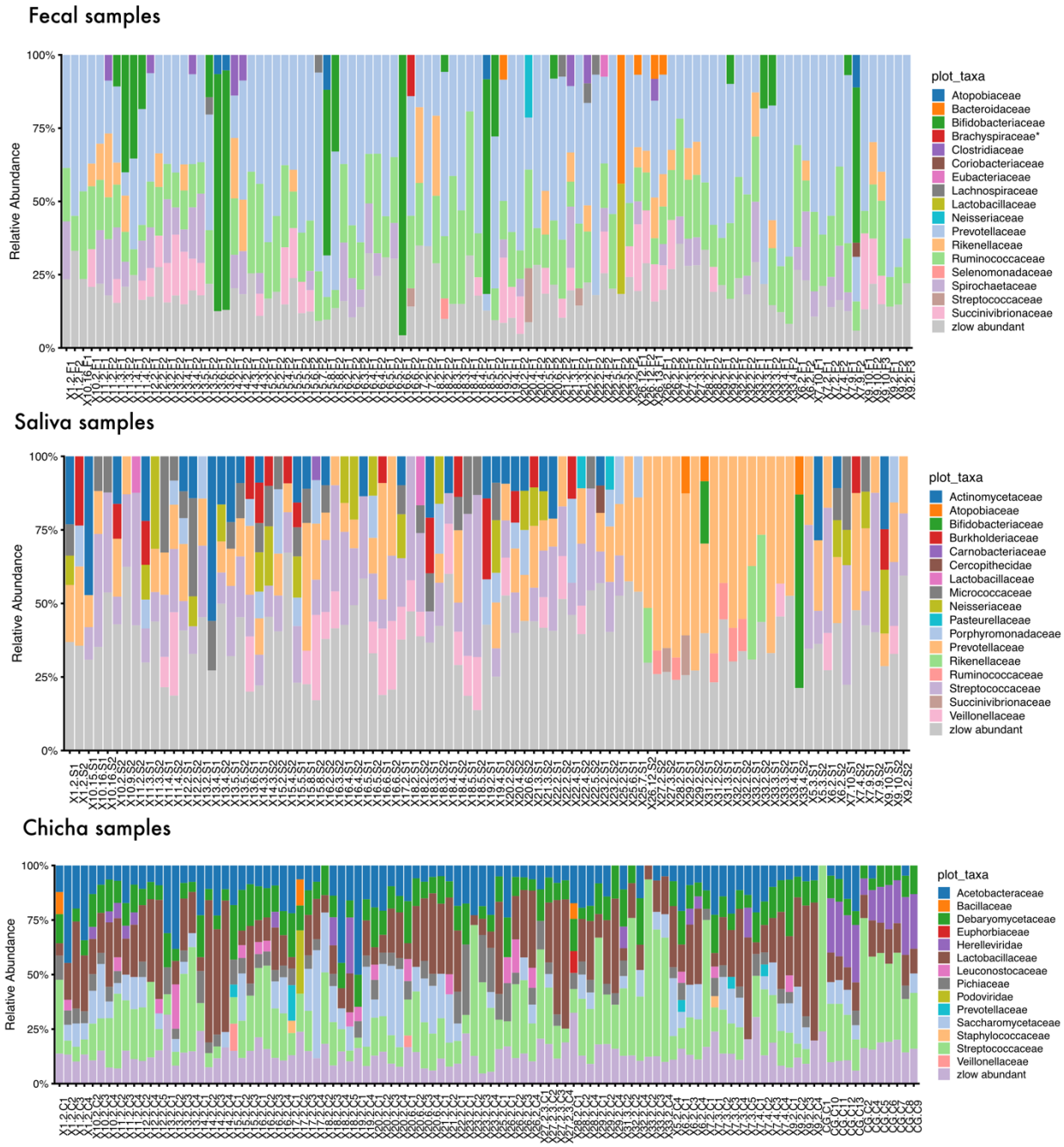
Figure 3.4: Composition and function of Conambo microbiomes in the context of industrialized and non-industrialized microbiomes. (A) The Conambo microbiomes clustered with other non-industrialized cohorts, and taxa contributing to these differences (indicated by vectors) included *Prevotella* and *Treponema*. (B) Richness (number of distinct species), k-mer (k=21) and CAZyme diversity of the Conambo gut microbiomes compared to other cohorts analyzed. (C) Strain level analysis of *P. copri* showed a loose clustering with Clade C of the *Prevotella copri*-Complex and very little spread in ordination space. (D) Similarly, strains of *S. salivarius* grouped together based on household (as previously shown) but were distinct from strains recovered from anthropologists living with the Conambo.



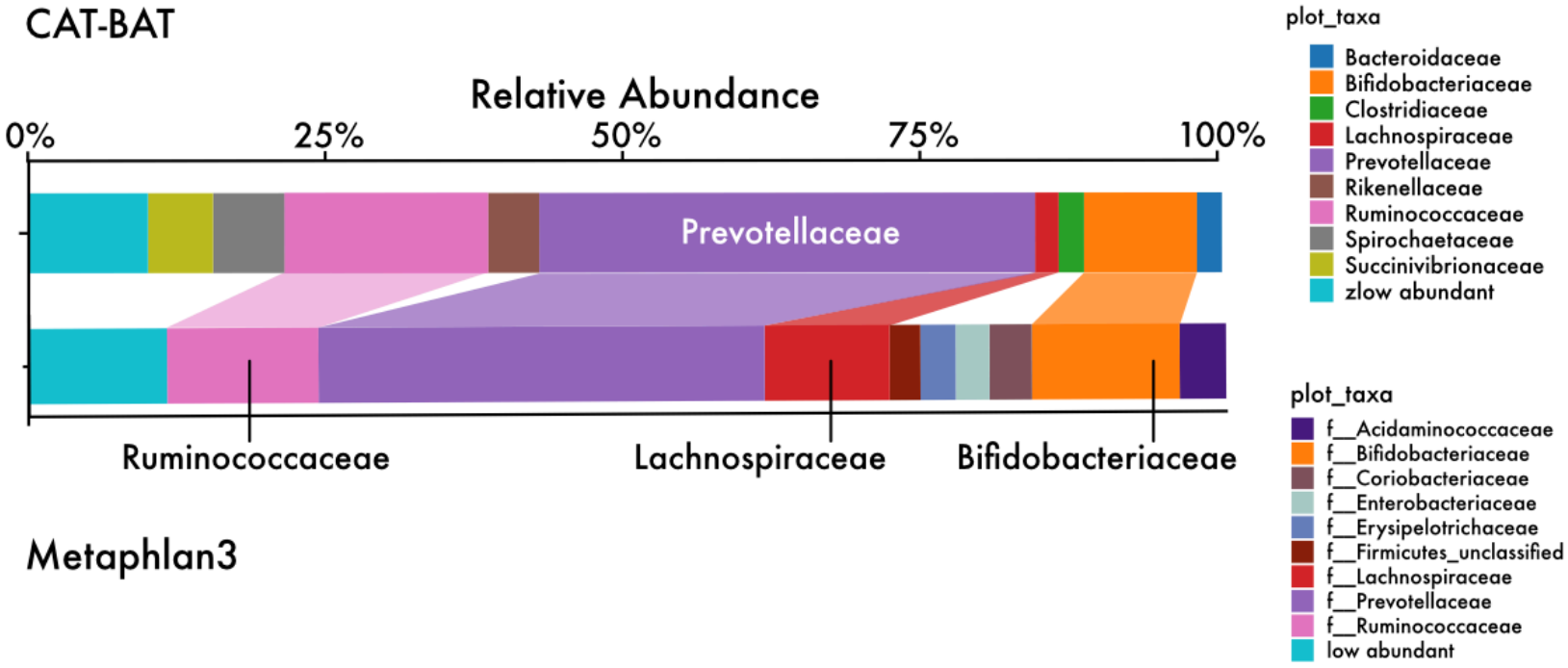
Supplemental Figure 3.1: Richness (number of distinct species) found using short-read taxonomic assignment (Metaphlan3) and assembly-based methods (CAT-BAT).



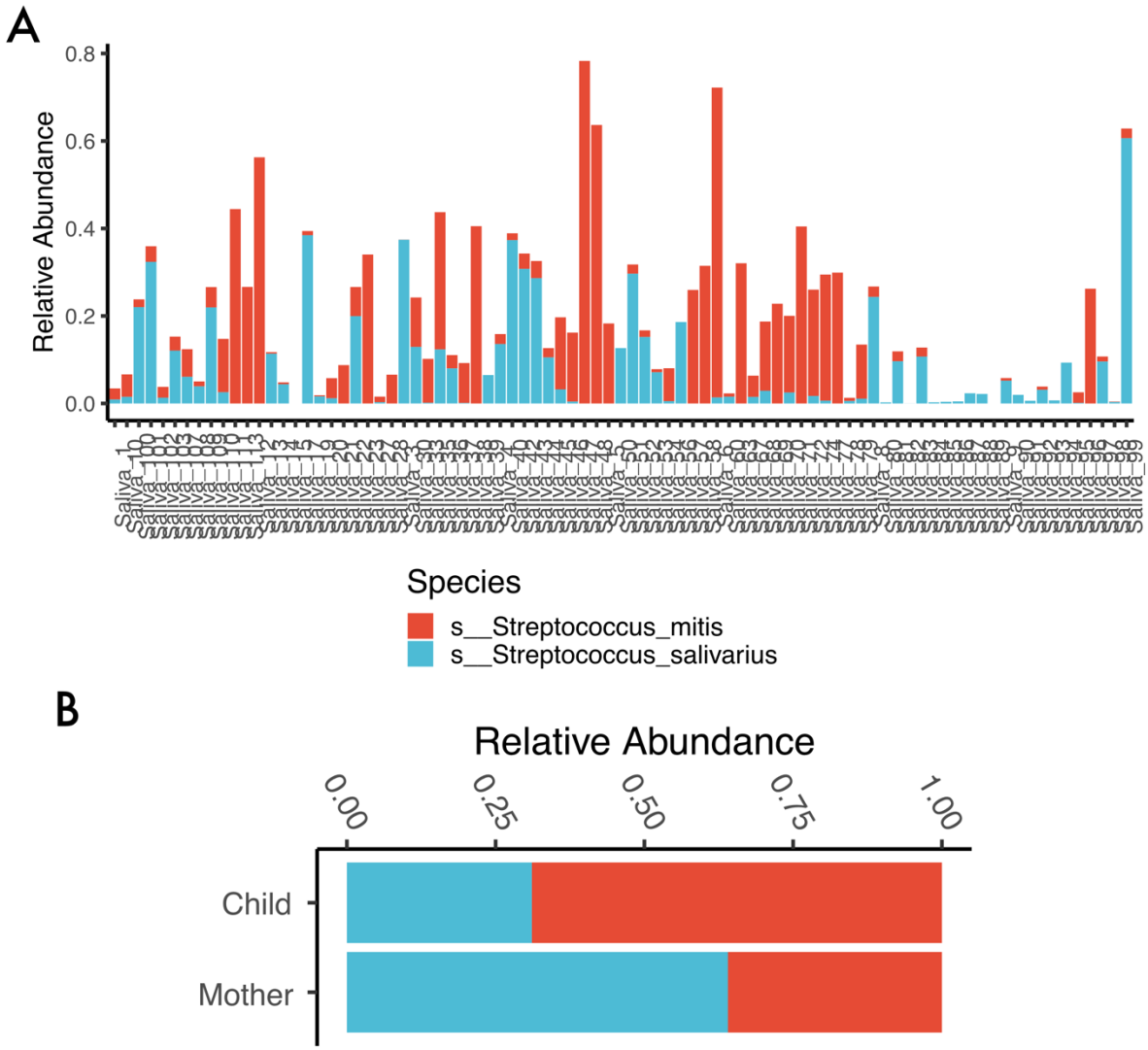
Supplemental Figure 3.2: Taxa bar plots, assigned using Metaphlan3 on short-reads, of fecal, saliva, and chicha samples, summarized at the family level. The 12 most abundant families (by average abundance) are plotted, and lower abundant taxa are grouped into the "other" categroy.



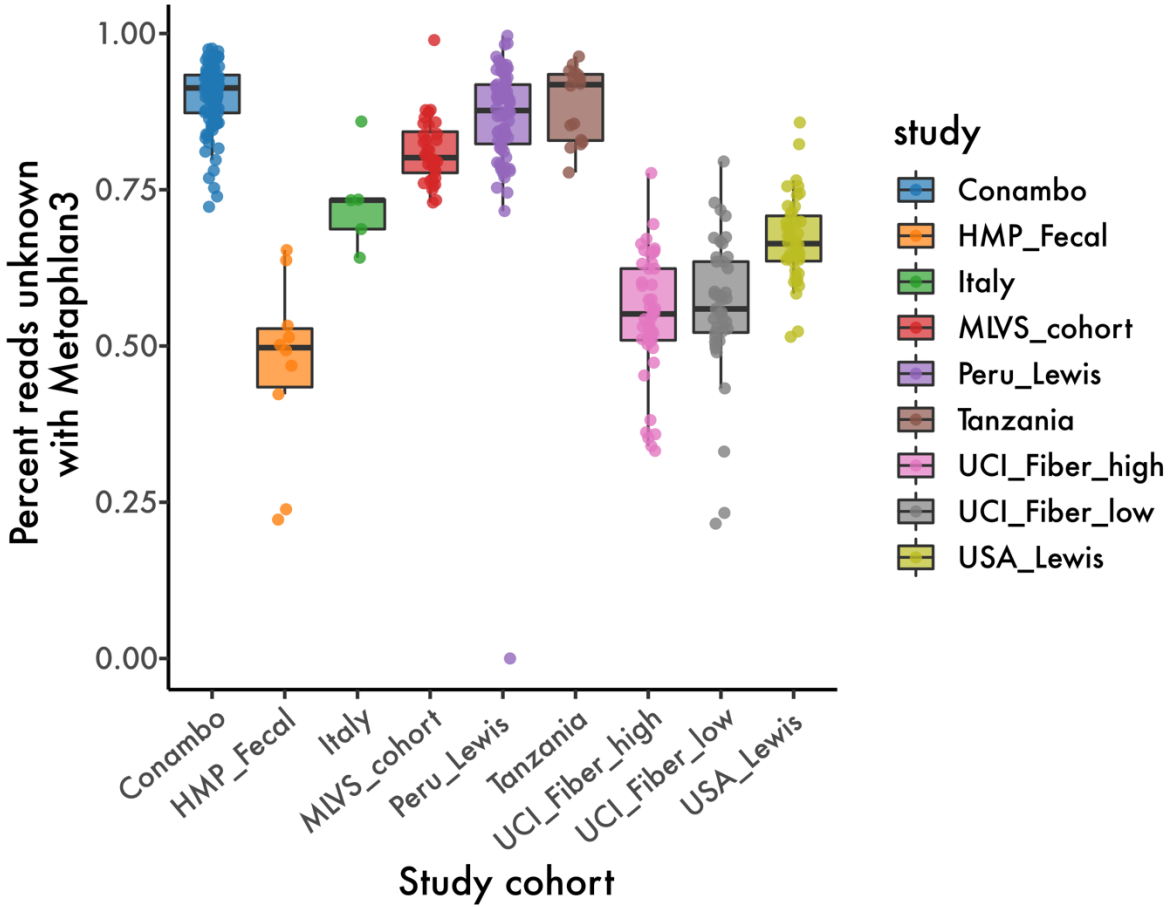
Supplemental Figure 3.3: Taxa bar plots, assigned using CAT-BAT on assemblies, of fecal, saliva, and chicha samples, summarized at the family level. The 12 most abundant families (by average abundance) are plotted, and lower abundant taxa are grouped into the "other" categroy.



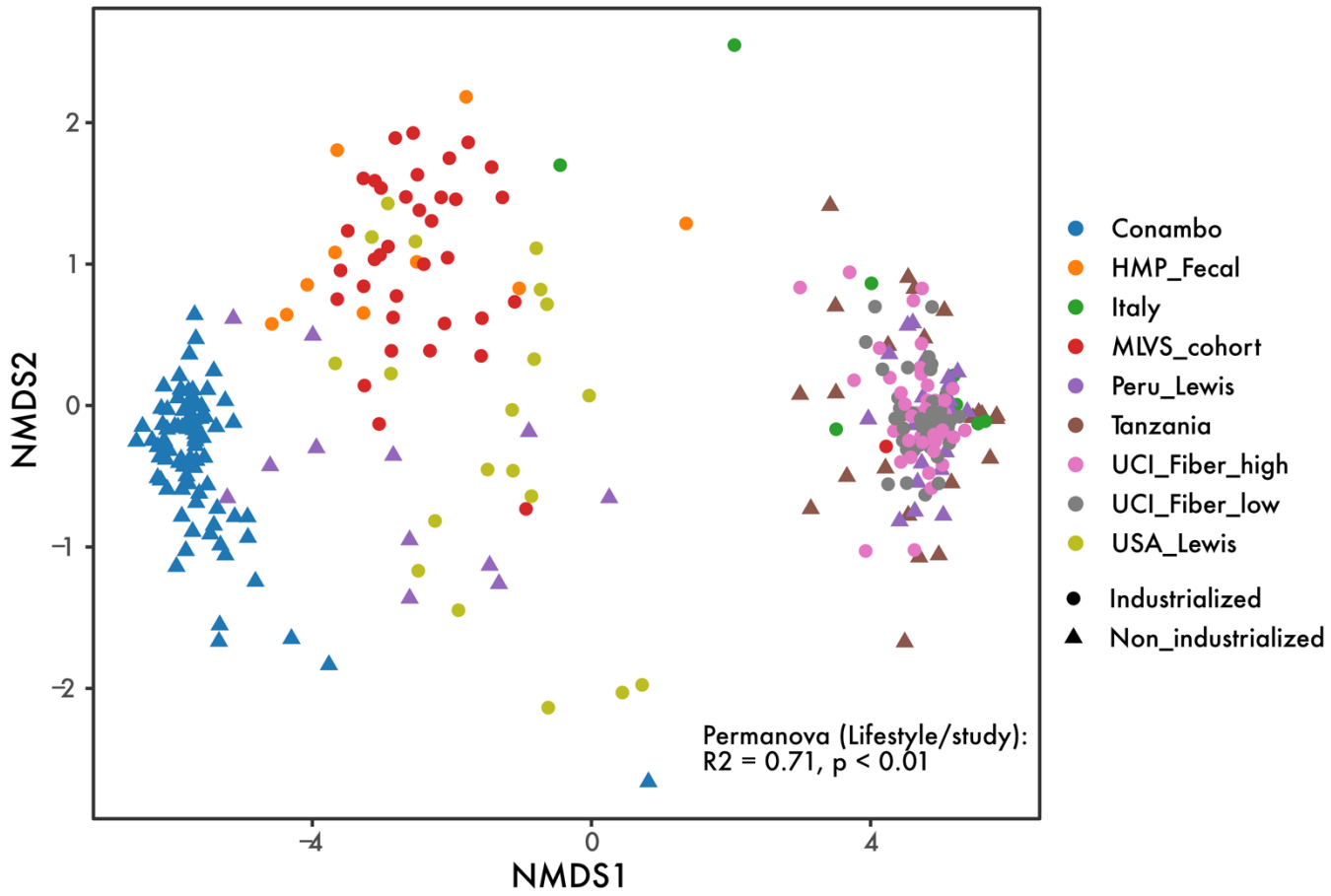
Supplemental Figure 3.4: Comparing the average fecal taxonomic assignment between Metaphlan3 and CAT-BAT (assemblies) using the Conambo samples.



Supplemental Figure 3.5: A) Relative abundance of Streptococcus salivarius and Streptococcus mitis across saliva samples. B) Relative abundance summarized between mother and child.



Supplemental Figure 3.6: Percent of unknown reads. These are reads that did sufficently map to marker genes in Metaphlan3 to be taxonomically assigned.



Supplemental Figure 3.7: NMDS ordination of glycoside hydrolase and polysaccharide lyase gene families from the Conambo and seven other cohorts

SUMMARY AND FUTURE DIRECTIONS

In the above work I have shown 1) the maternal microbiome and metabolome during pregnancy, and the implications of microbiome dysbiosis on female health and infant immune development (Chapter 1), 2) experimental evidence of microbiome modification through diets that more closely resemble pre-industrialized macronutrient consumption (Chapter 2), 3) and the microbiome of non-industrialized people, contextualizing the effects of industrialization on the microbiome (Chapter 3). I am certainly not the first to suggest that the advent of antibiotics, processed foods, and modern hygiene have put unique pressures on the ancient coevolution between human and microbe (231). Throughout this work I have highlighted how integral the microbiome is to acute and chronic disease, as well as the promotion of health. For example, increased consumption of processed foods have led to a precipitous drop in dietary fiber intake throughout many industrialized countries. Dietary fiber is a favored substrate for gut microbial metabolism, where bacteria ferment fiber to metabolites (short-chain fatty acids, SCFAs) critical to human health. It has been suggested that dietary changes may be a tractable therapeutic to promote a gut microbiome that functions to benefit human health (231). Therefore, understanding the factors that contribute to microbial assembly in humans will remain a critical area of research as we strive to understand how to engineer the microbiome to benefit health.

In Chapter 1, I discussed the vaginal microbiome during pregnancy. One fascinating result was the strong association between mannitol and vaginal microbiomes dominated by *Lactobacillus crispatus*. *L. crispatus* is thought to be a homofermentative lactic-acid bacteria, converting nearly all available glucose to lactic acid. An accumulation of mannitol in microbiomes dominated by *L. crispatus*, and not in microbiomes dominated by other organisms, suggests *L. crispatus* could be a source of mannitol production in the vagina. The most

parsimonious route for mannitol production is through fructose reduction. I hypothesize that in the absence of glucose, *L. crispatus* may switch to a previously underappreciated physiology of fructose reduction, which would allow for the regeneration of NAD+ in order to continue fermentation. This would allow the continued conversion of glyceraldehyde to pyruvate, which would continue ATP generation in the absence of glucose. The outcome of this metabolic shift would be a decrease in the production of lactic-acid, a molecule critical to vaginal health. Future work should determine if our *in vivo* observation of a mannitol–*L. crispatus* association is recapitulated *in vitro*. Feeding ^{13}C labeled fructose and measuring metabolic by-products would be an effective way to interrogate *L. crispatus* metabolism.

In Chapter 2, I investigated how the effect of a high-fiber diet intervention alters the gut microbial composition. While we illustrated that the microbiome is malleable to an influx of fiber, on the scale of two weeks, we failed to measure concomitant changes in fecal short-chain fatty acids. Others have speculated that circulating SCFAs may correlate more strongly with health benefits compared to fecal SCFAs (17, 187). In order to better understand the molecular mechanisms through which high fiber diets promote health, more work is needed measuring the flux between fecal and circulating SCFAs. Other promising ideas include synbiotics, or the formulation of a prebiotic (essentially food for microorganisms) together with probiotics. Synbiotics address one potential pitfall of a general dietary fiber increase; that is, increasing substrate availability (fiber) for health-beneficial microbes in the gut would only work if these microbes are present in the first place. To that end, finding optimum combinations of diet and probiotic may ultimately lead to therapeutics with broad effect.

Finally, in Chapter 3 I characterized the microbiomes of the Conambo village, deep in the Ecuadorian Amazon. These people live a distinctly non-industrialized lifestyle, extremely

isolated from urbanized environments. Populations such as these provide critical context to the role that industrialization has had in shaping the microbiome. Many of the organisms found are unique, rare, or highly divergent genetically compared to species found in microbiomes from industrialized countries. Future work should continue to catalog this sizable human-associated microbial diversity while endeavoring to be attentive to the impact this research has on isolated human populations.

1. **Sender R**, **Fuchs S**, **Milo R**. 2016. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLOS Biol **14**:e1002533.
2. **Dorrestein PC**, **Mazmanian SK**, **Knight R**. 2014. Finding the Missing Links among Metabolites, Microbes, and the Host. Immunity **40**:824–832.
3. **Costello EK**, **Stagaman K**, **Dethlefsen L**, **Bohannan BJM**, **Relman DA**. 2012. The application of ecological theory toward an understanding of the human microbiome. Science **336**:1255–62.
4. **Turnbaugh PJ**, **Ley RE**, **Hamady M**, **Fraser-Liggett CM**, **Knight R**, **Gordon JI**. 2007. The Human Microbiome Project. Nature **449**:804–810.
5. **Jašarević E**, **Howard CD**, **Morrison K**, **Misic A**, **Weinkopff T**, **Scott P**, **Hunter C**, **Beiting D**, **Bale TL**. 2018. The maternal vaginal microbiome partially mediates the effects of prenatal stress on offspring gut and hypothalamus. Nat Neurosci 1.
6. **Stanislawski MA**, **Dabelea D**, **Wagner BD**, **Sontag MK**, **Lozupone CA**, **Eggesbø M**. 2017. Pre-pregnancy weight, gestational weight gain, and the gut microbiota of mothers and their infants. Microbiome **5**:113.
7. **Ege MJ**, **Bieli C**, **Frei R**, **van Strien RT**, **Riedler J**, **Üblagger E**, **Schram-Bijkerk D**, **Brunekreef B**, **van Hage M**, **Scheynius A**, **Pershagen G**, **Benz MR**, **Lauener R**, **von Mutius E**, **Braun-Fahrländer C**, **the PARSIFAL Study team**. 2006. Prenatal farm exposure is related to the expression of receptors of the innate immunity and to atopic sensitization in school-age children. J Allergy Clin Immunol **117**:817–823.
8. **Gomez de Agüero M**, **Ganal-Vonarburg SC**, **Fuhrer T**, **Rupp S**, **Uchimura Y**, **Li H**, **Steinert A**, **Heikenwalder M**, **Hapfelmeier S**, **Sauer U**, **McCoy KD**, **Macpherson AJ**. 2016. The maternal microbiota drives early postnatal innate immune development. Science **351**:1296–302.
9. **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.
10. **Gomez A**, **Espinoza JL**, **Harkins DM**, **Leong P**, **Saffery R**, **Bockmann M**, **Torralba M**, **Kuelbs C**, **Kodukula R**, **Inman J**, **Hughes T**, **Craig JM**, **Highlander SK**, **Jones MB**, **Dupont CL**, **Nelson KE**. 2017. Host Genetic Control of the Oral Microbiome in Health and Disease. Cell Host Microbe **22**:269-278.e3.
11. **Scott KP**, **Gratz SW**, **Sheridan PO**, **Flint HJ**, **Duncan SH**. 2013. The influence of diet on the gut microbiota. Pharmacol Res **69**:52–60.
12. **O'Keefe SJD**, **Li J V.**, **Lahti L**, **Ou J**, **Carbonero F**, **Mohammed K**, **Posma JM**, **Kinross J**, **Wahl E**, **Ruder E**, **Vipperla K**, **Naidoo V**, **Mtshali L**, **Tims S**, **Puylaert PGB**, **DeLany J**, **Krasinskas A**, **Benefiel AC**, **Kaseb HO**, **Newton K**, **Nicholson JK**, **de Vos WM**, **Gaskins HR**, **Zoetendal EG**. 2015. Fat, fibre and cancer risk in African Americans and rural Africans. Nat Commun **6**:6342.
13. **den Besten G**, **van Eunen K**, **Groen AK**, **Venema K**, **Reijngoud D-J**, **Bakker BM**. 2013. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. J Lipid Res **54**:2325–40.
14. **Clemente JC**, **Ursell LK**, **Parfrey LW**, **Knight R**. 2012. The Impact of the Gut Microbiota on Human Health: An Integrative View. Cell **148**:1258–1270.
15. **Makki K**, **Deehan EC**, **Walter J**, **Bäckhed F**. 2018. The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. Cell Host Microbe **23**:705–715.

16. **Byndloss MX**, **Olsan EE**, **Rivera-Chávez F**, **Tiffany CR**, **Cevallos SA**, **Lokken KL**, **Torres TP**, **Byndloss AJ**, **Faber F**, **Gao Y**, **Litvak Y**, **Lopez CA**, **Xu G**, **Napoli E**, **Giulivi C**, **Tsolis RM**, **Revzin A**, **Lebrilla CB**, **Bäumler AJ**. 2017. Microbiota-activated PPAR-γ signaling inhibits dysbiotic Enterobacteriaceae expansion. Science (80-) **357**:570–575.
17. **de la Cuesta-Zuluaga J**, **Mueller N**, **Álvarez-Quintero R**, **Velásquez-Mejía E**, **Sierra J**, **Corrales-Agudelo V**, **Carmona J**, **Abad J**, **Escobar J**. 2018. Higher Fecal Short-Chain Fatty Acid Levels Are Associated with Gut Microbiome Dysbiosis, Obesity, Hypertension and Cardiometabolic Disease Risk Factors. Nutrients **11**:51.
18. **Tirosh A**, **Calay ES**, **Tuncman G**, **Claiborn KC**, **Inouye KE**, **Eguchi K**, **Alcala M**, **Rathaus M**, **Hollander KS**, **Ron I**, **Livne R**, **Heianza Y**, **Qi L**, **Shai I**, **Garg R**, **Hotamisligil GS**. 2019. The short-chain fatty acid propionate increases glucagon and FABP4 production, impairing insulin action in mice and humansSci. Transl. Med.
19. **Obregon-Tito AJ**, **Tito RY**, **Metcalf J**, **Sankaranarayanan K**, **Clemente JC**, **Ursell LK**, **Zech Xu Z**, **Van Treuren W**, **Knight R**, **Gaffney PM**, **Spicer P**, **Lawson P**, **Marin-Reyes L**, **Trujillo-Villarroel O**, **Foster M**, **Guija-Poma E**, **Troncoso-Corzo L**, **Warinner C**, **Ozga AT**, **Lewis CM**. 2015. Subsistence strategies in traditional societies distinguish gut microbiomes. Nat Commun **6**:6505.
20. **Clemente JC**, **Pehrsson EC**, **Blaser MJ**, **Sandhu K**, **Gao Z**, **Wang B**, **Magris M**, **Hidalgo G**, **Contreras M**, **Noya-Alarcón Ó**, **Lander O**, **McDonald J**, **Cox M**, **Walter J**, **Oh PL**, **Ruiz JF**, **Rodriguez S**, **Shen N**, **Song SJ**, **Metcalf J**, **Knight R**, **Dantas G**, **Dominguez-Bello MG**. 2015. The microbiome of uncontacted Amerindians. Sci Adv **1**:e1500183.
21. **De Filippo C**, **Cavalieri D**, **Di Paola M**, **Ramazzotti M**, **Poullet JB**, **Massart S**, **Collini S**, **Pieraccini G**, **Lionetti P**. 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc Natl Acad Sci **107**:14691–14696.
22. **Rampelli S**, **Schnorr SL**, **Consolandi C**, **Turroni S**, **Severgnini M**, **Peano C**, **Brigidi P**, **Crittenden AN**, **Henry AG**, **Candela M**. 2015. Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota. Curr Biol **25**:1682–1693.
23. **Smits SA**, **Leach J**, **Sonnenburg ED**, **Gonzalez CG**, **Lichtman JS**, **Reid G**, **Knight R**, **Manjurano A**, **Changalucha J**, **Elias JE**, **Dominguez-Bello MG**, **Sonnenburg JL**. 2017. Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. Science (80-) **357**:802–806.
24. **Brito IL**, **Gurry T**, **Zhao S**, **Huang K**, **Young SK**, **Shea TP**, **Naisilisili W**, **Jenkins AP**, **Jupiter SD**, **Gevers D**, **Alm EJ**. 2019. Transmission of human-associated microbiota along family and social networks. Nat Microbiol 1.
25. **Sonnenburg ED**, **Sonnenburg JL**. 2019. The ancestral and industrialized gut microbiota and implications for human health. Nat Rev Microbiol **17**:383–390.
26. **Patton JQ**. 2017. Reciprocal altruism and warfare a case from the Ecuadorian AmazonAdaptation and Human Behavior: An Anthropological Perspective.
27. **Bowser BJ**. 2000. From pottery to politics: An ethnoarchaeological study of political factionalism, ethnicity, and domestic pottery style in the Ecuadorian Amazon. J Archaeol Method Theory 219–248.
28. **Patton JQ**. 2005. Meat sharing for coalitional support. Evol Hum Behav **26**:137–157.
29. **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**, **De Sutter L**, **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**. 2016. Population-level analysis of gut microbiome variation. Science (80-) **352**:560–564.

30. **Zaneveld JR**, **Mcminds R**, **Vega Thurber R**. 2017. Stress and stability: applying the Anna Karenina principle to animal microbiomes. Nat Publ Gr **2**:17121.
31. **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.
32. **Cordain L**, **Eaton B**, **Sebastian A**, **Mann N**, **Lindeberg S**, **Watkins BA**, **O'keefe JH**, **Brand-Miller J**. 2005. Origins and evolution of the Western diet: health implications for the 21st century 1,2.
33. **Ly M**, **Jones MB**, **Abeles SR**, **Santiago-Rodriguez TM**, **Gao J**, **Chan IC**, **Ghose C**, **Pride DT**. 2016. Transmission of viruses via our microbiomes. Microbiome **4**:64.
34. **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-).
35. **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**, **Mohn WW**, **Turvey SE**, **Brett Finlay B**. 2015. Early infancy microbial and metabolic alterations affect risk of childhood asthma. Sci Transl Med **7**:307ra152-307ra152.
36. **Durack J**, **Kimes NE**, **Lin DL**, **Rauch M**, **McKean M**, **McCauley K**, **Panzer AR**, **Mar JS**, **Cabana MD**, **Lynch S V.** 2018. Delayed gut microbiota development in high-risk for asthma infants is temporarily modifiable by Lactobacillus supplementation. Nat Commun **9**:707.
37. **Fujimura KE**, **Sitarik AR**, **Havstad S**, **Lin DL**, **Levan S**, **Fadrosh D**, **Panzer AR**, **LaMere B**, **Rackaityte E**, **Lukacs NW**, **Wegienka G**, **Boushey HA**, **Ownby DR**, **Zoratti EM**, **Levin AM**, **Johnson CC**, **Lynch S V**. 2016. Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. Nat Med **22**:1187–1191.
38. **Jašarević E**, **Howerton CL**, **Howard CD**, **Bale TL**. 2015. Alterations in the Vaginal Microbiome by Maternal Stress Are Associated With Metabolic Reprogramming of the Offspring Gut and Brain. Endocrinology **156**:3265–3276.
39. **Walther-António MRS**, **Jeraldo P**, **Berg Miller ME**, **Yeoman CJ**, **Nelson KE**, **Wilson BA**, **White BA**, **Chia N**, **Creedon DJ**. 2014. Pregnancy's stronghold on the vaginal microbiome. PLoS One **9**:1–10.
40. **DiGiulio DB**, **Callahan BJ**, **McMurdie PJ**, **Costello EK**, **Lyell DJ**, **Robaczewska A**, **Sun CL**, **Goltsman DSA**, **Wong RJ**, **Shaw G**, **Stevenson DK**, **Holmes SP**, **Relman DA**. 2015. Temporal and spatial variation of the human microbiota during pregnancy. Proc Natl Acad Sci U S A **112**:11060–5.
41. **Nasioudis D**, **Forney LJ**, **Schneider GM**, **Gliniewicz K**, **France M**, **Boester A**, **Sawai M**, **Scholl J**, **Witkin SS**. 2017. Influence of Pregnancy History on the Vaginal

Microbiome of Pregnant Women in their First Trimester. Sci Rep **7**:10201.
42. **Noyes N**, **Cho K-C**, **Ravel J**, **Forney LJ**, **Abdo Z**. 2018. Associations between sexual habits, menstrual hygiene practices, demographics and the vaginal microbiome as revealed by Bayesian network analysis. PLoS One **13**:e0191625.
43. **Gajer P**, **Brotman RM**, **Bai G**, **Sakamoto J**, **Schütte UME**, **Zhong X**, **Koenig SSK**, **Fu L**, **Ma ZS**, **Zhou X**, **Abdo Z**, **Forney LJ**, **Ravel J**. 2012. Temporal dynamics of the human vaginal microbiota. Sci Transl Med **4**:132ra52.
44. **Serrano MG**, **Parikh HI**, **Brooks JP**, **Edwards DJ**, **Arodz TJ**, **Edupuganti L**, **Huang B**, **Girerd PH**, **Bokhari YA**, **Bradley SP**, **Brooks JL**, **Dickinson MR**, **Drake JI**, **Duckworth RA**, **Fong SS**, **Glascock AL**, **Jean S**, **Jimenez NR**, **Khoury J**, **Koparde VN**, **Lara AM**, **Lee V**, **Matveyev A V.**, **Milton SH**, **Mistry SD**, **Rozycki SK**, **Sheth NU**, **Smirnova E**, **Vivadelli SC**, **Wijesooriya NR**, **Xu J**, **Xu P**, **Chaffin DO**, **Sexton AL**, **Gravett MG**, **Rubens CE**, **Hendricks-Muñoz KD**, **Jefferson KK**, **Strauss JF**, **Fettweis JM**, **Buck GA**. 2019. Racioethnic diversity in the dynamics of the vaginal microbiome during pregnancy. Nat Med **25**:1001–1011.
45. **Petrova MI**, **Reid G**, **Vaneechoutte M**, **Lebeer S**. 2017. Lactobacillus iners: Friend or Foe? Trends Microbiol **25**:182–191.
46. **Ravel J**, **Gajer P**, **Abdo Z**, **Schneider GM**, **Koenig SSK**, **McCulle SL**, **Karlebach S**, **Gorle R**, **Russell J**, **Tacket CO**, **Brotman RM**, **Davis CC**, **Ault K**, **Peralta L**, **Forney LJ**. 2011. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci **108**:4680–4687.
47. **O'Hanlon DE**, **Moench TR**, **Cone RA**. 2011. In vaginal fluid, bacteria associated with bacterial vaginosis can be suppressed with lactic acid but not hydrogen peroxide. BMC Infect Dis **11**:200.
48. **Rampersaud R**, **Planet PJ**, **Randis TM**, **Kulkarni R**, **Aguilar JL**, **Lehrer RI**, **Ratner AJ**. 2011. Inerolysin, a cholesterol-dependent cytolysin produced by Lactobacillus iners. J Bacteriol **193**:1034–41.
49. **Macklaim JM**, **Gloor GB**, **Anukam KC**, **Cribby S**, **Reid G**. 2011. At the crossroads of vaginal health and disease, the genome sequence of Lactobacillus iners AB-1. Proc Natl Acad Sci U S A **108 Suppl**:4688–95.
50. **Verstraelen H**, **Verhelst R**, **Claeys G**, **De Backer E**, **Temmerman M**, **Vaneechoutte M**. 2009. Longitudinal analysis of the vaginal microflora in pregnancy suggests that L. crispatus promotes the stability of the normal vaginal microflora and that L. gasseri and/or L. iners are more conducive to the occurrence of abnormal vaginal microflora. BMC Microbiol **9**:116.
51. **Sobel JD**. 2000. Bacterial Vaginosis. Annu Rev Med **51**:349–356.
52. **Fettweis JM**, **Serrano MG**, **Brooks JP**, **Edwards DJ**, **Girerd PH**, **Parikh HI**, **Huang B**, **Arodz TJ**, **Edupuganti L**, **Glascock AL**, **Xu J**, **Jimenez NR**, **Vivadelli SC**, **Fong SS**, **Sheth NU**, **Jean S**, **Lee V**, **Bokhari YA**, **Lara AM**, **Mistry SD**, **Duckworth RA**, **Bradley SP**, **Koparde VN**, **Orenda XV**, **Milton SH**, **Rozycki SK**, **Matveyev A V.**, **Wright ML**, **Huzurbazar S V.**, **Jackson EM**, **Smirnova E**, **Korlach J**, **Tsai Y-C**, **Dickinson MR**, **Brooks JL**, **Drake JI**, **Chaffin DO**, **Sexton AL**, **Gravett MG**, **Rubens CE**, **Wijesooriya NR**, **Hendricks-Muñoz KD**, **Jefferson KK**, **Strauss JF**, **Buck GA**. 2019. The vaginal microbiome and preterm birth. Nat Med **25**:1012–1021.
53. **Watts DH**, **Krohn MA**, **Hillier SL**, **Eschenbach DA**. 1990. Bacterial vaginosis as a risk factor for post-cesarean endometritis. Obstet Gynecol.

54. **Acobsson BJ**, **Ernevi PP**, **Hidekel LC**, **J J**. 2002. Bacterial vaginosis in early pregnancy may predispose for preterm birth and postpartum endometritis. Acta Obstet Gynecol Scand.
55. **Hay PE**, **Lamont RF**, **Taylor-Robinson D**, **Morgan DJ**, **Ison C**, **Pearson J**. 1994. Abnormal bacterial colonisation of the genital tract and subsequent preterm delivery and late miscarriage. BMJ.
56. **Llahí-Camp JM**, **Rai R**, **Ison C**, **Regan L**, **Taylor-Robinson D**. 1996. Association of bacterial vaginosis with a history of second trimester miscarriage. Hum Reprod.
57. **Donders GG**, **Van Bulck B**, **Caudron J**, **Londers L**, **Vereecken A**, **Spitz B**. 2000. Relationship of bacterial vaginosis and mycoplasmas to the risk of spontaneous abortion. Am J Obstet Gynecol.
58. **Ralph SG**, **Rutherford AJ**, **Wilson JD**. 1999. Influence of bacterial vaginosis on conception and miscarriage in the first trimester: cohort study. BMJ.
59. **Lev-Sagie A**, **Goldman-Wohl D**, **Cohen Y**, **Dori-Bachash M**, **Leshem A**, **Mor U**, **Strahilevitz J**, **Moses AE**, **Shapiro H**, **Yagel S**, **Elinav E**. 2019. Vaginal microbiome transplantation in women with intractable bacterial vaginosis. Nat Med **25**:1500–1504.
60. **Masfari AN**, **Duerden BI**, **Kinghorn GR**. 1986. Quantitative studies of vaginal bacteria. Sex Transm Infect.
61. **Lindsay KL**, **Hellmuth C**, **Uhl O**, **Buss C**, **Wadhwa PD**, **Koletzko B**, **Entringer S**. 2015. Longitudinal Metabolomic Profiling of Amino Acids and Lipids across Healthy Pregnancy. PLoS One **10**.
62. **Entringer S**, **Buss C**, **Rasmussen JM**, **Lindsay K**, **Gillen DL**, **Cooper DM**, **Wadhwa PD**. 2017. Maternal Cortisol During Pregnancy and Infant Adiposity: A Prospective Investigation. J Clin Endocrinol Metab **102**:1366–1374.
63. **Fiehn O**. 2016. Metabolomics by Gas Chromatography-Mass Spectrometry: Combined Targeted and Untargeted Profiling, p. 30.4.1-30.4.32. *In* Current Protocols in Molecular Biology. John Wiley & Sons, Inc., Hoboken, NJ, USA.
64. **Cajka T**, **Fiehn O**. 2016. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. Anal Chem **88**:524–545.
65. **Cajka T**, **Smilowitz JT**, **Fiehn O**. 2017. Validating Quantitative Untargeted Lipidomics Across Nine Liquid Chromatography–High-Resolution Mass Spectrometry Platforms. Anal Chem **89**:12360–12368.
66. **de Raad M**, **de Rond T**, **Rübel O**, **Keasling JD**, **Northen TR**, **Bowen BP**. 2017. OpenMSI Arrayed Analysis Toolkit: Analyzing Spatially Defined Samples Using Mass Spectrometry Imaging. Anal Chem **89**:5818–5823.
67. **Caporaso JG**, **Lauber CL**, **Walters WA**, **Berg-Lyons D**, **Lozupone CA**, **Turnbaugh PJ**, **Fierer N**, **Knight R**. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci **108**:4516–4522.
68. **Baym M**, **Kryazhimskiy S**, **Lieberman TD**, **Chung H**, **Desai MM**, **Kishony RK**. 2015. Inexpensive multiplexed library preparation for megabase-sized genomes. PLoS One **10**:1–15.
69. **Abarenkov K**, **Henrik Nilsson R**, **Larsson K-H**, **Alexander IJ**, **Eberhardt U**, **Erland S**, **Høiland K**, **Kjøller R**, **Larsson E**, **Pennanen T**, **Sen R**, **Taylor AFS**, **Tedersoo L**, **Ursing BM**, **Vrålstad T**, **Liimatainen K**, **Peintner U**, **Kõljalg U**. 2010. The UNITE database for molecular identification of fungi - recent updates and future perspectives. New Phytol **186**:281–285.

70. **Caporaso JG**, **Kuczynski J**, **Stombaugh J**, **Bittinger K**, **Bushman FD**, **Costello EK**, **Fierer N**, **Peña 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.
71. **Pinheiro J**, **Bates D**, **DebRoy S**, **Sarkar D**, **R Core Team**. 2019. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-141.
72. **R Core Team**. 2018. No TitleR: A language and environment for statistical computing. Vienna, Austria.
73. **Schmieder R**, **Edwards R**. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics **27**:863–864.
74. **Langmead B**, **Salzberg SL**. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9**:357–359.
75. **Franzosa EA**, **McIver LJ**, **Rahnavard G**, **Thompson LR**, **Schirmer M**, **Weingart G**, **Lipson KS**, **Knight R**, **Caporaso JG**, **Segata N**, **Huttenhower C**. 2018. Species-level functional profiling of metagenomes and metatranscriptomes. Nat Methods **15**:962–968.
76. **Segata N**, **Izard J**, **Waldron L**, **Gevers D**, **Miropolsky L**, **Garrett WS**, **Huttenhower C**. 2011. Metagenomic biomarker discovery and explanation. Genome Biol **12**:R60.
77. **Bankevich A**, **Nurk S**, **Antipov D**, **Gurevich AA**, **Dvorkin M**, **Kulikov AS**, **Lesin VM**, **Nikolenko SI**, **Pham S**, **Prjibelski AD**, **Pyshkin A V.**, **Sirotkin A V.**, **Vyahhi N**, **Tesler G**, **Alekseyev MA**, **Pevzner PA**. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol **19**:455–477.
78. **Eren AM**, **Esen ÖC**, **Quince C**, **Vineis JH**, **Morrison HG**, **Sogin ML**, **Delmont TO**. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ **3**:e1319.
79. **Menzel P**, **Ng KL**, **Krogh A**. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun **7**:11257.
80. **Clark KR**, **Warwick RM**, **Gorley RN**, **Somerfield PJ**. 2014. Change in marine communities: an approach to statistical analysis and interpretation, 3nd edition. PRIMER-E: Plymouth.
81. **Archer E**. 2019. rfPermute: Estimate Permutation p-Values for Random Forest Importance Metrics. 2.1.7.
82. **Barton K**. 2019. MuMIn: Multi-Model Inference. 1.43.6.
83. **Pielou EC**. 1966. The measurement of diversity in different types of biological collections. J Theor Biol **13**:131–144.
84. **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.
85. **Brooks JP**, **Buck GA**, **Chen G**, **Diao L**, **Edwards DJ**, **Fettweis JM**, **Huzurbazar S**, **Rakitin A**, **Satten GA**, **Smirnova E**, **Waks Z**, **Wright ML**, **Yanover C**, **Zhou Y-H**. 2017. Changes in vaginal community state types reflect major shifts in the microbiome. Microb Ecol Health Dis **28**:1303265.
86. **Grigoriou O**, **Baka S**, **Makrakis E**, **Hassiakos D**, **Kapparos G**, **Kouskouni E**. 2006. Prevalence of clinical vaginal candidiasis in a university hospital and possible risk factors.

Eur J Obstet Gynecol Reprod Biol **126**:121–125.
87. **Wang S**, **Wang Q**, **Yang E**, **Yan L**, **Li T**, **Zhuang H**. 2017. Antimicrobial Compounds Produced by Vaginal Lactobacillus crispatus Are Able to Strongly Inhibit Candida albicans Growth, Hyphal Formation and Regulate Virulence-related Gene Expressions. Front Microbiol **08**:564.
88. **Narayanan TK**, **Rao GR**. 1976. Beta-indoleethanol and beta-indolelactic acid production by Candida species: their antibacterial and autoantibiotic action. Antimicrob Agents Chemother **9**:375–80.
89. **Naz S**, **Cretenet M**, **Vernoux JP**. Current knowledge on antimicrobial metabolites produced from aromatic amino acid metabolism in fermented products.
90. **Esser C**, **Rannug A**. 2015. The Aryl Hydrocarbon Receptor in Barrier Organ Physiology, Immunology, and Toxicology. Pharmacol Rev **67**:259–279.
91. **Zelante T**, **Iannitti RG**, **Cunha C**, **De Luca A**, **Giovannini G**, **Pieraccini G**, **Zecchi R**, **D'Angelo C**, **Massi-Benedetti C**, **Fallarino F**, **Carvalho A**, **Puccetti P**, **Romani L**. 2013. Tryptophan Catabolites from Microbiota Engage Aryl Hydrocarbon Receptor and Balance Mucosal Reactivity via Interleukin-22. Immunity **39**:372–385.
92. **Yang J**, **Chawla R**, **Rhee KY**, **Gupta R**, **Manson MD**, **Jayaraman A**, **Lele PP**. 2020. Biphasic chemotaxis of *Escherichia coli* to the microbiota metabolite indole. Proc Natl Acad Sci 201916974.
93. **Yeoman CJ**, **Yildirim S**, **Thomas SM**, **Durkin AS**, **Torralba M**, **Sutton G**, **Buhay CJ**, **Ding Y**, **Dugan-Rocha SP**, **Muzny DM**, **Qin X**, **Gibbs RA**, **Leigh SR**, **Stumpf R**, **White BA**, **Highlander SK**, **Nelson KE**, **Wilson BA**. 2010. Comparative Genomics of Gardnerella vaginalis Strains Reveals Substantial Differences in Metabolic and Virulence Potential. PLoS One **5**:e12411.
94. **Pybus V**, **Onderdonk AB**. 1997. Evidence for a Commensal, Symbiotic Relationship between Gardnerella vaginalis and Prevotella bivia Involving Ammonia: Potential Significance forSource: The Journal of Infectious Diseases.
95. **Anahtar MN**, **Byrne EH**, **Doherty KE**, **Bowman BA**, **Yamamoto HS**, **Soumillon M**, **Padavattan N**, **Ismail N**, **Moodley A**, **Sabatini ME**, **Ghebremichael MS**, **Nusbaum C**, **Huttenhower C**, **Virgin HW**, **Ndung'u T**, **Dong KL**, **Walker BD**, **Fichorova RN**, **Kwon DS**. 2015. Cervicovaginal Bacteria Are a Major Modulator of Host Inflammatory Responses in the Female Genital Tract. Immunity **42**:965–976.
96. **Forney LJ**, **Gliniewicz KS**. 2018. Use of estrogenic compounds to manipulate the bacterial composition of vaginal communities. US20180305742A1. United States.
97. **Huang L**, **Martin SM**, **Villanueva J**, **Greene S**, **Arehart K**, **Sayre C**, **Johnson RB**. 2006. Therapeutic agents for inhibiting and/or treating vaginal infection. 0105963 A1.
98. **Wu N**, **Zhang X**, **Li F**, **Zhang T**, **Gan Y**, **Li J**. 2015. Spray-dried powders enhance vaginal siRNA delivery by potentially modulating the mucus molecular sieve structure. Int J Nanomedicine **10**:5383–96.
99. **Eun Chang C**, **Kim S-C**, **So J-S**, **Shik Yun H**. 2001. Cultivation of Lactobacillus crispatus KLB46 Isolated from Human VaginaBiotechnol. Bioprocess Eng.
100. **Van Der Linden PJQ**, **Wiegerinck MAHM**. 1992. Cyclic changes in the concentration of glucose and fructose in human cervical mucusFertility and Sterility.
101. **Donnarumma G**, **Molinaro A**, **Cimini D**, **De Castro C**, **Valli V**, **De Gregorio V**, **De Rosa M**, **Schiraldi C**. 2014. Lactobacillus crispatus L1: high cell density cultivation and exopolysaccharide structure characterization to highlight potentially beneficial effects

against vaginal pathogens. BMC Microbiol **14**:137.
102. **Ojala T**, **Kankainen M**, **Castro J**, **Cerca N**, **Edelman S**, **Westerlund-Wikström B**, **Paulin L**, **Holm L**, **Auvinen P**. 2014. Comparative genomics of Lactobacillus crispatus suggests novel mechanisms for the competitive exclusion of Gardnerella vaginalis. BMC Genomics **15**:1070.
103. **Wisselink H.**, **Weusthuis R.**, **Eggink G**, **Hugenholtz J**, **Grobben G.** 2002. Mannitol production by lactic acid bacteria: a review. Int Dairy J **12**:151–161.
104. **Wisselink HW**, **Moers APHA**, **Mars AE**, **Hoefnagel MHN**, **De Vos WM**, **Hugenholtz J**. 2005. Overproduction of heterologous mannitol 1-phosphatase: A key factor for engineering mannitol production by Lactococcus lactis. Appl Environ Microbiol **71**:1507–1514.
105. **Gomez A**, **Nelson KE**. The Oral Microbiome of Children: Development, Disease, and Implications Beyond Oral Health.
106. **Loening-Baucke V**. 2005. Prevalence, symptoms and outcome of constipation in infants and toddlers. J Pediatr **146**:359–363.
107. **Adachi S**, **Patton S**. 1961. Presence and Significance of Lactulose in Milk Products: A Review. J Dairy Sci **44**:1375–1393.
108. **Scalabre A**, **Jobard E**, **Deme D**, **Gaillard ne**, **Pontoizeau ment**, **Mouriquand P**, **Elena-Herrmann dicte**, **Mure P-Y**. 2017. Evolution of Newborns' Urinary Metabolomic Profiles According to Age and Growth.
109. **Angeletti S**, **Ciccozzi M**. 2019. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry in clinical microbiology: An updating review. Infect Genet Evol **76**:104063.
110. **Jones JM**. 2014. CODEX-aligned dietary fiber definitions help to bridge the "fiber gap." Nutr J **13**:34.
111. **US Department of Agriculture; Agricultural Research Service**. 2012. Nutrient Intakes from Food: Mean Amounts Consumed per Individual, by Gender and Age, What We Eat in America, NHANES 2009-2010. Natl Heal Nutr Exam Surv 2009-10.
112. **Zhao L**, **Zhang F**, **Ding X**, **Wu G**, **Lam YY**, **Wang X**, **Fu H**, **Xue X**, **Lu C**, **Ma J**, **Yu L**, **Xu C**, **Ren Z**, **Xu Y**, **Xu S**, **Shen H**, **Zhu X**, **Shi Y**, **Shen Q**, **Dong W**, **Liu R**, **Ling Y**, **Zeng Y**, **Wang X**, **Zhang Q**, **Wang J**, **Wang L**, **Wu Y**, **Zeng B**, **Wei H**, **Zhang M**, **Peng Y**, **Zhang C**. 2018. Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. Science (80-) **359**:1151–1156.
113. **Reynolds A**, **Mann J**, **Cummings J**, **Winter N**, **Mete E**, **Te Morenga L**. 2019. Carbohydrate quality and human health: a series of systematic reviews and meta-analyses. Lancet (London, England) **0**.
114. **O'Keefe SJD**. 2016. Diet, microorganisms and their metabolites, and colon cancer. Nat Rev Gastroenterol Hepatol **13**:691–706.
115. **O'Keefe SJD**, **Li J V**, **Lahti L**, **Ou J**, **Carbonero F**, **Mohammed K**, **Posma JM**, **Kinross J**, **Wahl E**, **Ruder E**, **Vipperla K**, **Naidoo V**, **Mtshali L**, **Tims S**, **Puylaert PGB**, **DeLany J**, **Krasinskas A**, **Benefiel AC**, **Kaseb HO**, **Newton K**, **Nicholson JK**, **de Vos WM**, **Gaskins HR**, **Zoetendal EG**. 2015. Fat, fibre and cancer risk in African Americans and rural Africans. Nat Commun **6**:6342.
116. **Trompette A**, **Gollwitzer ES**, **Pattaroni C**, **Lopez-Mejia IC**, **Riva E**, **Pernot J**, **Ubags N**, **Fajas L**, **Nicod LP**, **Marsland BJ**. 2018. Dietary Fiber Confers Protection against Flu by Shaping Ly6c− Patrolling Monocyte Hematopoiesis and CD8+ T Cell Metabolism.

Immunity **48**:992-1005.e8.
117. **Oh JZ**, **Ravindran R**, **Chassaing B**, **Carvalho FA**, **Maddur MS**, **Bower M**, **Hakimpour P**, **Gill KP**, **Nakaya HI**, **Yarovinsky F**, **Sartor RB**, **Gewirtz AT**, **Pulendran B**. 2014. TLR5-mediated sensing of gut microbiota is necessary for antibody responses to seasonal influenza vaccination. Immunity **41**:478–492.
118. **Utzschneider KM**, **Kratz M**, **Damman CJ**, **Hullarg M**. 2016. Mechanisms Linking the Gut Microbiome and Glucose Metabolism. J Clin Endocrinol Metab. Oxford Academic.
119. **Cantarel BL**, **Lombard V**, **Henrissat B**. 2012. Complex Carbohydrate Utilization by the Healthy Human Microbiome. PLoS One **7**:e28742.
120. **Kaoutari A El**, **Armougom F**, **Gordon JI**, **Raoult D**, **Henrissat B**. 2013. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nat Rev Microbiol **11**:497–504.
121. **Koh A**, **De Vadder F**, **Kovatcheva-Datchary P**, **Bäckhed F**. 2016. From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites. Cell **165**:1332–1345.
122. **Roediger WEW**. 1982. Utilization of Nutrients by Isolated Epithelial Cells of the Rat ColonGastroenterology.
123. **Hryckowian AJ**, **Van Treuren W**, **Smits SA**, **Davis NM**, **Gardner JO**, **Bouley DM**, **Sonnenburg JL**. 2018. Microbiota-Accessible carbohydrates suppress Clostridium difficile infection in a murine model. Nat Microbiol **3**:662–669.
124. **Baxter NT**, **Schmidt AW**, **Venkataraman A**, **Kim KS**, **Waldron C**, **Schmidt TM**. 2019. Dynamics of Human Gut Microbiota and Short-Chain Fatty Acids in Response to Dietary Interventions with Three Fermentable Fibers. MBio **10**:e02566-18.
125. **Venkataraman A**, **Sieber JR**, **Schmidt AW**, **Waldron C**, **Theis KR**, **Schmidt TM**. 2016. Variable responses of human microbiomes to dietary supplementation with resistant starch. Microbiome **4**:33.
126. **Sawicki CM**, **Livingston KA**, **Obin M**, **Roberts SB**, **Chung M**, **Mckeown NM**. 2017. Dietary Fiber and the Human Gut Microbiota: Application of Evidence Mapping Methodology.
127. **Velázquez M**, **Davies C**, **Marett R**, **Slavin JL**, **Feirtag JM**. 2000. Effect of Oligosaccharides and Fibre Substitutes on Short-chain Fatty Acid Production by Human Faecal Microflora. Anaerobe **6**:87–92.
128. **Martínez I**, **Lattimer JM**, **Hubach KL**, **Case JA**, **Yang J**, **Weber CG**, **Louk JA**, **Rose DJ**, **Kyureghian G**, **Peterson DA**, **Haub MD**, **Walter J**. 2013. Gut microbiome composition is linked to whole grain-induced immunological improvements. ISME J **7**:269–280.
129. **McDonald D**, **Hyde E**, **Debelius JW**, **Morton JT**, **Gonzalez A**, **Ackermann G**, **Aksenov AA**, **Behsaz B**, **Brennan C**, **Chen Y**, **DeRight Goldasich L**, **Dorrestein PC**, **Dunn RR**, **Fahimipour AK**, **Gaffney J**, **Gilbert JA**, **Gogul G**, **Green JL**, **Hugenholtz P**, **Humphrey G**, **Huttenhower C**, **Jackson MA**, **Janssen S**, **Jeste D V.**, **Jiang L**, **Kelley ST**, **Knights D**, **Kosciolek T**, **Ladau J**, **Leach J**, **Marotz C**, **Meleshko D**, **Melnik A V.**, **Metcalf JL**, **Mohimani H**, **Montassier E**, **Navas-Molina J**, **Nguyen TT**, **Peddada S**, **Pevzner P**, **Pollard KS**, **Rahnavard G**, **Robbins-Pianka A**, **Sangwan N**, **Shorenstein J**, **Smarr L**, **Song SJ**, **Spector T**, **Swafford AD**, **Thackray VG**, **Thompson LR**, **Tripathi A**, **Vázquez-Baeza Y**, **Vrbanac A**, **Wischmeyer P**, **Wolfe E**, **Zhu Q**, **Knight R**, **Mann AE**, **Amir A**, **Frazier A**, **Martino C**, **Lebrilla C**, **Lozupone C**, **Lewis CM**, **Raison C**,

Zhang C, **Lauber CL**, **Warinner C**, **Lowry CA**, **Callewaert C**, **Bloss C**, **Willner D**, **Galzerani DD**, **Gonzalez DJ**, **Mills DA**, **Chopra D**, **Gevers D**, **Berg-Lyons D**, **Sears DD**, **Wendel D**, **Lovelace E**, **Pierce E**, **TerAvest E**, **Bolyen E**, **Bushman FD**, **Wu GD**, **Church GM**, **Saxe G**, **Holscher HD**, **Ugrina I**, **German JB**, **Caporaso JG**, **Wozniak JM**, **Kerr J**, **Ravel J**, **Lewis JD**, **Suchodolski JS**, **Jansson JK**, **Hampton-Marcell JT**, **Bobe J**, **Raes J**, **Chase JH**, **Eisen JA**, **Monk J**, **Clemente JC**, **Petrosino J**, **Goodrich J**, **Gauglitz J**, **Jacobs J**, **Zengler K**, **Swanson KS**, **Lewis K**, **Mayer K**, **Bittinger K**, **Dillon L**, **Zaramela LS**, **Schriml LM**, **Dominguez-Bello MG**, **Jankowska MM**, **Blaser M**, **Pirrung M**, **Minson M**, **Kurisu M**, **Ajami N**, **Gottel NR**, **Chia N**, **Fierer N**, **White O**, **Cani PD**, **Gajer P**, **Strandwitz P**, **Kashyap P**, **Dutton R**, **Park RS**, **Xavier RJ**, **Mills RH**, **Krajmalnik-Brown R**, **Ley R**, **Owens SM**, **Klemmer S**, **Matamoros S**, **Mirarab S**, **Moorman S**, **Holmes S**, **Schwartz T**, **Eshoo-Anton TW**, **Vigers T**, **Pandey V**, **Treuren W Van**, **Fang X**, **Zech Xu Z**, **Jarmusch A**, **Geier J**, **Reeve N**, **Silva R**, **Kopylova E**, **Nguyen D**, **Sanders K**, **Salido Benitez RA**, **Heale AC**, **Abramson M**, **Waldispühl J**, **Butyaev A**, **Drogaris C**, **Nazarova E**, **Ball M**, **Gunderson B**. 2018. American Gut: an Open Platform for Citizen Science Microbiome Research. mSystems **3**:2020.

130. **Claesson MJ**, **Jeffery IB**, **Conde S**, **Power SE**, **O'connor EM**, **Cusack S**, **Harris HMB**, **Coakley M**, **Lakshminarayanan B**, **O'sullivan O**, **Fitzgerald GF**, **Deane J**, **O'connor M**, **Harnedy N**, **O'connor K**, **O'mahony D**, **Van Sinderen D**, **Wallace M**, **Brennan L**, **Stanton C**, **Marchesi JR**, **Fitzgerald AP**, **Shanahan F**, **Hill C**, **Paul Ross R**, **O'toole PW**. 2012. Gut microbiota composition correlates with diet and health in the elderly. Nature **488**:178–184.
131. **Sewall JM**, **Oliver A**, **Denaro K**, **Chase AB**, **Weihe C**, **Lay M**, **Martiny JBH**, **Whiteson K**. 2020. Fiber Force: A Fiber Diet Intervention in an Advanced Course-Based Undergraduate Research Experience (CURE) Course †. J Microbiol Biol Educ.
132. **Merkel S**. 2012. The Development of Curricular Guidelines for Introductory Microbiology that Focus on Understanding. J Microbiol Biol Educ **13**:32–38.
133. **National Research Council**. 2003. BIO 2010: Transforming Undergraduate Education for Future Research Biologists. Education.
134. **American Association for the Advancement of Science**. 2009. No TitleVision and change in undergraduate biology education: a call to action: a summary of recommendations made at a national conference organized by the American Association for the Advancement of Science. Washington, DC.
135. **Junick J**, **Blaut M**. 2012. Quantification of Human Fecal Bifidobacterium Species by Use of Quantitative Real-Time PCR Analysis Targeting the groEL Gene. Appl Environ Microbiol **78**:2613–2622.
136. **Zhao G**, **Nyman M**, **Jönsson JÅ**. 2006. Rapid determination of short-chain fatty acids in colonic contents and faeces of humans and rats by acidified water-extraction and direct-injection gas chromatography. Biomed Chromatogr **20**:674–682.
137. **Nayfach S**, **Rodriguez-Mueller B**, **Garud N**, **Pollard KS**. 2016. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. Genome Res **26**:1612–1625.
138. **Nayfach S**, **Shi ZJ**, **Seshadri R**, **Pollard KS**, **Kyrpides NC**. 2019. New insights from uncultivated genomes of the global human gut microbiome. Nature **568**:505–510.
139. **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**.

2018. The MetaCyc database of metabolic pathways and enzymes. Nucleic Acids Res **46**:D633–D639.
140. **Hyatt D**, **Chen GL**, **LoCascio PF**, **Land ML**, **Larimer FW**, **Hauser LJ**. 2010. Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics.
141. **Finn RD**, **Bateman A**, **Clements J**, **Coggill P**, **Eberhardt RY**, **Eddy SR**, **Heger A**, **Hetherington K**, **Holm L**, **Mistry J**, **Sonnhammer ELL**, **Tate J**, **Punta M**. 2014. Pfam: the protein families database. Nucleic Acids Res **42**:D222–D230.
142. **EDDY SR**. 2009. A NEW GENERATION OF HOMOLOGY SEARCH TOOLS BASED ON PROBABILISTIC INFERENCE, p. 205–211. *In* Genome Informatics 2009. PUBLISHED BY IMPERIAL COLLEGE PRESS AND DISTRIBUTED BY WORLD SCIENTIFIC PUBLISHING CO.
143. **Lombard V**, **Golaconda Ramulu H**, **Drula E**, **Coutinho PM**, **Henrissat B**. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res.
144. **Camacho C**, **Coulouris G**, **Avagyan V**, **Ma N**, **Papadopoulos J**, **Bealer K**, **Madden TL**. 2009. BLAST+: architecture and applications. BMC Bioinformatics **10**:421.
145. **Oksanen J**, **Blanchet FG**, **Friendly M**, **Kindt R**, **Legendre P**, **McGlinn D**, **Minchin PR**, **O'Hara RB**, **Simpson GL**, **Solymos P**, **Stevens MHH**, **Szoecs E**, **Wagner H**. 2019. vegan: Community Ecology Package. R package version 2.5-2. Cran R.
146. **Zhang Z**, **Ersoz E**, **Lai C-Q**, **Todhunter RJ**, **Tiwari HK**, **Gore M a**, **Bradbury PJ**, **Yu J**, **Arnett DK**, **Ordovas JM**, **Buckler ES**, **Cho RJ**, **Mindrinos M**, **Richards DR**, **Sapolsky RJ**, **Anderson M**, **Drenkard E**, **Dewdney J**, **Reuber TL**, **Stammers M**, **Federspiel N**, **Theologis A**, **Yang WH**, **Hubbell E**, **Au M**, **Chung EY**, **Lashkari D**, **Lemieux B**, **Dean C**, **Lipshutz RJ**, **Ausubel FM**, **Davis RW**, **Oefner PJ**, **Bradbury PJ**, **Zhang Z**, **Kroon DE**, **Casstevens TM**, **Ramdoss Y**, **Buckler ES**, **Glaubitz JC**, **Casstevens TM**, **Lu F**, **Harriman J**, **Elshire RJ**, **Sun Q**, **Buckler ES**, **Lenné JM**, **Takan JP**, **Mgonja MA**, **Manyasa EO**, **Kaloki P**, **Wanyera N**, **Okwadi J**, **Muthumeenakshi S**, **Brown AE**, **Tamale M**, **Sreenivasaprasad S**, **Murray SC**, **Rooney WL**, **Hamblin MT**, **Mitchell SE**, **Kresovich S**, **Dida MM**, **Wanyera N**, **Dunn MLH**, **Bennetzen JL**, **Devos KM**, **Murray MG**, **Thompson WF**, **Kant S**, **Bi Y**, **Rothstein SJ**, **Crossa J**, **Burgueño J**, **Dreisigacker S**, **Vargas M**, **Herrera-Foessel SA**, **Lillemo M**, **Singh RP**, **Trethowan R**, **Warburton M**, **Franco J**, **Reynolds M**, **Crouch JH**, **Ortiz R**, **Yan WG**, **Li Y**, **Agrama HA**, **Luo D**, **Gao F**, **Lu X**, **Ren G**, **Yamori W**, **Kondo E**, **Sugiura D**, **Terashima I**, **Suzuki Y**, **Makino A**, **Goodstein DM**, **Shu S**, **Howson R**, **Neupane R**, **Hayes RD**, **Fazo J**, **Mitros T**, **Dirks W**, **Hellsten U**, **Putnam N**, **Rokhsar DS**, **Gupta, P. K., Roy, J. K. & Prasad M**, **Yadav S**, **Gaur VS**, **Jaiswal JPP**, **Kumar A**, **Furlotte N a**, **Eskin E**, **Gao H**, **Zhang T**, **Wu Y**, **Jiang L**, **Zhan J**, **Li J**, **Yang R**, **Altschul SF**, **Gish W**, **Miller W**, **Myers EW**, **Lipman DJ**, **Stothard P**, **Zhu C**, **Gore M a**, **Buckler ES**, **Yu J**, **Kalyana Babu B**, **Agrawal PK**, **Pandey D**, **Jaiswal JPP**, **Kumar A**, **Babu BK**, **Dinesh P**, **Agrawal PK**, **Sood S**, **Chandrashekara C**, **Bhatt JC**, **Kumar A**, **Begum H**, **Spindel JE**, **Lalusin A**, **Borromeo T**, **Gregorio G**, **Hernandez J**, **Virk P**, **Collard B**, **McCouch SR**, **Segura V**, **Vilhjálmsson BJ**, **Platt A**, **Korte A**, **Seren Ü**, **Long Q**, **Nordborg M**, **Sachidanandam R**, **Weissman D**, **Schmidt SC**, **Kakol JM**, **Stein LD**, **Marth G**, **Sherry S**, **Mullikin JC**, **Mortimore BJ**, **Willey DL**, **Hunt SE**, **Cole CG**, **Coggill PC**, **Rice CM**, **Ning Z**, **Rogers J**, **Bentley DR**, **Kwok PY**, **Mardis ER**, **Yeh RT**, **Schultz B**, **Cook L**, **Davenport R**, **Dante M**, **Fulton L**, **Hillier L**, **Waterston RH**,

McPherson JD, **Gilman B**, **Schaffner S**, **Van Etten WJ**, **Reich D**, **Higgins J**, **Daly MJ**, **Blumenstiel B**, **Baldwin J**, **Stange-Thomann N**, **Zody MC**, **Linton L**, **Lander ES**, **Altshuler D**, **Patterson HD**, **Williams ER**, **Rafalski A**, **R Development Core Team**, **Rott M**, **Martins NF**, **Thiele W**, **Lein W**, **Bock R**, **Kramer DM**, **Schöttler MA**, **Korte A**, **Vilhjálmsson BJ**, **Segura V**, **Platt A**, **Long Q**, **Nordborg M**, **Reimer S**, **Pozniak CJ**, **Clarke FR**, **Clarke JM**, **Somers DJ**, **Knox RE**, **Singh a K**, **Lu F**, **Lipka AE**, **Glaubitz JC**, **Elshire RJ**, **Cherney JH**, **Casler MD**, **Buckler ES**, **Costich DE**, **Kumar A**, **Sharma D**, **Tiwari A**, **Jaiswal JPP**, **Singh NK**, **Sood S**, **Gupta PK**, **Rustgi S**, **Panwar P**, **Jha AK**, **Pandey PK**, **Gupta AK**, **Kumar A**. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. Nat Genet.

147. **Truong DT**, **Tett A**, **Pasolli E**, **Huttenhower C**, **Segata N**. 2017. Microbial strain-level population structure & genetic diversity from metagenomes. Genome Res **27**:626–638.
148. **Wattam AR**, **Abraham D**, **Dalay O**, **Disz TL**, **Driscoll T**, **Gabbard JL**, **Gillespie JJ**, **Gough R**, **Hix D**, **Kenyon R**, **MacHi D**, **Mao C**, **Nordberg EK**, **Olson R**, **Overbeek R**, **Pusch GD**, **Shukla M**, **Schulman J**, **Stevens RL**, **Sullivan DE**, **Vonstein V**, **Warren A**, **Will R**, **Wilson MJC**, **Yoo HS**, **Zhang C**, **Zhang Y**, **Sobral BW**. 2014. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res.
149. **Sievers F**, **Wilm A**, **Dineen D**, **Gibson TJ**, **Karplus K**, **Li W**, **Lopez R**, **McWilliam H**, **Remmert M**, **Söding J**, **Thompson JD**, **Higgins DG**. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol.
150. **Stamatakis A**. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics.
151. **Bushnell B**. 2015. BBMap. https://sourceforge.net/projects/bbmap/.
152. **Kent WJ**. 2002. BLAT---The BLAST-Like Alignment Tool. Genome Res.
153. **Matsen FA**, **Kodner RB**, **Armbrust EV**. 2010. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics **11**:538.
154. **Jackson S**, **Cimentada J**, **Ruiz E**. corrr: Correlations in R. R package version 0.3.1.9000.
155. **Flint HJ**, **Scott KP**, **Duncan SH**, **Louis P**, **Forano E**. 2012. Microbial degradation of complex carbohydrates in the gut. Gut Microbes **3**.
156. **So D**, **Whelan K**, **Rossi M**, **Morrison M**, **Holtmann G**, **Kelly JT**, **Shanahan ER**, **Staudacher HM**, **Campbell KL**. 2018. Dietary fiber intervention on gut microbiota composition in healthy adults: a systematic review and meta-analysis. Am J Clin Nutr **107**:965–983.
157. **Nemoto H**, **Ikata K**, **Arimochi H**, **Iwasaki T**, **Ohnishi Y**, **Kuwahara T**, **Kataoka K**. 2011. Effects of fermented brown rice on the intestinal environments in healthy adult. J Med Investig **58**:235–245.
158. **Tap J**, **Furet J-P**, **Bensaada M**, **Philippe C**, **Roth H**, **Rabot S**, **Lakhdari O**, **Lombard V**, **Henrissat B**, **Corthier G**, **Fontaine E**, **Doré J**, **Leclerc M**. 2015. Gut microbiota richness promotes its stability upon increased dietary fibre intake in healthy adults. Environ Microbiol **17**:4954–4964.
159. **Cooper D**, **Kable M**, **Marco M**, **De Leon A**, **Rust B**, **Baker J**, **Horn W**, **Burnett D**, **Keim N**. 2017. The Effects of Moderate Whole Grain Consumption on Fasting Glucose and Lipids, Gastrointestinal Symptoms, and Microbiota. Nutrients **9**:173.
160. **Karl JP**, **Meydani M**, **Barnett JB**, **Vanegas SM**, **Goldin B**, **Kane A**, **Rasmussen H**,

Saltzman E, **Vangay P**, **Knights D**, **Chen C-YO**, **Das SK**, **Jonnalagadda SS**, **Meydani SN**, **Roberts SB**. 2017. Substituting whole grains for refined grains in a 6-wk randomized trial favorably affects energy-balance metrics in healthy men and postmenopausal women. Am J Clin Nutr **105**:589–599.

161. **Ampatzoglou A**, **Atwal KK**, **Maidens CM**, **Williams CL**, **Ross AB**, **Thielecke F**, **Jonnalagadda SS**, **Kennedy OB**, **Yaqoob P**. 2015. Increased Whole Grain Consumption Does Not Affect Blood Biochemistry, Body Composition, or Gut Microbiology in Healthy, Low-Habitual Whole Grain Consumers. J Nutr **145**:215–221.
162. **Carvalho-Wells AL**, **Helmolz K**, **Nodet C**, **Molzer C**, **Leonard C**, **McKevith B**, **Thielecke F**, **Jackson KG**, **Tuohy KM**. 2010. Determination of the in vivo prebiotic potential of a maize-based whole grain breakfast cereal: a human feeding study. Br J Nutr **104**:1353–1356.
163. **Costabile A**, **Klinder A**, **Fava F**, **Napolitano A**, **Fogliano V**, **Leonard C**, **Gibson GR**, **Tuohy KM**. 2008. Whole-grain wheat breakfast cereal has a prebiotic effect on the human gut microbiota: a double-blind, placebo-controlled, crossover study. Br J Nutr **99**:110–120.
164. **Gråsten SM**, **Juntunen KS**, **Mättö J**, **Mykkänen OT**, **El-Nezami H**, **Adlercreutz H**, **Poutanen KS**, **Mykkänen HM**. 2007. High-fiber rye bread improves bowel function in postmenopausal women but does not cause other putatively positive changes in the metabolic activity of intestinal microbiota. Nutr Res.
165. **Jenkins DJA**, **Vuksan V**, **Rao AV**, **Vidgen E**, **Kendall CWC**, **Tariq N**, **Würsch P**, **Koellreutter B**, **Shiwnarain N**, **Jeffcoat R**. 1999. Colonic bacterial activity and serum lipid risk factors for cardiovascular disease. Metabolism **48**:264–268.
166. **Ross AB**, **Bruce SJ**, **Blondel-Lubrano A**, **Oguey-Araymon S**, **Beaumont M**, **Bourgeois A**, **Nielsen-Moennoz C**, **Vigo M**, **Fay L-B**, **Kochhar S**, **Bibiloni R**, **Pittet A-C**, **Emady-Azar S**, **Grathwohl D**, **Rezzi S**. 2011. A whole-grain cereal-rich diet increases plasma betaine, and tends to decrease total and LDL-cholesterol compared with a refined-grain diet in healthy subjects. Br J Nutr **105**:1492–1502.
167. **Smith SC**, **Choy R**, **Johnson SK**, **Hall RS**, **Wildeboer-Veloo ACM**, **Welling GW**. 2006. Lupin kernel fiber consumption modifies fecal microbiota in healthy men as determined by rRNA gene fluorescent in situ hybridization. Eur J Nutr **45**:335–341.
168. **Zeng Y**, **Huang S**, **Mu G**, **Zeng X**, **Zhou X**. 2015. Effects of whole grain-bean mixed staple food on intestinal microecology and metabolic parameters of obese people. Chinese J Clin Nutr.
169. **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.
170. **De Filippo C**, **Cavalieri D**, **Di Paola M**, **Ramazzotti M**, **Poullet JB**, **Massart S**, **Collini S**, **Pieraccini G**, **Lionetti P**. 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc Natl Acad Sci U S A **107**:14691–6.
171. **Pallav K**, **Dowd SE**, **Villafuerte J**, **Yang X**, **Kabbani T**, **Hansen J**, **Dennis M**, **Leffler DA**, **Newburg DS**, **Kelly CP**. 2014. Effects of polysaccharopeptide from Trametes Versicolor and amoxicillin on the gut microbiome of healthy volunteers. Gut Microbes

5:458–467.
172. **Hooda S**, **Boler BMV**, **Serao MCR**, **Brulc JM**, **Staeger MA**, **Boileau TW**, **Dowd SE**, **Fahey GC**, **Swanson KS**. 2012. 454 Pyrosequencing Reveals a Shift in Fecal Microbiota of Healthy Adult Men Consuming Polydextrose or Soluble Corn Fiber. J Nutr **142**:1259–1265.
173. **Herman DR**, **Rhoades N**, **Mercado J**, **Argueta P**, **Lopez U**, **Flores GE**. 2020. Dietary Habits of 2- to 9-Year-Old American Children Are Associated with Gut Microbiome Composition. J Acad Nutr Diet **120**:517–534.
174. **Wastyk HC**, **Fragiadakis GK**, **Perelman D**, **Dahan D**, **Merrill BD**, **Yu FB**, **Topf M**, **Gonzalez CG**, **Robinson JL**, **Elias JE**, **Sonnenburg ED**, **Gardner CD**, **Sonnenburg JL**. 2020. Gut Microbiota-Targeted Diets Modulate Human Immune Status. bioRxiv 2020.09.30.321448.
175. **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.
176. **Graf E**, **Eaton JW**. 1985. Dietary suppression of colonic cancer fiber or phytate? Cancer **56**:717–718.
177. **Dinicola S**, **Minini M**, **Unfer V**, **Verna R**, **Cucina A**, **Bizzarri M**. 2017. Nutritional and Acquired Deficiencies in Inositol Bioavailability. Correlations with Metabolic Disorders. Int J Mol Sci **18**:2187.
178. **Belenguer A**, **Duncan SH**, **Calder AG**, **Holtrop G**, **Louis P**, **Lobley GE**, **Flint HJ**. 2006. Two routes of metabolic cross-feeding between Bifidobacterium adolescentis and butyrate-producing anaerobes from the human gut. Appl Environ Microbiol **72**:3593–3599.
179. **Rodriguez CI**, **Martiny JBH**. 2020. Evolutionary relationships among bifidobacteria and their hosts and environments. BMC Genomics **21**:26.
180. **Hu L**, **Lu W**, **Wang L**, **Pan M**, **Zhang H**, **Zhao J**, **Chen W**. 2017. Assessment of Bifidobacterium Species Using groEL Gene on the Basis of Illumina MiSeq High-Throughput Sequencing. Genes (Basel) **8**:336.
181. **Rossi M**, **Corradini C**, **Amaretti A**, **Nicolini M**, **Pompei A**, **Zanoni S**, **Matteuzzi D**. 2005. Fermentation of fructooligosaccharides and inulin by bifidobacteria: A comparative study of pure and fecal cultures. Appl Environ Microbiol **71**:6150–6158.
182. **Ramirez-Farias C**, **Slezak K**, **Fuller Z**, **Duncan A**, **Holtrop G**, **Louis P**. 2008. Effect of inulin on the human gut microbiota: stimulation of Bifidobacterium adolescentis and Faecalibacterium prausnitzii. Br J Nutr **101**:541–550.
183. **Topping DL**, **Clifton PM**. 2001. Short-Chain Fatty Acids and Human Colonic Function: Roles of Resistant Starch and Nonstarch Polysaccharides. Physiol Rev **81**:1031–1064.
184. **Cummings JH**. 1981. Short chain fatty acids in the human colon. Gut **22**:763–779.
185. **Ueyama J**, **Oda M**, **Hirayama M**, **Sugitate K**, **Sakui N**, **Hamada R**, **Ito M**, **Saito I**, **Ohno K**. 2020. Freeze-drying enables homogeneous and stable sample preparation for determination of fecal short-chain fatty acids. Anal Biochem **589**:113508.
186. **Rivière A**, **Selak M**, **Lantin D**, **Leroy F**, **De Vuyst L**. 2016. Bifidobacteria and butyrate-producing colon bacteria: Importance and strategies for their stimulation in the human gut. Front Microbiol.
187. **Sze MA**, **Topçuoğlu BD**, **Lesniak NA**, **Ruffin MT**, **Schloss PD**. 2019. Fecal Short-Chain

Fatty Acids Are Not Predictive of Colonic Tumor Status and Cannot Be Predicted Based on Bacterial Community Structure. MBio **10**.
188. **Vogt JA**, **Wolever TMS**. 2003. Fecal Acetate Is Inversely Related to Acetate Absorption from the Human Rectum and Distal Colon. J Nutr **133**:3145–3148.
189. **Sonnenburg ED**, **Sonnenburg JL**. 2014. Starving our Microbial Self: The Deleterious Consequences of a Diet Deficient in Microbiota-Accessible Carbohydrates. Cell Metab **20**:779–786.
190. **Sonnenburg JL**, **Sonnenburg ED**. 2019. Vulnerability of the industrialized microbiota. Science (80-). American Association for the Advancement of Science.
191. **Bowser BJ**. 2000. From Pottery to Politics : An Ethnoarchaeological Study of Political Factionalism , Ethnicity , and Domestic Pottery Style in the Ecuadorian Amazon. J Archaeol Method Theory **7**:219–248.
192. **Roach M**. 1998. The yuckiest food in the Amazon. Salon.
193. **Freire AL**, **Zapata S**, **Mosquera J**, **Mejia ML**, **Trueba G**. 2016. Bacteria associated with human saliva are major microbial components of Ecuadorian indigenous beers (chicha). PeerJ **2016**.
194. **Beghini F**, **McIver LJ**, **Blanco-Míguez A**, **Dubois L**, **Asnicar F**, **Maharjan S**, **Mailyan A**, **Thomas AM**, **Manghi P**, **Valles-Colomer M**, **Weingart G**, **Zhang Y**, **Zolfo M**, **Huttenhower C**, **Franzosa EA**, **Segata N**. 2020. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. bioRxiv. bioRxiv.
195. **Salazar G**. EcolUtils: Utilities for community ecology analysis. 0.1.
196. **Benoit G**, **Peterlongo P**, **Mariadassou M**, **Drezen E**, **Schbath S**, **Lavenier D**, **Lemaitre C**. 2016. Multiple comparative metagenomics using multiset k-mer counting. PeerJ Comput Sci **2016**:e94.
197. **Li D**, **Liu CM**, **Luo R**, **Sadakane K**, **Lam TW**. 2015. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics **31**:1674–1676.
198. **Li H**, **Handsaker B**, **Wysoker A**, **Fennell T**, **Ruan J**, **Homer N**, **Marth G**, **Abecasis G**, **Durbin R**, **1000 Genome Project Data Processing Subgroup 1000 Genome Project Data Processing**. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**:2078–9.
199. **Von Meijenfeldt FAB**, **Arkhipova K**, **Cambuy DD**, **Coutinho FH**, **Dutilh BE**. 2019. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. Genome Biol **20**:1–14.
200. **Finn RD**, **Clements J**, **Eddy SR**. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res **39**:W29–W37.
201. **Huang L**, **Zhang H**, **Wu P**, **Entwistle S**, **Li X**, **Yohe T**, **Yi H**, **Yang Z**, **Yin Y**. 2018. dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. Nucleic Acids Res **46**:D516–D521.
202. **Yu G**. 2020. Using ggtree to Visualize Data on Tree-Like Structures. Curr Protoc Bioinforma **69**.
203. **Mehta RS**, **Abu-Ali GS**, **Drew DA**, **Lloyd-Price J**, **Subramanian A**, **Lochhead P**, **Joshi AD**, **Ivey KL**, **Khalili H**, **Brown GT**, **Dulong C**, **Song M**, **Nguyen LH**, **Mallick H**, **Rimm EB**, **Izard J**, **Huttenhower C**, **Chan AT**. 2018. Stability of the human faecal microbiome in a cohort of adult men. Nat Microbiol **3**:347–355.

204. **Oliver A**, **Chase AB**, **Weihe C**, **Orchanian SB**, **Riedel SF**, **Hendrickson CL**, **Lay M**, **Sewall JM**, **Martiny JBH**, **Whiteson K**. 2021. High-Fiber, Whole-Food Dietary Intervention Alters the Human Gut Microbiome but Not Fecal Short-Chain Fatty Acids. mSystems **6**.
205. **Lloyd-Price J**, **Mahurkar A**, **Rahnavard G**, **Crabtree J**, **Orvis J**, **Brantley Hall A**, **Brady A**, **Creasy HH**, **Mccracken C**, **Giglio MG**, **Mcdonald D**, **Franzosa EA**, **Knight R**, **White O**, **Curtis Huttenhower &**. 2017. Strains, functions and dynamics in the expanded Human Microbiome Project. Nat Publ Gr **550**.
206. **Tett A**, **Huang KD**, **Asnicar F**, **Fehlner-Peach H**, **Pasolli E**, **Karcher N**, **Armanini F**, **Manghi P**, **Bonham K**, **Zolfo M**, **De Filippis F**, **Magnabosco C**, **Bonneau R**, **Lusingu J**, **Amuasi J**, **Reinhard K**, **Rattei T**, **Boulund F**, **Engstrand L**, **Zink A**, **Collado MC**, **Littman DR**, **Eibach D**, **Ercolini D**, **Rota-Stabelli O**, **Huttenhower C**, **Maixner F**, **Segata N**. 2019. The Prevotella copri Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations. Cell Host Microbe **26**:666-679.e7.
207. **Soto-Girón MJ**, **Peña-Gonzalez A**, **Hatt JK**, **Montero L**, **Páez M**, **Ortega E**, **Smith S**, **Cevallos W**, **Trueba G**, **Konstantinidis KT**, **Levy K**. 2021. Gut Microbiome Changes with Acute Diarrheal Disease in Urban Versus Rural Settings in Northern Ecuador. Am J Trop Med Hyg.
208. **Fragiadakis GK**, **Smits SA**, **Sonnenburg ED**, **Van Treuren W**, **Reid G**, **Knight R**, **Manjurano A**, **Changalucha J**, **Dominguez-Bello MG**, **Leach J**, **Sonnenburg JL**. 2019. Links between environment, diet, and the hunter-gatherer microbiome. Gut Microbes **10**:216–227.
209. **Geissinger O**, **Herlemann DPR**, **Morschel E**, **Maier UG**, **Brune A**. 2009. The ultramicrobacterium "Elusimicrobium minutum" gen. nov., sp. nov., the first cultivated representative of the termite group 1 phylum. Appl Environ Microbiol **75**:2831–2840.
210. **Méheust R**, **Castelle CJ**, **Matheus Carnevali PB**, **Farag IF**, **He C**, **Chen LX**, **Amano Y**, **Hug LA**, **Banfield JF**. 2020. Groundwater Elusimicrobia are metabolically diverse compared to gut microbiome Elusimicrobia and some have a novel nitrogenase paralog. ISME J **14**:2907–2922.
211. **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. Nature Publishing Group.
212. **Schnorr SL**, **Candela M**, **Rampelli S**, **Centanni M**, **Consolandi C**, **Basaglia G**, **Turroni S**, **Biagi E**, **Peano C**, **Severgnini M**, **Fiori J**, **Gotti R**, **De Bellis G**, **Luiselli D**, **Brigidi P**, **Mabulla A**, **Marlowe F**, **Henry AG**, **Crittenden AN**. 2014. Gut microbiome of the Hadza hunter-gatherers. Nat Commun **5**:1–12.
213. **Hansen MEB**, **Rubel MA**, **Bailey AG**, **Ranciaro A**, **Thompson SR**, **Campbell MC**, **Beggs W**, **Dave JR**, **Mokone GG**, **Mpoloka SW**, **Nyambo T**, **Abnet C**, **Chanock SJ**, **Bushman FD**, **Tishkoff SA**. 2019. Population structure of human gut bacteria in a diverse cohort from rural Tanzania and Botswana. Genome Biol **20**:1–21.
214. **De Filippo C**, **Cavalieri D**, **Di Paola M**, **Ramazzotti M**, **Poullet JB**, **Massart S**, **Collini S**, **Pieraccini G**, **Lionetti P**. 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc Natl Acad Sci U S A.
215. **Chen T**, **Long W**, **Zhang C**, **Liu S**, **Zhao L**, **Hamaker BR**. 2017. Fiber-utilizing

capacity varies in Prevotella- versus Bacteroides-dominated gut microbiota. Sci Rep **7**.
216. **De Vadder F**, **Kovatcheva-Datchary P**, **Zitoun C**, **Duchampt A**, **Bäckhed F**, **Mithieux G**. 2016. Microbiota-Produced Succinate Improves Glucose Homeostasis via Intestinal Gluconeogenesis. Cell Metab **24**:151–157.
217. **Scher JU**, **Sczesnak A**, **Longman RS**, **Segata N**, **Ubeda C**, **Bielski C**, **Rostron T**, **Cerundolo V**, **Pamer EG**, **Abramson SB**, **Huttenhower C**, **Littman DR**. 2013. Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. Elife **2**.
218. **Abranches J**, **Zeng L**, **Kajfasz JK**, **Palmer SR**, **Chakraborty B**, **Wen ZT**, **Richards VP**, **Brady LJ**, **Lemos JA**. 2018. Biology of Oral Streptococci. Microbiol Spectr **6**.
219. **Sprockett DD**, **Martin M**, **Costello EK**, **Burns AR**, **Holmes SP**, **Gurven MD**, **Relman DA**. 2020. Microbiota assembly, structure, and dynamics among Tsimane horticulturalists of the Bolivian Amazon. Nat Commun **11**:1–14.
220. **Colehour AM**, **Meadow JF**, **Liebert MA**, **Cepon-Robins TJ**, **Gildner TE**, **Urlacher SS**, **Bohannan BJM**, **Snodgrass JJ**, **Sugiyama LS**. 2014. Local domestication of lactic acid bacteria via cassava beer fermentation. PeerJ **2014**:e479.
221. **Elizaquível P**, **Pérez-Cataluña A**, **Yépez A**, **Aristimuño C**, **Jiménez E**, **Cocconcelli PS**, **Vignolo G**, **Aznar R**. 2015. Pyrosequencing vs. culture-dependent approaches to analyze lactic acid bacteria associated to chicha, a traditional maize-based fermented beverage from Northwestern Argentina. Int J Food Microbiol **198**:9–18.
222. **Puerari C**, **Magalhães-Guedes KT**, **Schwan RF**. 2015. Physicochemical and microbiological characterization of chicha, a rice-based fermented beverage produced by Umutina Brazilian Amerindians. Food Microbiol.
223. **Faria-Oliveira F**, **Diniz RHS**, **Godoy-Santos F**, **Piló FB**, **Mezadri H**, **Castro IM**, **Brandão RL**. 2015. The Role of Yeast and Lactic Acid Bacteria in the Production of Fermented Beverages in South America, p. 13. *In* Food Production and Industry. InTech.
224. **Brewster R**, **Tamburini FB**, **Asiimwe E**, **Oduaran O**, **Hazelhurst S**, **Bhatt AS**. 2019. Surveying Gut Microbiome Research in Africans: Toward Improved Diversity and Representation. Trends Microbiol. Elsevier Ltd.
225. **Dubinkina VB**, **Ischenko DS**, **Ulyantsev VI**, **Tyakht A V.**, **Alexeev DG**. 2016. Assessment of k-mer spectrum applicability for metagenomic dissimilarity analysis. BMC Bioinformatics **17**:1–11.
226. **Wibowo MC**, **Yang Z**, **Borry M**, **Hübner A**, **Huang KD**, **Tierney BT**, **Zimmerman S**, **Barajas-Olmos F**, **Contreras-Cubas C**, **García-Ortiz H**, **Martínez-Hernández A**, **Luber JM**, **Kirstahler P**, **Blohm T**, **Smiley FE**, **Arnold R**, **Ballal SA**, **Pamp SJ**, **Russ J**, **Maixner F**, **Rota-Stabelli O**, **Segata N**, **Reinhard K**, **Orozco L**, **Warinner C**, **Snow M**, **LeBlanc S**, **Kostic AD**. 2021. Reconstruction of ancient microbial genomes from the human gut. Nature.
227. **Jacobson DK**, **Honap TP**, **Ozga AT**, **Meda N**, **Kagoné TS**, **Carabin H**, **Spicer P**, **Tito RY**, **Obregon-Tito AJ**, **Reyes LM**, **Troncoso-Corzo L**, **Guija-Poma E**, **Sankaranarayanan K**, **Lewis CM**. 2021. Analysis of global human gut metagenomes shows that metabolic resilience potential for short-chain fatty acid production is strongly influenced by lifestyle. Sci Rep.
228. **Rabesandratana T**. 2018. Microbiome conservancy stores global fecal samples. Science (80-) **362**:510–511.
229. **Groussin M**, **Poyet M**, **Sistiaga A**, **Kearney SM**, **Moniz K**, **Noel M**, **Hooker J**,

Gibbons SM, **Segurel L**, **Froment A**, **Mohamed RS**, **Fezeu A**, **Juimo VA**, **Lafosse S**, **Tabe FE**, **Girard C**, **Iqaluk D**, **Nguyen LTT**, **Shapiro BJ**, **Lehtimäki J**, **Ruokolainen L**, **Kettunen PP**, **Vatanen T**, **Sigwazi S**, **Mabulla A**, **Domínguez-Rodrigo M**, **Nartey YA**, **Agyei-Nkansah A**, **Duah A**, **Awuku YA**, **Valles KA**, **Asibey SO**, **Afihene MY**, **Roberts LR**, **Plymoth A**, **Onyekwere CA**, **Summons RE**, **Xavier RJ**, **Alm EJ**. 2021. Elevated rates of horizontal gene transfer in the industrialized human microbiome. Cell **184**:2053-2067.e18.

230. 2019. World Urbanization Prospects: The 2018 RevisionWorld Urbanization Prospects: The 2018 Revision. UN.

231. **Sonnenburg ED**, **Sonnenburg JL**. 2019. The ancestral and industrialized gut microbiota and implications for human health. Nat Rev Microbiol **17**:383–390.