UNIVERSITY OF CALIFORNIA SAN DIEGO

Temporal Microbiome Dynamics: From Circadian Rhythms to Probiotic Evolution

A dissertation submitted in partial satisfaction

of the requirements for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Celeste Frances Allaband

Committee in charge:

Professor Rob Knight, Co-Chair
Professor Amir Zarrinpar, Co-Chair
Professor Susan Golden
Professor Victor Nizet
Professor Bernhard Palsson

2022

The dissertation of Celeste Frances Allaband is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

I would like to dedicate this work to people who take non-traditional paths as well as the many amazing women in STEM.

EPIGRAPH

“Progress is made by trial and failure; the failures are generally a hundred times more numerous than the successes; yet they are usually left unchronicled.”

— **William Ramsay** (1852-1916)

“What you do makes a difference, and you have to decide what kind of difference you want to make.”

— **Jane Goodall**

“When you believe you have found an important scientific fact, and are feverishly curious to publish it, constrain yourself for days, weeks, years sometimes, fight yourself, try and ruin your own experiments, and only proclaim your discovery after having exhausted all contrary hypotheses. But when, after so many efforts you have at last arrived at a certainty, your joy is one of the greatest which can be felt by a human soul.”

— Louis Pasteur. A speech given at the Pasteur Institute’s inauguration (14 November, 1888).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Firstly, I would like to acknowledge my many mentors who shepherded me during the course of my graduate studies at UC San Diego, including Rob Knight, Amir Zarrinpar, Victor Nizet, Susan Golden, Bernhard Palsson, and Peter Ernst. Each of these individuals met with me on a regular basis and guided and/or influenced much of the work that is presented in this dissertation. Rob and Amir were especially kind after the death of my father, my favorite cat, my uncle, and the suicide of one of my veterinary mentors.

Secondly, I would like to thank the leadership of the Biomedical Sciences Program (BMS), Arshad Desai, Kevin Corbett, and Åsa Gustafsson, for believing and investing in me. I also want to thank all of the administrative support team of BMS, including Leanne Nordeman and Patricia Leutmer, who have provided a constant source of encouragement and practical help over the many years. My BMS classmates have also been a wonderful source of camaraderie and friendship during these years, especially Tiani Louis, Rebecca Lau, Gabriela Goldberg, Margaret Burns, Alanna Koehler, Dominic McGrosso, and Chelsea Blankenship.

Thirdly, I would like to acknowledge Peter Ernst, my mentor in the UCSD Research Training Program for Veterinarians. Without his help and guidance, I would not have obtained the National Institutes of Health T32 grant that funded much of the work presented in this dissertation. I would also like to acknowledge the tireless efforts of Lanisa Medina, my grant coordinator for the T32 who assisted in all of the associated paperwork and administration. Of course, related to these points, I would also like to acknowledge the generous funding of the Office of the Director, National Institutes of Health that made this DVM-specific grant possible in the first place. My veterinary colleagues have given me a source of community and connection to my first profession and have become friends. I especially want to acknowledge Katherine Makowski, Cassandra Tang Wing, and Kent Osborn.

Fourthly, this work could not have been completed without the strong support of my colleagues in both the Zarrinpar and Knight laboratories. I am extremely grateful for those who constantly managed the finances, administrative logistics, computational hardware, and more, including Jeff DeReus, Michiko Souza, Gail Ackermann, Yna Villanueva, Jerry Kennedy, and Peggy Castaneda. I would also like the thank friends, collaborators, and informal mentors within both labs who shaped my skill set, provided feedback on my work, sparked creativity, and further protected me from going astray at various points; they are some of the most intelligent and hardworking people I know: (Knight Lab) Cameron Martino, Justin Shaffer, Daniel McDonald, Gibs Rahman, Victor Cantu, George Armstrong, Caitlin Guccione, Hazel Dilmore, Omar Din, Greg Humphrey, MacKenzie Bryant, Priya Tripathi, Antonio Gonzalez, Shi Huang, Qiyun Zhu, Franck Lejzerowicz, Greg Sepich-Poore, Jake Minich, and Jamie Morton; (Dorrestein Lab) Kathleen Dorrestein, Pieter Dorrestein, Allegra Aron, and Emmanuel Elijah; (Zarrinpar Lab) Baylee Russell, Maria D. Tiu, Erica Maissy, Irene Mai, Ana Carolina Dantas Machado, Stephany Flores-Ramos, Nicole Siguenza, and Alex Richter; (Palsson lab) Richard Szubin. I want to make special mention of Joe Pogliano, who listened and gave me great advice during a difficult time.

Lastly, I want to thank my many friends near and far —Sarah Navarro, Kathleen Navarro, Sarah Nelson, Maricarmen Perez Rodriguez, Tracy Keppel-Kolb, Brittany McCauslin, Claire Northall, Jaime Zlamal, Jungwoo Lee, and Tyler Poston—for their constant love, support, and encouragement during various seasons of this doctoral work, particularly during the COVID-19 pandemic. They have fostered and deepened my curiosity, creativity, daring, and joy.

Chapter 1, in full, is a reprint of the material as it appears in “Microbiome 101: Studying, Analyzing, and Interpreting Gut Microbiome Data for Clinicians.” Celeste Allaband, Daniel McDonald, Yoshiki Vázquez-Baeza, Jeremiah J Minich, Anupriya Tripathi, David A Brenner, Rohit Loomba, Larry Smarr, William J Sandborn, Bernd Schnabl, Pieter Dorrestein, Amir Zarrinpar, and

Rob Knight. Clinical Gastroenterology and Hepatology, 17 (2) , 218-230. The dissertation author was the primary investigator and first author of this paper.

Chapter 2, in full, is a reprint of the material as it appears in, “Intermittent Hypoxia and Hypercapnia Alter Diurnal Rhythms of Luminal Gut Microbiome and Metabolome.” Celeste Allaband, Amulya Lingaraju, Cameron Martino, Baylee Russell, Anupriya Tripathi, Orit Poulsen, Ana Carolina Dantas Machado, Dan Zhou, Jin Xue, Emmanuel Elijah, Atul Malhotra, Pieter C Dorrestein, Rob Knight, Gabriel G Haddad, and Amir Zarrinpar. mSystems, e0011621. The dissertation author was the primary investigator and first author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in “Influence of intermittent hypoxia/hypercapnia on atherosclerosis, gut microbiome, and metabolome.” Jin Xue, Celeste Allaband, Dan Zhou, Orit Poulsen, Cameron Martino, Lingjing Jiang, Anupriya Tripathi, Emmanuel Elijah, Pieter C Dorrestein, Rob Knight, Amir Zarrinpar, and Gabriel G Haddad. Frontiers in Physiology, 12:663950. The dissertation author was a primary investigator and a first author of this paper.

Chapter 4 has been submitted for publication of the material as it may appear in *Nature*, “Microbiome Diurnal Dynamics Dominate Phenotype Effects, Enabling Replicability When Controlled.” Celeste Allaband, Amulya Lingaraju, Stephany Flores-Ramos, Tanya Kumar, Haniyeh Javaheri, Maria D. Tiu, Ana Carolina Dantas Machado, Roland A. Richter, Emmanuel Elijah, Gabriel G. Haddad, Vanessa A. Leone, Pieter C. Dorrestein, Rob Knight, and Amir Zarrinpar. The dissertation author was the primary investigator and first author of this paper.

Chapter 5 is the unsubmitted manuscript for my final project, “Engineered Native Bacteria are More Efficient at Engraftment and Colonization of a Host and Require Fewer Mutations for Maintenance Long Term.” Celeste Allaband, Maria D. Tiu, Omar Din, Berhard Palsson, Rob Knight, and Amir Zarrinpar. The dissertation author was the primary investigator and first author of this paper.

VITA

2008 B.S., Biological Sciences with an emphasis in Medical Microbiology with Honors, University of California Davis

2014 D.V.M., Colorado State University

2022 Ph. D., Biomedical Sciences, University of California San Diego

PUBLICATIONS

Author names marked with † indicate shared first co-authorship.

Allaband C, Tiu MD, (…) Knight R, Zarrinpar A. (2022). "Engineered Native Bacteria are More Efficient at Engraftment and Colonization of a Host and Require Less Mutations for Maintenance Long Term." *Nature*. MANUSCRIPT IN PROCESS

Allaband C, Lingaraju A, Flores-Ramos S, Kumar T, Javaheri H, Tiu MD, Dantas Machado AC, Richter RA, Elijah E, Haddad GG, Leone VA, Dorrestein PC, Knight R, Zarrinpar A. (2022). "Microbiome Diurnal Dynamics Dominate Phenotype Effects, Enabling Replicability When Controlled." *Cell*. UNDER REVIEW

Allaband C, Lingaraju A, Martino C, Russell B, Tripathi A, Poulsen O, Dantas Machado AC, Zhou D, Xue J, Elijah E, Malhotra A, Dorrestein PC, Knight R, Haddad GG & Zarrinpar A. (2021). "Intermittent Hypoxia and Hypercapnia Alter Diurnal Rhythms of Luminal Gut Microbiome and Metabolome." *mSystems*, e0011621. https://doi.org/10.1128/mSystems.00116-21

Xue J†, **Allaband C†**, Zhou D, Poulsen O, Martino C, Jiang L, Tripathi A, Dorrestein PC, Knight R, Zarrinpar A and Haddad GG. (2021). "Influence of Intermittent Hypoxia/Hypercapnia on Atherosclerosis, Gut Microbiome and Metabolome." *Frontiers in Physiology*, 12:663950. https://doi.org/10.3389/fphys.2021.663950

Allaband C, McDonald D, Vazquez-Baeza Y, Minich JJ, Tripathi A, Brenner DA, Loomba R, Smarr L, Sandborn WJ, Schnabl B, Dorrestein P, Zarrinpar A, Knight R. (2019). "Microbiome 101: Studying, Analyzing, and Interpreting Gut Microbiome Data for Clinicians." *Clinical Gastroenterology and Hepatology*, 17 (2) , 218-230. https://doi.org/10.1016/j.cgh.2018.09.01.

The following publications were not included as part of this dissertation, but were also significant byproducts of my doctoral training.

Martino C, McDonald D, Cantrell K, Dilmore AH, Vázquez-Baeza Y, Shenhav L, Shaffer JP, Rahman G, Armstrong G, **Allaband C**, Song SJ, Knight R. (2022) Compositionally Aware Phylogenetic Beta-Diversity Measures Better Resolve Microbiomes Associated with Phenotype. mSystems. Apr 28:e0005022. https://doi.org/10.1128/msystems.00050-22 PMID: 35477286

Huang S, Jiang S, Huo D, **Allaband C**, Estaki M, Cantu V, Belda-Ferre P, Vázquez-Baeza Y, Zhu Q, Ma C, Li C, Zarrinpar A, Liu YY, Knight R, & Zhang J. (2021). "Candidate probiotic *Lactiplantibacillus plantarum* HNU082 rapidly and convergently evolves within human, mice, and zebrafish gut but differentially influences the resident microbiome." *Microbiome*, 9(1), 151. https://doi.org/10.1186/s40168-021-01102-0.

Huang S, Haiminen N, Carrieri A-P, Hu R, Jiang L, Parida L, Russell B, **Allaband C**, Zarrinpar A, Vázquez-Baeza Y, Belda-Ferre P, Zhou H, Kim H-C, Swafford AD, Knight R, Xu ZZ. (2020). "Human skin, oral, and gut microbiomes predict chronological age." *mSystems* 5:e00630-19. https://doi.org/10.1128/mSystems.00630-19.

Song SJ, Woodhams DC, Martino C, **Allaband C**, Mu A, Javorschi-Miller-Montgomery S, Sucholdolski JS, Knight, R. (2019). "Engineering the microbiome for animal health and conservation." *Experimental Biology and Medicine*, 244(6), 494–504. https://doi.org/10.1177/1535370219830075.

Williams CL, Caraballo-Rodríguez AM, **Allaband C**, Zarrinpar A, Knight R, Gauglitz JM. (2019). "Wildlife-microbiome interactions and disease: exploring opportunities for disease mitigation across ecological scales." *Drug Discovery Today: Disease Models*, 28, 105-115. https://doi.org/10.1016/j.ddmod.2019.08.012.

ABSTRACT OF THE DISSERTATION

Microbiome Dynamics: From Circadian Rhythms to Probiotic Evolution

by

Celeste Frances Allaband

Doctor of Philosophy in Biomedical Sciences

University of California San Diego, 2022

Professor Rob Knight, Co-Chair

Professor Amir Zarrinpar, Co-Chair

Microbial life is all around us and inside us in an unseen vast and varied ecosystem. For the past several decades, the microbiome field has attempted to characterize this microscopic web of life using next generation sequencing technology. It turns out that the number of host organism cells is generally outnumbered by the number of microbial cells that live on and within them. Moreover, these microbes collectively have two orders of magnitude more genes than their

host organism. While the host organism's genome stays constant during its life, the microbial community and its associated genes change much more frequently. Since microbial lifespans are sometimes measured in minutes, selective pressures in the host micro-environments can rapidly affect the evolution and gene content of the microorganisms present. Over the course of a lifetime, the host winds up with a unique set of microbes selected to fit their unique individual environment, from immune system quirks to specialized diets to biological rhythms. However, despite biological rhythms having always played a role in host behavior and survival, the impact of time on health has been a widely neglected realm until recently. The basic argument of this thesis is that time is an underappreciated but critical variable that impacts both experimental study design as well as host health.

Altogether, these chapters and analyses comprehensively characterize the impact of time and biological rhythms on the microbiome and host health, experimental study design, as well as microbial evolution.

Chapter 1. Microbiome 101: Studying, Analyzing, and Interpreting Gut Microbiome Data for Clinicians

Abstract

Advances in technical capabilities for reading out complex human microbiomes are leading to an explosion of microbiome research, leading in turn to intense interest among clinicians in applying these techniques to their patients. In this review, we discuss the content of the human microbiome, including inter- and intra-subject variability, considerations of study design including important confounding factors, and different methods in the laboratory and on the computer to read out the microbiome and its resulting gene products and metabolites. We highlight several common pitfalls for clinicians, including the expectation that an individual's microbiome will be stable, that diet can induce rapid changes that are large compared to the differences among subjects, that everyone has essentially the same core stool microbiome, and that different laboratory and computational methods will yield essentially the same results. We also highlight the current limitations and future promise of these techniques, with the expectation that an understanding of these considerations will help accelerate the path towards routine clinical application of these techniques developed in research settings.

1.1 Introduction

In the first part of this thesis, we comprehensively outline the literature landscape of microbiome data as is relevant to diagnostic and clinical applications. Despite the interest and prevalence in products claiming to improve the microbiome, there are still many limitations in study design, analysis, and interpretation of microbiome data that prevent immediate clinical implementation. We highlight several common misconceptions, including that there is a "normal"

or single ideal “core” microbiome, that microbiomes are stable over an individual’s life, and that diet induces large rapid changes to microbial communities. We describe several different methods used in research to characterize microbiome samples as well as computational and analytic pipelines. We also provide an example of a report provided to individuals that participate in the citizen science project, Microsetta (formerly the American Gut Project). Finally, we discuss the future promise of microbiome data as we move toward routine clinical applications of research methods.

Interest in the microbiome is at an all-time high, with the microbiome connected to an increasing range of diseases of interest to gastroenterologists and hepatologists. For example, obesity 1,2,3,4, inflammatory bowel disease5,6,7, alcoholic and nonalcoholic fatty liver disease8,9,10, and hepatocellular carcinoma11,12,13,14 have all been linked to the microbiome in humans, and changes in the microbiome have been shown to induce or modify these diseases in animal models. However, moving this linkage to a clinical relevant diagnostic is still in the research phase.

We have seen enormous progress in the last decade in using genomic sequencing coupled with computational pipelines to decipher the human gut microbiome15. These tools are necessary because of the incredible information density of the microbiome. Each teaspoon of stool contains in its bacterial DNA alone the amount of data that it would take 100,000 of today’s highest-capacity thumb drives to store. (This number was reached using the following calculations: 1 gram of stool contains 100 billion microbes16. We then assume 5 million bases per microbe (~1 million bytes), which then yields 10E11x10E6=10E17 or 100,000 Terabytes (Tb). The current highest capacity thumb drive size is 1 Tb, so at approx. 12g each is 1200kg (1.3 tons), about the weight of a young giraffe.) This information is also dynamic, as the microbiome profile changes with diet and medical interventions. These problems create challenges for clinicians in deciding whether it will be medically informative to ask a patient to collect stool or colonoscopy

biopsies and send it off for sequencing. Interpreting and discussing the results with patients can be challenging, especially with a lack of standard parameters and reference data for comparison.

In this review, we cover what the microbiome is, how it can be collected, what molecular methods can be used to analyze it, how the data can be interpreted, and what are some of the limitations in combining conclusions from different studies. Our goal is to highlight which areas are solid, which areas are emerging, and where the greatest potential is for future work to provide actionable information that benefits patients.

1.2 What is the microbiome?

The human gut is home to a variety of microbes, including bacteria, archaea (single-celled organisms without nuclei that are more closely related to eukaryotes than to bacteria), fungi (mostly yeasts), microbial eukaryotes (usually *Blastocystis* in the US, but a variety of pathogenic and non-pathogenic taxa in developing countries), and viruses/phages. This collection of microbes is called the microbiota; their genes are called the microbiome[17]. However, the term "microbiome" has come in popular usage to refer to the microbes themselves. Whether the microbiome includes the virome (the repertoire of viral genes) is open to debate. Because of the technical ease and widespread utility of approaches that just read out the bacteria (see below), many have concluded that the microbiome refers only to the bacteria, but this is not correct. Rather, if a difference is demonstrated in the bacterial compartment of the microbiome between cases and controls, it is necessarily true that the microbiome is different; however, if no difference in the bacteria is found, there might still be a difference in other kinds of microbes, e.g. the yeast or viruses.

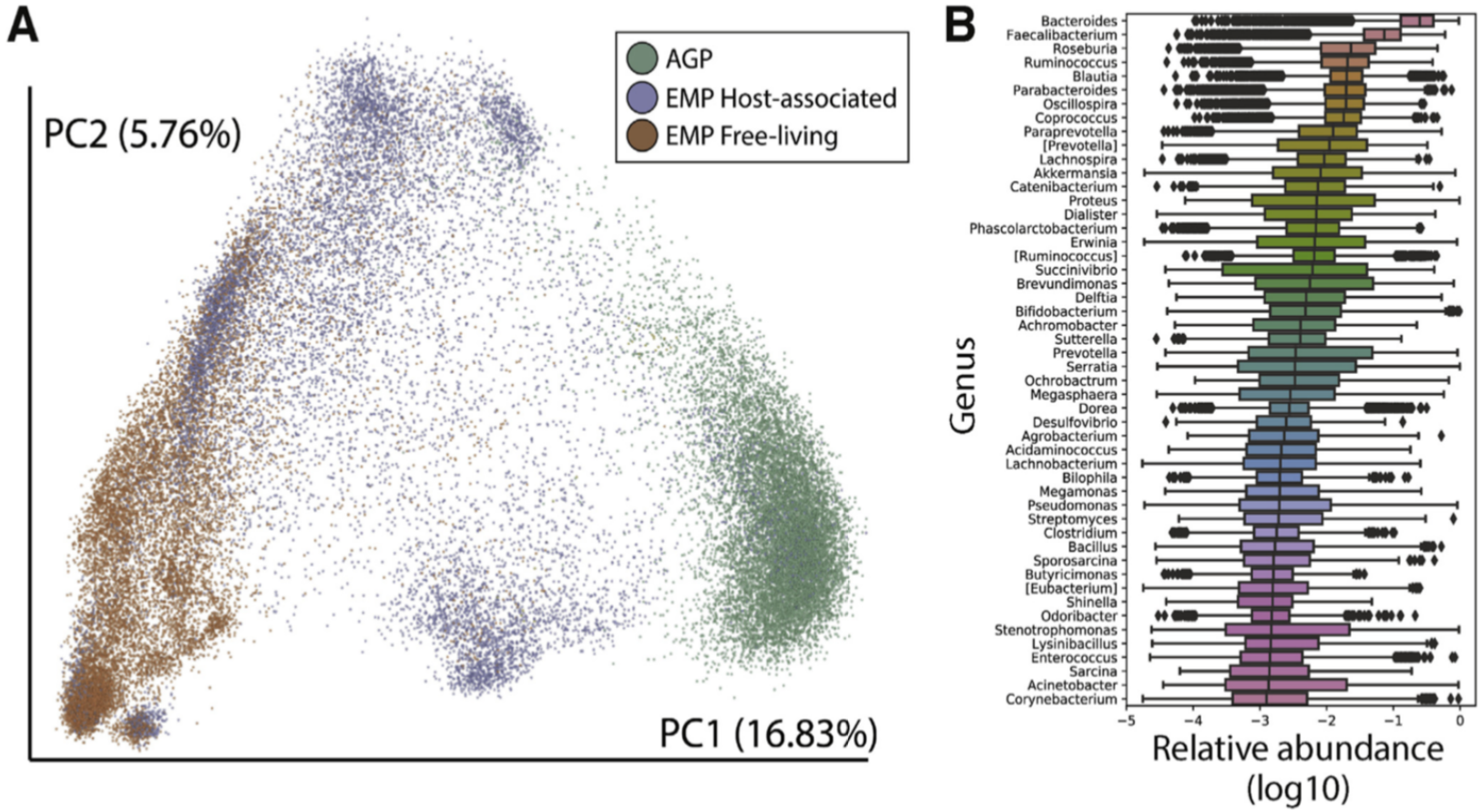
Until recently, a frequently repeated slogan was that the human microbiome contained ten times as many cells as the human body. This figure was based on a back-of-the-envelope calculation 40 years ago[18], and the correct claim was that the true figure was somewhere between

1:10 and 10:1, but could be as much as 10:1. Since then, the errors on the estimates of the number of human cells and microbial cells have narrowed considerably, with the true figure being much closer to 1:1, with the balance slightly in favor of the microbes[16]. Therefore, it is fascinating to consider that one can tip the balance in an individual's body from having more microbial cells to having more human cells by simply administering the bowel prep for a colonoscopy. In this article, we will focus on the gut microbiome, although microbiomes in other parts of the body (e.g. the skin, mouth, and vagina) are also important for health and in numerous diseases.

In most healthy humans, the gut microbiome is dominated in cellular relative abundance by bacteria, specifically members of the phyla Bacteroidetes and Firmicutes, with only small amounts of non-bacterial microbes. It is important to recognize that among healthy people, their percentage of each of these two dominant phyla can vary from 10 to 90%, even though the combined percentage tends to be around 95%[19]. However, some individuals, particularly in the disease state, can have large percentages of other bacterial phyla, such as Proteobacteria (which E. coli belongs to), Verrucomicrobia, Actinobacteria, or Fusobacteria.

The earliest culture-independent projects revealed that different people can differ greatly from one another in terms of their microbiomes[20,1,21,19], and the diversity spanned by human stool is comparable to the diversity spanned by completely different kinds of environments in the Earth Microbiome Project (Figure. 1A). In fact, some bacterial species that were as abundant as 5% of the total in one individual, turned out to be no more abundant than 0.01% in another individual, even in a small cohort[21]. We see the same in the American Gut Project data (Fig. 1.1B). Therefore, there is no "standard" microbiome ecology that all healthy people share. However, because of this high variability among individuals, extreme caution must be taken in interpreting results from fewer than hundreds of people, and the reference range approach that has worked for blood tests will not work for the microbiome[22,23].

Figure 1.1. Intersubject variability of the gut microbiome. (A) A principal coordinates plot of unweighted UniFrac distances computed using the Earth Microbiome Project dataset24 and the fecal samples from the American Gut Project dataset25. Even though the EMP data include samples from many of the environments on the planet, including hydrothermal vents, soils, marine sediment, and many others, the extent of diversity associated with just the large intestine of a single mammal is one of the dominating clusters of microbial diversity. (B) Dynamic ranges of the 50 most abundant genera in the human fecal microbiome from 9316 individuals. These data are based off of a single sample per person, and only considering organisms observed in at least 100 people. Even though Bacteroides are ranked the highest, there are individuals with up to three orders of magnitude lower relative abundance of those genera, and that genera was not detected in approximately 1% of the individuals.


A
AGP
EMP Host-associated
EMP Free-living
PC2 (5.76%)
PC1 (16.83%)
B
Genus
Bacteroides
Faecalibacterium
Roseburia
Ruminococcus
Blautia
Parabacteroides
Oscillospira
Coprococcus
Paraprevotella
[Prevotella]
Lachnospira
Akkermansia
Catenibacterium
Proteus
Dialister
Phascolarctobacterium
Erwinia
[Ruminococcus]
Succinivibrio
Brevundimonas
Delftia
Bifidobacterium
Achromobacter
Sutterella
Prevotella
Serratia
Ochrobactrum
Megasphaera
Dorea
Desulfovibrio
Agrobacterium
Acidaminococcus
Lachnobacterium
Bilophila
Megamonas
Pseudomonas
Streptomyces
Clostridium
Bacillus
Sporosarcina
Butyricimonas
[Eubacterium]
Shinella
Odoribacter
Stenotrophomonas
Lysinibacillus
Enterococcus
Sarcina
Acinetobacter
Corynebacterium
−5
−4
−3
−2
−1
0
Relative abundance
(log10)

1.3 What is the best way to collect a sample for microbiome analysis?

The first topic a clinician faces is: what is the optimal protocol for collecting a microbiome sample for analysis? There is still an ongoing debate on the best way to collect and store a sample for analysis of the microbiome. In short, there is no perfect method because the choice will depend on feasibility, cost, patient acceptance, and which methods will be used to read out the microbiome downstream.

The first important question is what to sample. Stool is by far the most accessible material, and can be collected as often as your subjects produce stool, enabling longitudinal studies (e.g. of daily samples) that would not be feasible with biopsy specimens. For studying gastrointestinal and liver diseases, assessing the gut microbiome using stool provides a unique opportunity to study pathophysiology and disease states in both cross-sectional and longitudinal study designs. However, stool does not capture all the microbes in the gut[20,26], and in particular mucosally-adherent microbes and microbes in the small intestine, particularly the ileum, can be missed. In addition, stool is often quite distant from the GI region of the pathology being investigated, and has been stored in the rectum, where there is active dehydration and where fermentation selects for bacteria that are not commonly found in other parts of the lumen. This implies that it is difficult to use the stool microbiome to understand the pathophysiology of a disease, because it likely fails to reflect the microbiome of the region of pathology, and it is imperative to choose a sample collection method that is inherently consistent with the scientific or clinical question being asked.

"Culturomics" approaches[27], in which large numbers of cells are isolated and cultured, show that metagenomics approaches miss many rare bacteria that are not well represented in the reference databases or that are below the filtering thresholds used to eliminate noise (see below). They also suggest that even the most aggressive homogenizing procedure to break bacterial cells walls may still miss important organisms. On the other hand, some 85% of microbes in the human gut are anaerobic and therefore do not culture in an open petri dish, although they

can be grown in research laboratory anaerobic chambers. However, despite advances in culturing methods[28,29], what can be cultured is still biased, especially because any given culture condition will allow some bacteria to grow much faster than others.

In spite of these limitations, the gold standard protocol for stool sampling is to collect the whole stool, homogenize it immediately (e.g. with a blender or a tissue homogenizer), then flash freeze the homogenate in liquid nitrogen or in dry ice/ethanol slurry, with an aliquot preserved in 20% glycerol in LB for culturing. Nucleic acid protectors such as RNAlater, although popular, have had mixed success in different laboratories, and render the sample unsuitable for metabolomics, so should be used on a separate aliquot. However, this protocol is expensive and often impractical, especially given the limitations inherent in subjects' ability to produce stool on demand. Although stool is not homogeneous, in general the differences between whole stool and a small sample of stool are small compared to the differences between individuals. While stool consistency is correlated with microbiome changes[30], stool consistency does not interfere with DNA extraction in people with chronic GI conditions such as IBS, IBD, and constipation.

For DNA analyses, several studies have shown that Flinders Technology Associate (FTA) and fecal occult blood test (FOBT) cards are stable at room temperature for at least days[31,32,33], and although they induce small, systematic shifts in the resulting taxon profiles compared to flash-frozen samples, the practical ease of use of these methods is a considerable attraction. Another widely used method is dry swabs of fecal material left behind on bathroom tissue, such as those used in the American Gut Project[25], can be used for amplicon analysis (e.g. for 16S rRNA gene profiling by PCR, see below) with appropriate filtering for overgrowth, but are problematic for shotgun metagenomics, and only cotton-based swabs, not polyester-based, can be used for metabolomics because of issues with polymers. An important practical consideration for using swabs in the mail is that polycarbonate housings are not nearly as robust to the vagaries of mail

handling as polyethylene, and require padded envelopes to arrive intact. However, despite these limitations, dry swabs from bathroom tissue have yielded useful results in many studies[34,35,25].

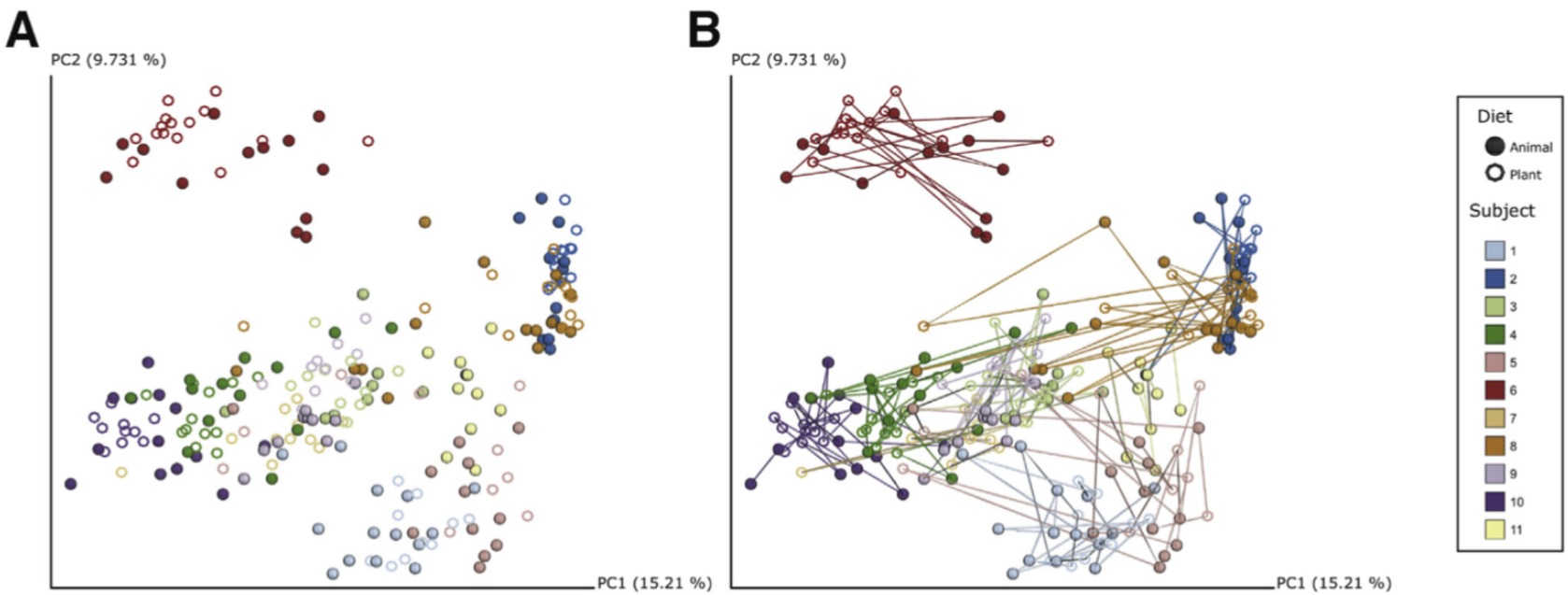
Going beyond the stool, many studies have demonstrated that the mucosa and lumen differ in their microbiomes from each other at a given site in the gut[36,37], and that the microbiome varies dramatically along the length of the gut, with the stomach and small intestine being essentially entirely distinct from the large intestine. More subtle inter-sample variations are then found within the small intestine and within the large intestine. This raises the question of where one should look for microbiome associations. However, practically speaking, getting biopsy samples from the small intestine is quite challenging clinically, and obtaining them from the large intestine during colonoscopy requires skill and protocols for extracting microbial contents from the biopsies.

Some studies (e.g. Gevers 2014[6]) have shown that better classifiers for inflammatory bowel disease can be developed using samples of luminal content collected directly from the gut rather than stool, but this is contradicted by other studies that show high classifier accuracy for stool (see below for explanation of these terms). In an ideal world, sampling design would be driven by a hypothesis about mechanism. Most microbial biomass and therefore metabolism occurs in the luminal contents of the large intestine, so microbes that produce and release small-molecule metabolites that enter the bloodstream would be expected to be most important there. In contrast, microbes that interact directly with epithelial cells or dendritic cells would be expected to be concentrated in mucosal biopsy. Microbes that produce metabolites from dietary components that are absorbed in the ileum, duodenum or jejunum should be sought there. However, we still lack the general understanding about the distribution of microbes and metabolism along the length of the gut to draw general conclusions about where to take samples. The advent of very low biomass protocols, such as KatharoSeq (which uses a series of positive

control spike-ins to define what is real and what is contamination at different stages), allows even tiny specimens to be processed[38].

An important question is how often to sample stool, because the microbiome ecology is intrinsically dynamic. This largely comes down to what question you are trying to answer. Remarkable changes have been observed between one day and the next, especially in the times surrounding colonoscopy and surgery[39,25] as well as during clinical situations such as IBD flares. These changes would be missed entirely with a less frequent study design. For episodic diseases, such as IBD, it is known that patients can have large changes in the microbiome composition on time scales of weeks to months[40]. On the other hand, changes induced by diet, e.g. those associated with weight loss, take place on a timescale closer to months than days in humans[41,42,43] (Fig. 1.2). Having several serial samples provides considerable insight into microbiome dynamics[40,44], with samples up to half a dozen providing substantially better classifiers from stool regardless of sampling interval. However, answering this question conclusively will require detailed study of many patients, prohibitively expensive at present and impossible to perform with anything beyond a stool sample. However, although it is difficult to obtain serial mucosal/biopsy or luminal samples from individuals due to cost and invasiveness of the procedures, this may be the best strategy for patients who are receiving multiple, often scheduled, endoscopies as part of their routine care or in event of exacerbations (e.g. variceal screening EGDs for patients with cirrhosis; colonoscopy for patients with IBD).

Figure 1.2. Interindividual variability is a stronger discriminatory factor than diet, even under extreme dietary changes. (A) Principal coordinates (PCoA) plot of unweighted UniFrac distances of the subjects (color) and their diets (shape). (B) PCoA plot with traces to show the individual variation over time, each edge s connected in order according to the collection timepoint. [43]

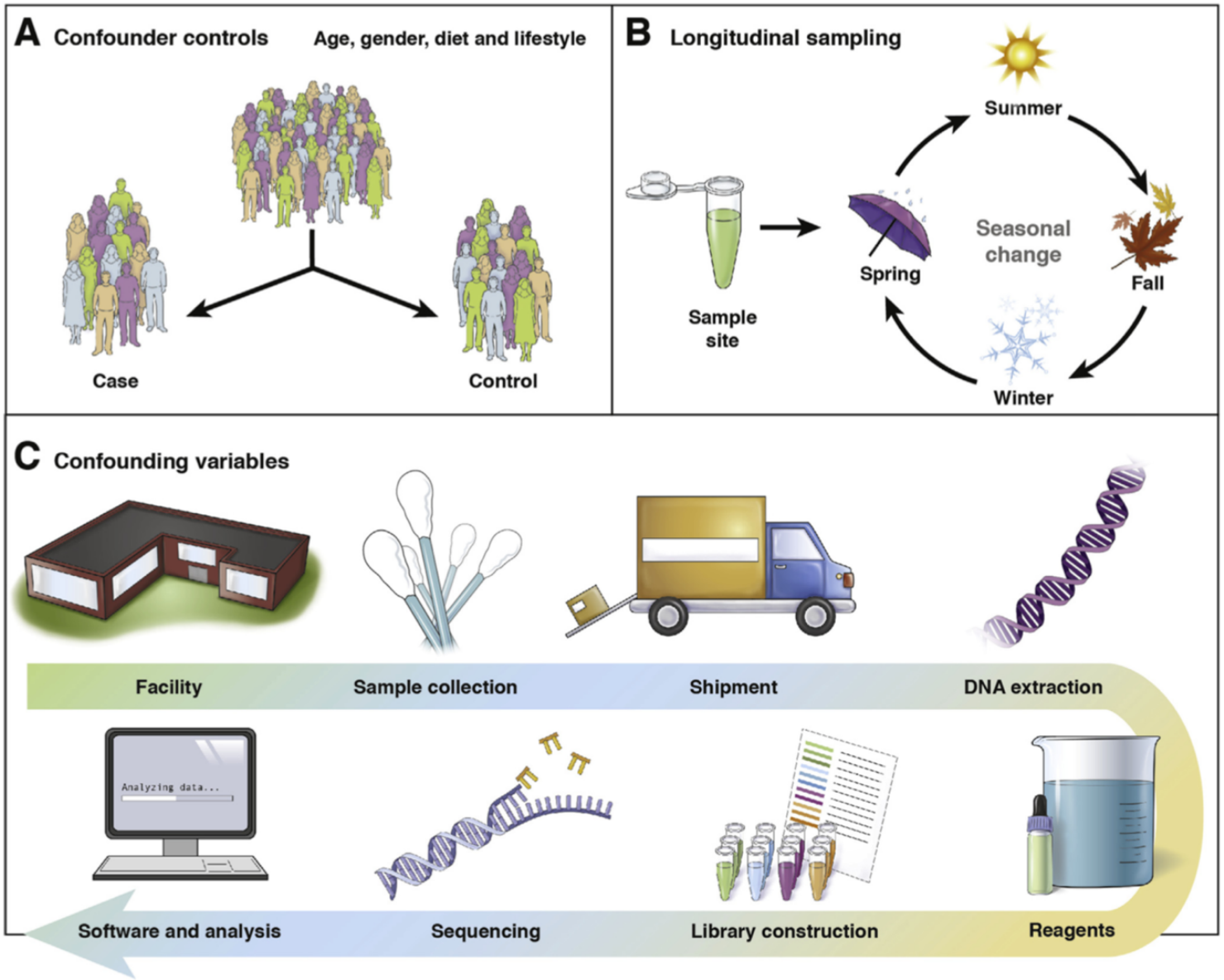

A
PC2 (9.731 %)
PC1 (15.21 %)
B
PC2 (9.731 %)
PC1 (15.21 %)
Diet
Animal
Plant
Subject
1
2
3
4
5
6
7
8
9
10
11

On the other hand, adequately collected and optimally stored fecal samples from chronic liver disease patients, such as nonalcoholic fatty liver disease, can provide unique insights into differentiation between those with milder form of fibrosis versus advanced fibrosis in cross-sectional setting[45]. Furthermore, integrating the gut microbiome with the metabolome may offer deeper insights into the metabolic perturbations linking the gut microbiome with disease states[46]. Recent studies have also suggested that certain bacterially derived metabolites may be associated with shared gene effects with disease states of interest. Longitudinal studies are needed to assess causality, and will be discussed later in this review.

1.4 What sort of microbiome data should I collect?

There is a bewildering diversity of microbiome-relevant molecular analyses that can be performed on biological specimens today, each with strengths and weaknesses (Fig. 1.3). The correct type of analyses for an experiment is completely dependent on the scientific question and hypothesis. Some of the more traditional methods focus on species identification or toxin presence for pathogens[47], while newer methods seek to describe and detect whole communities rather than individual organisms.

Figure 1.3. Clinical microbiome studies need to be planned with a critical eye towards many details.A) Groups of interest will need to be evaluated for confounders - age, gender, diet, lifestyle factors, medications, etc. Accounting for such variables can help resolve differences in microbiota between cohorts that might otherwise be masked.[48] (B) Confounding factors can be further reduced by performing longitudinal studies that allow for the assessment of community stability.[40] (C) Standardization of all stages of sample processing is essential to control for variation introduced by every step of the process: kit reagents, primers, sample storage, and other factors. Recording and maintaining detailed metadata notes about all aspects of each sample, from clinical variables to sample processing, are crucial for data interpretation; without metadata, it is difficult to draw meaningful conclusions from sequencing data.


A Confounder controls
Age, gender, diet and lifestyle
Case
Control
B Longitudinal sampling
Summer
Spring
Seasonal change
Fall
Winter
Sample site
C Confounding variables
Facility
Sample collection
Shipment
DNA extraction
Analyzing data...
Software and analysis
Sequencing
Library construction
Reagents

For known organisms with well-characterized selective culture conditions, culturing is still the most sensitive detection method, and comparisons of colony-forming units per milliliter is the best way to obtain the absolute abundance of viable organisms. This method can be used on a variety of sample types including stool, blood, and skin. Various organisms found in the stool are susceptible for antibiotic resistance including *Clostridium difficile* and *Enterococcus* spp. which are also highly infectious pathogens[49]. Culturing enables a phenotypic classification of an isolate including pathogenicity, antibiotic resistance mechanisms, and antibiotic susceptibility[47]. However, this method is best suited to reading out a small number of well-known organisms that can survive in the presence of oxygen, not to characterizing the entire, complex and largely anaerobic gut microbiome.

A broader view can be obtained by assay panels that target a set of known bacteria, viruses, parasites, or functional genes such as toxins or antibiotic resistance. Stool samples are generally processed through nucleic acid extraction followed by cDNA synthesis and subsequent amplification using mixtures of primers specific for a given range of organisms. Either gDNA or PCR product is then qualified and quantified across the organism panel either through a hybridization array using a fluorescence based measure or a melt curve analysis[50,51]. Both qPCR and RT-qPCR are also examples of these methods which are used to detect and quantify specific organisms[52]. Various companies (Verigene, Biofire, Luminex) have developed FDA-approved platforms to detect microbial pathogens from bulk stool samples[53]. The platforms can be microfluidic chips that perform multiple processes including DNA extraction, PCR, and readout[54]. Throughput ranges from 1 to 24 samples, while time is 1-5 hours[53]. The mentioned technologies targets between 14 - 22 analytes including 7 - 14 bacteria, 2 toxins, 2 - 5 viruses, and 0 - 4 parasites. The advantage to these assays is that they provide absolute abundance of each taxon per gram or mL of input material, and have high dynamic range. The disadvantage is that there are many undiscovered taxa in the gut that may be important, and these will be missed in a

targeted panel. However, as we understand more about the specific microbes that make the difference between clinical indications, these targeted panels will be increasingly valuable. However, one important concern is whether panels developed in one population will apply to another (see below).

Amplicon analyses, in which a specific piece of DNA is amplified by orders of magnitude using various methods including the polymerase chain reaction (PCR), have been the workhorse of the microbiome for the past 15 years[55]. In these analyses, PCR primers that match a specific gene, usually the 16S rRNA for bacteria and archaea and the ITS for fungi, are used to amplify all the variants that occur between the highly conserved regions used to construct the primers. For example, bacterial 16S ribosomal RNA (rRNA) genes contain 9 “hypervariable regions” (V1 – V9) that exhibit sequence diversity and therefore are often used as a “barcode” to differentiate many bacterial taxa, sometimes but not always at the species level. Then next-generation sequencing, typically on the Illumina platform[56], is used to read all the sequences, which can then be placed into a phylogenetic tree or matched to a database. There are many considerations in choosing which primers to use, and the difference between the microbiome profiles obtained with different PCR primers is much greater than the difference between the stool of different healthy individuals[19]. Consequently, the best option is to use the same PCR primers as other studies that you would like to compare your results to, or if there is no specific study in mind then using widely used primers such as the V1-3 or V3-5 primers from the Human Microbiome Project or the V4 primers from the Earth Microbiome Project (which have the advantage that they pick up archaea such as Methanobrevibacter and Methanosphaera, both important in the gut) is the best plan. Critically, many primers can target the same variable region, so it is important to know not just which region is being sequenced but the specific primers themselves. In general, the specific region is much more important than the length of the fragment[57,58], and a long sequence with biased primers can give a spectacularly incorrect result. Therefore, it is important to beware of

claims about the value of long-read sequencing that are not backed by extensive validation in the form of peer-reviewed papers. Many species of bacteria are identical along the full length of the 16S rRNA gene, and it is therefore in principle impossible to distinguish all bacterial species using that gene - despite claims of some vendors. In general, genus-level resolution is possible for most bacterial taxa, but species resolution is difficult[59]. Amplicon analyses in general are challenging to apply to viruses, which is mostly due to the fact that there is no gene common to all viruses like there is in bacteria.

Although 16S rRNA sequencing has enabled a great deal of scientific research on microbiomes, simply knowing the genera of bacteria and its relative abundance is not as useful for clinical analysis. This is because each genus can have a wide range of strains which are genomically distinct. This is true even within a species: *E. coli*, for instance, has a genome that can vary from 4-6 million DNA bases[60], which group into several thousand distinct genes, some of which can be quite virulent. As a result, there are thousands of known strains of *E. coli* that have been sequenced (only about 1/3 of the *E. coli* genome is core to all its strains) and found to be genomically-distinct, with at least one strain considered a probiotic and another that can cause debilitating illness. .

In contrast to the use of one gene, such as 16S rRNA, shotgun metagenomics is a method that fragments all the DNA from a sample into small pieces, sequences these fragments, then tries to puzzle these fragments together into a view of the microbiome[61]. The advantage to shotgun metagenomics is that it is very easy to explain what it does: you are trying to infer the complete list of microbial strains present in a microbiome, including the fungi and viruses that are missed by 16S rRNA amplicon analysis, and how abundant each of those strains is. However, the technical challenges are considerable: for example, analyses rely on genomes of the organisms in the gut, many of which are unknown (especially outside the bacteria). Shotgun metagenomics was traditionally orders of magnitude more expensive than amplicon analyses, but

with rapid declines in the cost of DNA sequencing and library preparation this technique is becoming much more accessible on a large scale. Additionally, the amount of DNA required for shotgun metagenomics has recently dropped from micrograms to less than a nanogram, allowing it to be used on biopsy specimens. An important limitation to shotgun metagenomics is that *all* the DNA will be sequenced, including human DNA, which is a problem if your subjects are not consented for human DNA analysis or if your biopsy specimen is dominated by host tissue (resulting in very expensively re-sequencing the human genome, with only a small trace of microbial reads; this is common in biopsy specimens, which is which 16S rRNA amplicon analysis is typically used for such specimens; "host DNA depletion" techniques, although successful in saliva[62] have not yet worked for biopsies although this is an active area of methods development). Shotgun metagenomics is rapidly displacing 16S rRNA amplicon analysis because of its expanded taxonomic range and strain-level resolution, but is subject to many of the same reproducibility issues that have not yet been as well characterized because of the increased expense of the assays.

Metatranscriptomics, in which the transcribed RNA is sequenced, and metaproteomics, which using mass spectrometry to sort out the wide range of proteins in a sample, have tremendous promise because they read out gene expression, but are still very challenging. Most bacterial transcripts only last a few minutes[63], so the interpretation of RNA left in a stool sample is challenging. Moreover, in the few comparisons that have been done, the correlation between gene expression in the RNA and proteins at the whole-community level has been close to zero, complicating interpretation of the expression profiles. These should be considered emerging technologies rather than ready for routine use, although techniques are rapidly improving. Studies of expression require metagenomic data from the same sample to back them so that changes in the relative expression of particular genes can be distinguished from changes in the representation of these genes in the community[64].

Metabolomics, the study of the non-protein smaller molecules including products of metabolism, is a very exciting emerging area because it gets directly at the function of the community. The most common approaches separate metabolites by GC (gas chromatography) or LC (liquid chromatography) before analysis by mass spectrometry as charged ions. There are two main approaches of metabolomics analysis: targeted metabolomics, where we have a predetermined list of molecules, typically where the reference standards are available. It is usually the most sensitive approach for detecting molecules of interest and has better quantification compared to untargeted mass spectrometry but does not allow for discovery[65,66]. Most molecules that are made by the microbiome are not commercially available or still remain to be discovered and therefore cannot be analyzed via targeted methods. On the other hand, untargeted metabolomics aims to detect as many ions as possible. The main challenge for untargeted metabolomics is the annotation. For untargeted metabolomics tandem mass spectrometry (which weighs the ions, then breaks them into fragments, then weighs the fragments) is often used to provide annotations by matching against a reference library of known molecules. However this fails to annotate molecules that are modified by the microbiome or host metabolism. However fragmentation data from related spectra can be found by linking their mass spectra through a technique called molecular networking[67,68] (see below), allowing identification of new molecules that are related to known ones. An important consideration when choosing a metabolomics platform is whether the target molecules will be captured – for example, many standard untargeted LC/MS/MS approaches do not pick up short-chain fatty acids (SCFAs) such as butyrate and acetate which are known to play important physiological roles in the gut, on the other hand GC-MS does not pick up molecules from the host that are modified by microbes. Examples of such molecules include lithocholic acid, the oral bacteria produced fungal biofilm inhibitor mutanobactin A[69], and the microbial molecule 4-phenyl-ethyl sulfate that results in autism-like symptoms in rodent models[70]. The current preferred methods for stool are a combination of shotgun metagenomics and metabolomics. It is likely that metabolomics will not only be able to

report on microbially modified or microbially biosynthesized molecules, but also provide a direct read out of the medications as well as diet that affect the gut microbiome.

1.5 How should I analyze my data?

The main question clinicians usually have is either “how do my cases and controls differ?” or “is this sample from this patient indicative of a particular disease?” These questions can be difficult to answer with the current state of the science, especially given the many options for conducting the molecular analysis.

The wrong approach is to decide to do a microbiome study, pick a type of sample to collect, decide what molecular assay to run, and then decide to “analyze the data” yourself or hand it off to a bioinformatics or biostatistics collaborator, core facility or company. The greatest expense in many studies is data analysis, and if the study was not designed in a way that allows the data to be analyzed easily, this can take years and cost hundreds of thousands of dollars (if accurately accounted). We cover issues of study design extensively elsewhere in other recent reviews[71,72,73,74]. Briefly, it is important to consider confounding factors like age, drugs, diet, and co-housing, issues of causality. Your patients might be sick because their microbiomes are different or their microbiomes may be different as a consequence of their medical condition or treatment. Appreciating that study designs with equal numbers of samples per group, consistent timepoints, etc. are dramatically easier to analyze is also important. Furthermore, a common tactic is to use the microbiome “differences” to infer that they underlie a pathophysiological process that was not even part of the initial intent of the study. Not only does this assume a causative relationship between the microbiome and the pathology being investigated, but also our knowledge of the relationship of the gut microbiome on host processes is often not yet sufficient to support such conclusions. Lastly, for all NGS based methods of microbiome analysis, it is

paramount to include positive and negative controls to help distinguish between signal and noise[38].

The most important consideration with data analysis is that different methods will give different results, even using the same raw data from the DNA sequencing instrument. This issue stems from several distinct sources. First, algorithms for assigning DNA sequences to particular genomes or classes of organisms are approximate. For example the popular RDP classifier[59] has an accuracy of about 80% at the genus level using short 16S rRNA fragments. This means that about 20% of the assignments are wrong, which is not ideal. In shotgun metagenomics, approaches, such as Kraken[75] or Centrifuge[76], based on k-mers (short fragments of sequences, often only a few bases long) are much more sensitive (likely to find an organism if it is present, especially at low abundance) but less specific (likely to report an organism even if it is not present) than those based on profile matches to marker genes, such as PhyloPhlAn[77]. Whether it's more dangerous to miss an organism that is present or accidentally report an organism that's absent depends on the clinical application. In any case, none of these techniques is currently suitable for clinical use. Diagnosis of pathogens should still be performed by FDA-approved culture-based, PCR-based, or antibody-based assays.

In addition, most approaches rely on reference databases that are highly incomplete. Consequently, matches to a given sequence will vary depending on what sequences are actually in the reference database and the name given to the closest sequence, which results in different bacterial names given to the same DNA sequence depending on the database used. Because of this, you can get wildly different results. In the past this was an enormous problem, although cooperation among the rRNA-based taxonomy databases such as SILVA[78], RDP[79] and Greengenes[80] have reduced this problem and resulted in more consistency between results in recent years. However, taxonomy based on whole genomes rather than on single marker genes

is likely to prompt large-scale revision of taxonomy as we discover more about the relationships among major groups of organisms.

The major considerations in data analysis are as follows (Fig. 1.4):

1. How do I go from my raw DNA sequence data to a table of how many of each species (or gene/strain, for metagenomics) is observed in each sample?

2. How do I link this table up to relevant clinical variables for analysis?

3. How do I perform appropriate analyses either at the level of the whole microbiome (typically, “alpha diversity “ and “beta diversity” analyses) or at the level of individual taxa or genes?

Figure 1.4. In order to turn samples into data, you will first have to extract DNA and decide what type of sequencing you want to perform.Two common types of protocols are amplicon sequencing and shotgun sequencing. In amplicon sequencing, polymerase chain reaction primers are used to target a specific region of a specific gene, focusing sequencing effort on just those fragments. One of the most widely used protocols targets the V4 region of the 16S rRNA gene24. In shotgun sequencing, the DNA in the sample is randomly sheared and sequenced, generating data from many different parts of the genome. The specifics of the molecular protocol used prior to shotgun sequencing are important for what type of data are being examined, and this type of sequencing can be used for example for metagenomics and metatranscriptomics. The initial processing performed on the data after sequencing depends on the type of sequencing performed. For amplicon studies, one common strategy is to upload the data into Qiita81 and to use Deblur82 to resolve sequence data into single-sequence variants called sub-operational taxonomic units (sOTUs). Taxonomic assignments are generally performed using naive Bayes classifiers such as the RDP Classifier59 as implemented in q2-feature-classifier against reference databases such as Greengenes83, SILVA78, RDP79 or UNITE84 (fungal ITS) depending on the amplicon target. Shotgun sequencing of host-associated samples first requires preprocessing to remove host DNA prior to analysis. Typically, the shotgun data are then summarized using tools such as Kraken75, MEGAN85, or HUMAnN286 to generate taxonomic or functional profiles, or are assembled with tools such as metaSPAdes87 and MEGAHIT88. For both sequencing methods, higher-level analyses (for example, alpha and beta diversity, taxonomic profiling and machine learning) are subsequently used to assay patterns of microbiome variation in the context of the study design. Metagenomic assemblies can also be analyzed through platforms such as Anvi'o89. SourceTracker90 is a Bayesian estimator of the sources that make up each unknown community, is useful for classifying microbial samples according to environment of origin91. ITS, internal transcribed spacer. Citizen Science platforms, such as the American Gut Project25, standardize the molecular work and bioinformatic processing in order to generate a basic summary report of the content of an individuals sample. In the case of the American Gut, the samples are also placed into the context of a few other popular microbiome studies though data integration.

Broad view					Narrow view
Metabolomics (Non-protein small molecules) All small molecules made by all organisms present Targeted: better for known metabolites (i.e. bile acids) Non-targeted: better for novel compounds, discovery Good for looking at functional changes No link to specific organisms	**Metaproteomics** (Protein) **Metatranscriptomics** (RNA) All protein or RNA made by all organisms present Good for looking at functional changes No link to specific organisms	**Shotgun sequencing** (Complete genomes, "Metagenomics") Every organism present will have most of the genomes sequenced: all bacteria fungi, viruses, etc. This includes the host/patient, discuss if using biopsy samples All organisms present No functional changes	**Amplicon sequencing** (Partial genomes) Most selected organisms present, depending on method used (no viruses) Most selected organisms present - uses 16S, 18S or ITS as "barcode"	**PCR panels** (qPCR, RT-PCR) Can include a single type or a select combination of organisms Generally up to around 24 per sample Limited in scope to known specific organisms in selected panel	**Culture** (traditional method for bacteria; also some archaea and viruses) A small number of known organisms that will grow on specific media under aerobic conditions Anaerobes can be isolated and grown, but many difficulties are present Limited in scope to known organisms under specific conditions
All organisms (including host)	RNA viruses and all organisms (including host)	All organisms (including host)	Bacteria and some archaea for 16S Eukaryotes only for 18S Fungi only for ITS	Viruses and other selected organisms (depends on panel used)	Bacteria, fungi, archaea and viruses (depends on media used)
High throughput 96 samples per run ~ 48 hours $$$	High throughput 96-384 samples per run ~ 48 hours $$$$	High throughput 384 samples per run 48 hours $$$	High throughput 384 samples per run 48 hours $$	Low throughput Max 30 pooled samples per run 1 to 5 hours $$	Low throughput 1 sample per media used 24 to 48 hours $

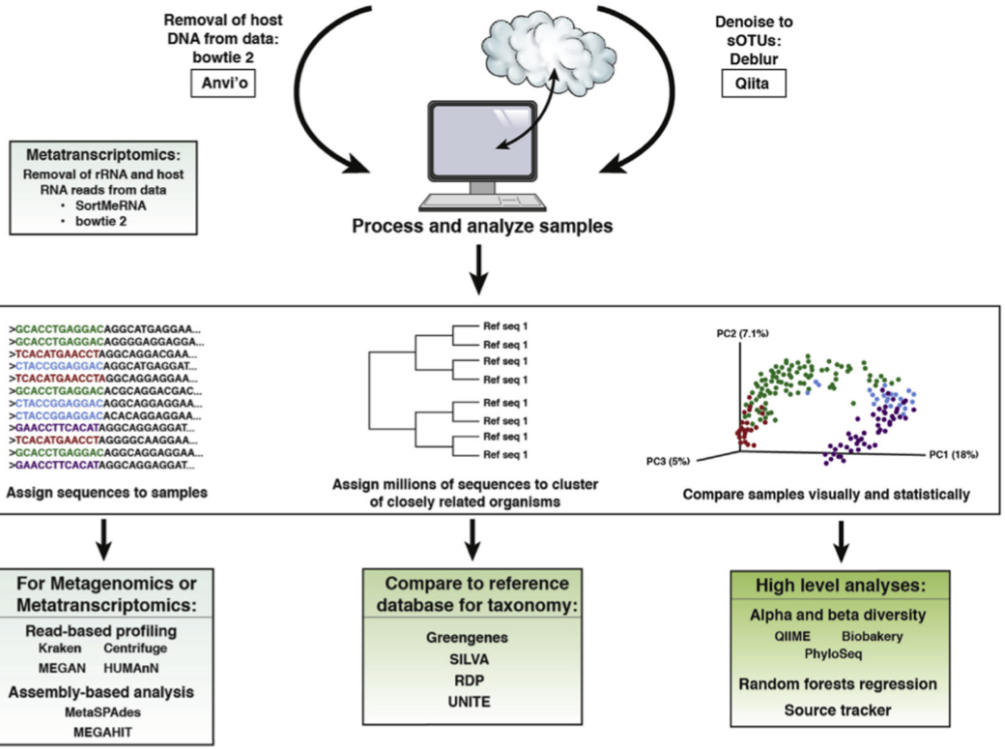


Figure. 1.4. In order to turn samples into data, you will first have to extract DNA and decide what type of sequencing you want to perform. *Continued*

Citizen science platforms:

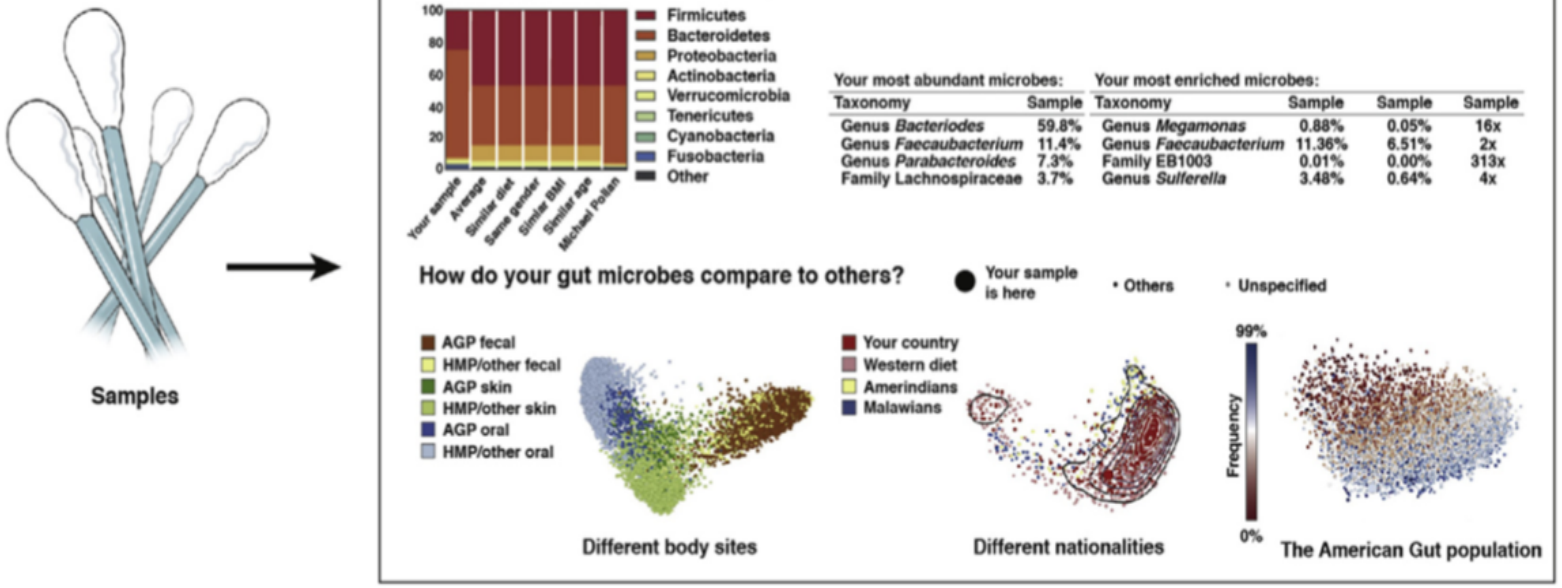


Figure 4. Continued

There are several features of microbiome data from a statistical standpoint such as sparsity, compositionality, and zero-inflation that make standard statistical tools inappropriate for most microbiome analyses. It is therefore critical to use tools designed for these analyses, such as QIIME/Qiita[81], the BioBakery[92], or PhyloSeq[93] that take these considerations into account. Providing details of how to analyze microbiome data is beyond the scope of this piece, but we have covered this topic recently in several other reviews that will be of interest to readers who want more details[74].

1.6 What are the limits to combining data from different studies?

One frequently encountered issue is reading an exciting research paper that links a particular microbe, pathway, or gene to a condition or treatment, then wanting to see if the same relationship holds true in a new cohort or a new individual patient. This apparently simple question turns out to be surprisingly difficult.

As noted above, a very large number of factors can affect the readout of the microbiome, especially at finer taxonomic levels, but they are by no means limited to these levels. Especially with PCR-based methods, including primers that target different hypervariable regions (e.g. V1-3 vs V4) or different primers that target the same region but pick up different taxa with different efficiency, the same samples can yield completely different assessments of which phyla are abundant in a given specimen, let alone assessments at the species level (to which current sequencing techniques are poorly suited). Consequently, if you are designing a new study and want to compare to an existing study, the safest approach to use exactly the same methods in every detail, including sample collection, sample storage, DNA extraction, PCR or library construction, sequencing, and bioinformatics analysis. Standardized reporting such as the Genomic Standards Consortium MIxS standards[94] help immensely with this task by capturing the

information in a structured way and, in the context of databases such as Qiita[95], allow automatic retrieval of studies that used comparable methods.

The Human Microbiome Project[19] demonstrated that even when everything else is kept exactly the same, the choice of PCR primers (V1-V3 vs V3-V5) and the choice of whether to perform shotgun metagenomics or 16S rRNA sequencing on the same samples can produce completely different results. Similarly, the Microbiome Quality Control (MBQC) project demonstrated that differences in computational pipeline, even on the same data, could lead to large differences in the inferred outcomes at levels from the species to the phylum[96]. However, one valuable outcome of MBQC was that many different labs could independently reproduce similar results on the same samples by following a consistent, written protocol[96]. Similarly, in the American Gut Project, we find that dozens of sequencing runs over many years yield consistent results when consistent protocols are used, and this is highlighted in the supplementary video file of that paper [https://figshare.com/articles/movie_s2_mp4/][97,25].

In general, whether and how studies can be combined depends on the subtlety of the effect and the type of analysis being performed. Different parts of the human body differ radically in their microbiomes, and neonatal microbiomes are completely different from adults. Therefore, even studies using different DNA extraction methods and sequencing techniques will often yield the same pattern in combined analysis, e.g. through principal coordinates analysis[98]. In contrast, subtle differences such as those yielded by day-to-day variation within a healthy individual are much smaller, and will be obscured by even minor technical variation such as lot numbers of sequencing reagents. A general guideline is that the more technical factors differ between two studies, the more obvious the difference will need to be in order to be visible. Although the American Gut Project and other recent projects have started to construct an effect size scale for factors that affect the microbiome in large or small ways, incorporating technical variation at these scales would be an arduous and expensive undertaking. One approach that is often useful is

asking whether particular taxa or gene functions are reliably increased or decreased with a given clinical state (e.g. ulcerative colitis, NAFLD) across many studies, although different methods can in principle lead to different conclusions even with data analysis at this level.

Of particular concern to clinicians is whether data from companies offering testing, or from citizen-science projects such as the American Gut Project[25], is comparable to studies performed in the scientific literature. The American Gut Project is part of the Earth Microbiome Project (EMP), and uses the EMP protocols[24] that have been applied in literally thousands of microbiome studies, including those that are clinically relevant. Unsurprisingly, testing services that use proprietary protocols produce different results even on the same biological specimens. In general, to understand these differences, it is necessary to have detailed information about all the protocols being used.

Another important issue is that although many associations between the microbiome and disease or between the microbiome and treatment have been found within the context of individual research studies, there are many reasons why these might not generalize to new individuals or populations. It is well-known in the field of human genetics that environmental factors have a major impact on which genes are important for a given trait, and the same is likely true for the microbiome, so validation cohorts are essential in order to prove generality of microbiome findings just as they are for human genetic findings. Some conditions, such as IBD, have very robust signatures across populations[99,6,100], with diagnostic models trained in humans working even on dogs[101]; in contrast, for obesity, although there are typically signatures that separate lean from obese within one population, these signatures do not apply across other cohorts[99,102,103,23]. This result is surprising given that obesity can be transmitted from humans into germ-free mice by transmitting the microbiome from obese people, demonstrating the direct effects of the microbiome[2,104]. Understanding which findings will generalize to new subjects, and which will not, remains an important outstanding challenge in the field. It is possible that new

ecosystem-level or pathway-level concepts and methods need to be developed in order to develop such an understanding.

1.7 Conclusions

Although there is great interest in the microbiome, there is still a long way to go before microbiome-based diagnostics become a routine part of clinical care. Microbiome studies have been enormously valuable both in understanding mechanisms of disease in animal models and finding associations with disease in humans. A good analogy is machine translation of natural languages: there has been interest since the 1950s, and poorly-functioning systems have been available since the 1980s, but only in the last couple of years has it been possible to have a conversation with someone who speaks no common language using an app on a smartphone, or to translate signs or menus from Chinese into English in real time using that smartphone's camera. In the same way, microbiome testing right now is primarily of interest as a science project. However, there will be rapid progress in the near term to develop better technical capability, including better user interfaces with readouts at the level of bacterial strains, and integration of ecological dynamics concepts to better understand the transitions from health to illness.

1.8 Acknowledgments

Chapter 1, in full, is a reprint of the material as it appears in “Microbiome 101: Studying, Analyzing, and Interpreting Gut Microbiome Data for Clinicians.” Celeste Allaband, Daniel McDonald, Yoshiki Vázquez-Baeza, Jeremiah J Minich, Anupriya Tripathi, David A Brenner, Rohit Loomba, Larry Smarr, William J Sandborn, Bernd Schnabl, Pieter Dorrestein, Amir Zarrinpar, and Rob Knight. Clinical Gastroenterology and Hepatology, 17 (2) , 218-230. The dissertation author was the primary investigator and first author of this paper.

1.9 References

1. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JIA core gut microbiome in obese and lean twins. Nature. 2009;457(7228):480-4. Epub 2008/12/02.

2. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V, Henrissat B, Bain JR, Muehlbauer MJ, Ilkayeva O, Semenkovich CF, Funai K, Hayashi DK, Lyle BJ, Martini MC, Ursell LK, Clemente JC, Van Treuren W, Walters WA, Knight R, Newgard CB, Heath AC, Gordon JI. Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science. 2013;341(6150):1241214. Epub 2013/09/07.

3. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto JM, Kennedy S, Leonard P, Li J, Burgdorf K, Grarup N, Jørgensen T, Brandslund I, Nielsen HB, Juncker AS, Bertalan M, Levenez F, Pons N, Rasmussen S, Sunagawa S, Tap J, Tims S, Zoetendal EG, Brunak S, Clément K, Doré J, Kleerebezem M, Kristiansen K, Renault P, Sicheritz-Ponten T, de Vos WM, Zucker JD, Raes J, Hansen T; MetaHIT consortium, Bork P, Wang J, Ehrlich SD, Pedersen O. Richness of human gut microbiome correlates with metabolic markers. Nature. 2013;500(7464):541-6. Epub 2013/08/30.

4. Cotillard A, Kennedy SP, Kong LC, Prifti E, Pons N, Le Chatelier E, Almeida M, Quinquis B, Levenez F, Galleron N, Gougis S, Rizkalla S, Batto JM, Renault P; ANR MicroObes consortium, Doré J, Zucker JD, Clément K, Ehrlich SD. Dietary intervention impact on gut microbial gene richness. Nature. 2013;500(7464):585-8. Epub 2013/08/30.

5. Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci U S A. 2007;104(34):13780-5. Epub 2007/08/19.

6. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, Schwager E, Knights D, Song SJ, Yassour M, Morgan XC, Kostic AD, Luo C, González A, McDonald D, Haberman Y, Walters T, Baker S, Rosh J, Stephens M, Heyman M,

Markowitz J, Baldassano R, Griffiths A, Sylvester F, Mack D, Kim S, Crandall W, Hyams J, Huttenhower C, Knight R, Xavier RJ.The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe. 2014;15(3):382-92. Epub 2014/03/19.

7. Schirmer M, Smeekens SP, Vlamakis H, Jaeger M, Oosting M, Franzosa EA, Horst RT, Jansen T, Jacobs L, Bonder MJ, Kurilshikov A, Fu J, Joosten LAB, Zhernakova A, Huttenhower C, Wijmenga C, Netea MG, Xavier RJ. Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity. Cell. 2016;167(4):1125-36 e8. Epub 2016/11/05.

8. Henao-Mejia J, Elinav E, Jin C, Hao L, Mehal WZ, Strowig T, Thaiss CA, Kau AL, Eisenbarth SC, Jurczak MJ, Camporez JP, Shulman GI, Gordon JI, Hoffman HM, Flavell RA.Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. Nature. 2012;482(7384):179-85. Epub 2012/02/03.

9. Hoyles L, Fernández-Real JM, Federici M, Serino M, Abbott J, Charpentier J, Heymes C, Luque JL, Anthony E, Barton RH, Chilloux J, Myridakis A, Martinez-Gili L, Moreno-Navarrete JM, Benhamed F, Azalbert V, Blasco-Baque V, Puig J, Xifra G, Ricart W, Tomlinson C, Woodbridge M, Cardellini M, Davato F, Cardolini I, Porzio O, Gentileschi P, Lopez F, Foufelle F, Butcher SA, Holmes E, Nicholson JK, Postic C, Burcelin R, Dumas ME. Molecular phenomics and metagenomics of hepatic steatosis in non-diabetic obese women. Nat Med. 2018;24(7):1070-80. Epub 2018/06/27.

10. Tripathi A, Debelius J, Brenner DA, Karin M, Loomba R, Schnabl B, Knight R.. The gut-liver axis and the intersection with the microbiome. Nat Rev Gastroenterol Hepatol. 2018;15(7):397-411. Epub 2018/05/12.

11. Xie G, Wang X, Liu P, Wei R, Chen W, Rajani C, Hernandez BY, Alegado R, Dong B, Li D, Jia W. Distinctly altered gut microbiota in the progression of liver disease. Oncotarget. 2016;7(15):19355-66. Epub 2016/04/02.

12. Grąt M, Wronka KM, Krasnodębski M, Masior Ł, Lewandowski Z, Kosińska I, Grąt K, Stypułkowski J, Rejowski S, Wasilewicz M, Gałęcka M, Szachta P, Krawczyk M. Profile of Gut Microbiota Associated With the Presence of Hepatocellular Cancer in Patients With Liver Cirrhosis. Transplant Proc. 2016;48(5):1687-91. Epub 2016/08/09.

13. Xie G, Wang X, Zhao A, Yan J, Chen W, Jiang R, Ji J, Huang F, Zhang Y, Lei S, Ge K, Zheng X, Rajani C, Alegado RA, Liu J, Liu P, Nicholson J, Jia W.Sex-dependent effects on gut microbiota regulate hepatic carcinogenic outcomes. Sci Rep. 2017;7:45232. Epub 2017/03/28.

14. Shalapour S, Lin XJ, Bastian IN, Brain J, Burt AD, Aksenov AA, Vrbanac AF, Li W, Perkins A, Matsutani T, Zhong Z, Dhar D, Navas-Molina JA, Xu J, Loomba R, Downes M, Yu RT, Evans RM, Dorrestein PC, Knight R, Benner C, Anstee QM, Karin M.Inflammation-induced IgA+ cells dismantle anti-liver cancer immunity. Nature. 2017;551(7680):340-5. Epub 2017/11/17.

15. Knight R, Callewaert C, Marotz C, Hyde ER, Debelius JW, McDonald D, Sogin ML. The Microbiome and Human Biology. Annu Rev Genomics Hum Genet. 2017;18:65-86. Epub 2017/04/05.

16. Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol. 2016;14(8):e1002533. Epub 2016/08/20.

17. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI.The human microbiome project. Nature. 2007;449(7164):804-10. Epub 2007/10/19.

18. Savage DC. Microbial ecology of the gastrointestinal tract. Annu Rev Microbiol. 1977;31:107-33. Epub 1977/01/01.

19. Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486(7402):207-14. Epub 2012/06/16.

20. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. Diversity of the human intestinal microbial flora. Science. 2005;308(5728):1635-8. Epub 2005/04/16.

21. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Doré J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J; MetaHIT Consortium, Bork P, Ehrlich SD, Wang J.A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010;464(7285):59-65. Epub 2010/03/06.

22. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Doré J; MetaHIT Consortium, Antolín M, Artiguenave F, Blottiere HM, Almeida M, Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariaz G, Dervyn R, Foerstner KU, Friss C, van de Guchte M, Guedon E, Haimet F, Huber W, van Hylckama-Vlieg J, Jamet A, Juste C, Kaci G, Knol J, Lakhdari O, Layec S, Le Roux K, Maguin E, Mérieux A, Melo Minardi R, M'rini C, Muller J, Oozeer R, Parkhill J, Renault P, Rescigno M, Sanchez N, Sunagawa S, Torrejon A, Turner K, Vandemeulebrouck G, Varela E, Winogradsky Y, Zeller G, Weissenbach J, Ehrlich SD, Bork P. Enterotypes of the human gut microbiome. Nature. 2011;473(7346):174-80. Epub 2011/04/22.

23. He Y, Wu W, Zheng HM, Li P, McDonald D, Sheng HF, Chen MX, Chen ZH, Ji GY, Zheng ZD, Mujagond P, Chen XJ, Rong ZH, Chen P, Lyu LY, Wang X, Wu CB, Yu N, Xu YJ, Yin J, Raes J, Knight R, Ma WJ, Zhou HWRegional variation limits applications of healthy gut microbiome reference ranges and disease models. Nat Med. 2018. Epub 2018/08/29.

24. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vázquez-Baeza Y, González A, Morton JT, Mirarab S, Zech Xu Z, Jiang L, Haroon MF, Kanbar J, Zhu Q, Jin Song S, Kosciolek T, Bokulich NA, Lefler J, Brislawn CJ, Humphrey G, Owens SM, Hampton-Marcell J, Berg-Lyons D, McKenzie V, Fierer N, Fuhrman JA, Clauset A, Stevens RL, Shade A, Pollard KS, Goodwin KD, Jansson JK, Gilbert JA, Knight R; Earth

Microbiome Project Consortium. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017;551(7681):457-63. Epub 2017/11/02.

25. 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 DV, Jiang L, Kelley ST, Knights D, Kosciolek T, Ladau J, Leach J, Marotz C, Meleshko D, Melnik AV, 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; American Gut Consortium, Knight R. American Gut: an Open Platform for Citizen Science Microbiome Research. mSystems. 2018;3(3). Epub 2018/05/26.

26. 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 DV, Jiang L, Kelley ST, Knights D, Kosciolek T, Ladau J, Leach J, Marotz C, Meleshko D, Melnik AV, 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; American Gut Consortium, Knight R. Numerical ecology validates a biogeographical distribution and gender-based effect on mucosa-associated bacteria along the human colon. ISME J. 2011;5(5):801-9. Epub 2010/12/03.

27. Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, Bittar F, Fournous G, Gimenez G, Maraninchi M, Trape JF, Koonin EV, La Scola B, Raoult D. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect. 2012;18(12):1185-93. Epub 2012/10/05.

28. Zengler K, Toledo G, Rappe M, Elkins J, Mathur EJ, Short JM, Keller M. Cultivating the uncultured. Proc Natl Acad Sci U S A. 2002;99(24):15681-6. Epub 2002/11/20.

29. Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, Goulding D, Lawley TD. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. Nature. 2016;533(7604):543-6. Epub 2016/05/05.

30. Vandeputte D, Falony G, Vieira-Silva S, Tito RY, Joossens M, Raes J.Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. Gut. 2016;65(1):57-62. Epub 2015/06/13.

31. Sinha R, Chen J, Amir A, Vogtmann E, Shi J, Inman KS, Flores R, Sampson J, Knight R, Chia N Collecting Fecal Samples for Microbiome Analyses in Epidemiology Studies. Cancer Epidemiol Biomarkers Prev. 2016;25(2):407-16. Epub 2015/11/26.

32. Loftfield E, Vogtmann E, Sampson JN, Moore SC, Nelson H, Knight R, Chia N, Sinha R. Comparison of Collection Methods for Fecal Samples for Discovery Metabolomics in Epidemiologic Studies. Cancer Epidemiol Biomarkers Prev. 2016;25(11):1483-90. Epub 2016/11/03.

33. Vogtmann E, Chen J, Amir A, Shi J, Abnet CC, Nelson H, Knight R, Chia N, Sinha R.Comparison of Collection Methods for Fecal Samples in Microbiome Studies. Am J Epidemiol. 2017;185(2):115-23. Epub 2016/12/18.

34. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial community variation in human body habitats across space and time. Bacterial community variation in human body habitats across space and time. Science. 2009;326(5960):1694-7. Epub 2009/11/07.

35. Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, Knights D, Gajer P, Ravel J, Fierer N, Gordon JI, Knight R.Moving pictures of the human microbiome. Genome Biol. 2011;12(5):R50. Epub 2011/06/01.

36. Zhang Z, Geng J, Tang X, Fan H, Xu J, Wen X, Ma ZS, Shi P. Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. ISME J. 2014;8(4):881-93. Epub 2013/10/18.

37. Hillman ET, Lu H, Yao T, Nakatsu CH. Microbial Ecology along the Gastrointestinal Tract. Microbes Environ. 2017;32(4):300-13. Epub 2017/11/14.

38. Minich JJ, Zhu Q, Janssen S, Hendrickson R, Amir A, Vetter R, Hyde J, Doty MM, Stillwell K, Benardini J, Kim JH, Allen EE, Venkateswaran K, Knight R. KatharoSeq Enables High-Throughput Microbiome Analysis from Low-Biomass Samples. mSystems. 2018;3(3). Epub 2018/03/27.

39. Smarr L, Hyde ER, McDonald D, Sandborn WJ, Knight R. Tracking Human Gut Microbiome Changes Resulting from a Colonoscopy. Methods Inf Med. 2017;56(6):442-7. Epub 2017/01/01.

40. Halfvarson J, Brislawn CJ, Lamendella R, Vázquez-Baeza Y, Walters WA, Bramer LM, D'Amato M, Bonfiglio F, McDonald D, Gonzalez A, McClure EE, Dunklebarger MF, Knight R, Jansson JK. Dynamics of the human gut microbiome in inflammatory bowel disease. Nat Microbiol. 2017;2:17004. Epub 2017/02/14.

41. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. Nature. 2006;444(7122):1022-3. Epub 2006/12/22.

42. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L, Li H, Bushman FD, Lewis JD. Linking long-term dietary patterns with gut microbial enterotypes. Science. 2011;334(6052):105-8. Epub 2011/09/03.

43. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, Biddinger SB, Dutton RJ, Turnbaugh PJ. Diet rapidly and reproducibly alters the human gut microbiome. Nature. 2014;505(7484):559-63. Epub 2013/12/18.

44. Vázquez-Baeza Y, Gonzalez A, Xu ZZ, Washburne A, Herfarth HH, Sartor RB, Knight R. Guiding longitudinal sampling in IBD cohorts. Gut. 2018;67(9):1743-5. Epub 2017/10/23.

45. Loomba R, Seguritan V, Li W, Long T, Klitgord N, Bhatt A, Dulai PS, Caussy C, Bettencourt R, Highlander SK, Jones MB, Sirlin CB, Schnabl B, Brinkac L, Schork N, Chen CH, Brenner DA, Biggs W, Yooseph S, Venter JC, Nelson KE. Gut Microbiome-Based Metagenomic Signature for Non-invasive Detection of Advanced Fibrosis in Human Nonalcoholic Fatty Liver Disease. Cell Metab. 2017;25(5):1054-62 e5. Epub 2017/05/04.

46. Caussy C, Hsu C, Lo MT, Liu A, Bettencourt R, Ajmera VH, Bassirian S, Hooker J, Sy E, Richards L, Schork N, Schnabl B, Brenner DA, Sirlin CB, Chen CH, Loomba R. Link between gut-microbiome derived metabolite and shared gene-effects with hepatic steatosis and fibrosis in NAFLD. Hepatology. 2018. Epub 2018/03/25.

47. Váradi L, Luo JL, Hibbs DE, Perry JD, Anderson RJ, Orenga S, Groundwater PW. Methods for the detection and identification of pathogenic bacteria: past, present, and future. Chem Soc Rev. 2017;46(16):4818-32. Epub 2017/06/24.

48. Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Kristiansen K, Voigt AY, Vestergaard H, Hercog R, Costea PI, Kultima JR, Li J, Jørgensen T, Levenez F, Dore J; MetaHIT consortium, Nielsen HB, Brunak S, Raes J, Hansen T, Wang J, Ehrlich SD, Bork P, Pedersen O. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature. 2015;528(7581):262-6. Epub 2015/12/04.

49. Bassetti M, Merelli M, Temperoni C, Astilean A. New antibiotics for bad bugs: where are we? Ann Clin Microbiol Antimicrob. 2013;12:22. Epub 2013/08/30.

50. Kempf VA, Trebesius K, Autenrieth IB. Fluorescent In situ hybridization allows rapid identification of microorganisms in blood cultures. J Clin Microbiol. 2000;38(2):830-8. Epub 2000/02/03.

51. Harris DM, Hata DJ. Rapid identification of bacteria and Candida using PNA-FISH from blood and peritoneal fluid cultures: a retrospective clinical study. Ann Clin Microbiol Antimicrob. 2013;12:2. Epub 2013/01/09.

52. Zautner AE, Groß U, Emele MF, Hagen RM, Frickmann H. More Pathogenicity or Just More Pathogens?-On the Interpretation Problem of Multiple Pathogen Detections with Diagnostic Multiplex Assays. Front Microbiol. 2017;8:1210. Epub 2017/07/15.

53. Huang RS, Johnson CL, Pritchard L, Hepler R, Ton TT, Dunn JJ. Performance of the Verigene(R) enteric pathogens test, Biofire FilmArray gastrointestinal panel and Luminex xTAG(R) gastrointestinal pathogen panel for detection of common enteric pathogens. Diagn Microbiol Infect Dis. 2016;86(4):336-9. Epub 2016/11/05.

54. Liu RH, Yang J, Lenigk R, Bonanno J, Grodzinski P.Self-contained, fully integrated biochip for sample preparation, polymerase chain reaction amplification, and DNA microarray detection. Anal Chem. 2004;76(7):1824-31. Epub 2004/04/01.

55. Tringe SG, Hugenholtz P. A renaissance for the pioneering 16S rRNA gene. Curr Opin Microbiol. 2008;11(5):442-6. Epub 2008/09/27.

56. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. Ultra-high-

throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012;6(8):1621-4. Epub 2012/03/10.

57. Liu Z, DeSantis TZ, Andersen GL, Knight R.Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. Nucleic Acids Res. 2008;36(18):e120. Epub 2008/08/30.

58. Soergel DA, Dey N, Knight R, Brenner SE. Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. ISME J. 2012;6(7):1440-4. Epub 2012/01/13.

59. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73(16):5261-7. Epub 2007/06/26.

60. Ogura Y, Ooka T, Asadulghani, Terajima J, Nougayrède JP, Kurokawa K, Tashiro K, Tobe T, Nakayama K, Kuhara S, Oswald E, Watanabe H, Hayashi T. Extensive genomic diversity and selective conservation of virulence-determinants in enterohemorrhagic Escherichia coli strains of O157 and non-O157 serotypes. Genome Biol. 2007;8(7):R138. Epub 2007/08/23.

61. Riesenfeld CS, Schloss PD, Handelsman J. Metagenomics: genomic analysis of microbial communities. Annu Rev Genet. 2004;38:525-52. Epub 2004/12/01.

62. Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K. Improving saliva shotgun metagenomics by chemical host DNA depletion. Microbiome. 2018;6(1):42. Epub 2018/02/28.

63. Har-El R, Silberstein A, Kuhn J, Tal MSynthesis and degradation of lac mRNA in E. coli depleted of 30S ribosomal subunits. Mol Gen Genet. 1979;173(2):135-44. Epub 1979/06/07.

64. Segata N, Boernigen D, Tickle TL, Morgan XC, Garrett WS, Huttenhower C. Computational meta'omics for microbial community studies. Mol Syst Biol. 2013;9:666. Epub 2013/05/15.

65. Melnik AV, da Silva RR, Hyde ER, Aksenov AA, Vargas F, Bouslimani A, Protsyuk I, Jarmusch AK, Tripathi A, Alexandrov T, Knight R, Dorrestein PC. Coupling Targeted and Untargeted Mass Spectrometry for Metabolome-Microbiome-Wide Association Studies of Human Fecal Samples. Anal Chem. 2017;89(14):7549-59. Epub 2017/06/20.

66. Aksenov AA, da Silva R, Knight R, Lopez NP, Dorrestein PC. Global chemical analysis of biology by mass spectrometry. Nature Reviews Chemistry. 2017;1:0054.

67. Watrous J, Roach P, Alexandrov T, Heath BS, Yang JY, Kersten RD, van der Voort M, Pogliano K, Gross H, Raaijmakers JM, Moore BS, Laskin J, Bandeira N, Dorrestein PCMass spectral molecular networking of living microbial colonies. Proc Natl Acad Sci U S A. 2012;109(26):E1743-52. Epub 2012/05/16.

68. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT,

Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, P CAB, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard PM, Phapale P, Nothias LF, Alexandrov T, Litaudon M, Wolfender JL, Kyle JE, Metz TO, Peryea T, Nguyen DT, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BO, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol. 2016;34(8):828-37. Epub 2016/08/10.

69. Wang X, Du L, You J, King JB, Cichewicz RH. Fungal biofilm inhibitors from a human oral microbiome-derived bacterium. Org Biomol Chem. 2012;10(10):2044-50. Epub 2012/01/28.

70. Hsiao EY, McBride SW, Hsien S, Sharon G, Hyde ER, McCue T, Codelli JA, Chow J, Reisman SE, Petrosino JF, Patterson PH, Mazmanian SK. Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. Cell. 2013;155(7):1451-63. Epub 2013/12/10.

71. Goodrich JK, Di Rienzi SC, Poole AC, Koren O, Walters WA, Caporaso JG, Knight R, Ley RE.Conducting a microbiome study. Cell. 2014;158(2):250-62. Epub 2014/07/19.

72. Gilbert JA, Quinn RA, Debelius J, Xu ZZ, Morton J, Garg N, Jansson JK, Dorrestein PC, Knight RMicrobiome-wide association studies link dynamic microbial consortia to disease. Nature. 2016;535(7610):94-103. Epub 2016/07/08.

73. Debelius J, Song SJ, Vazquez-Baeza Y, Xu ZZ, Gonzalez A, Knight R. Tiny microbes, enormous impacts: what matters in gut microbiome studies? Genome Biol. 2016;17(1):217. Epub 2016/10/21.

74. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciolek T, McCall LI, McDonald D, Melnik AV, Morton JT, Navas J, Quinn RA, Sanders JG, Swafford AD, Thompson LR, Tripathi A, Xu ZZ, Zaneveld JR, Zhu Q, Caporaso JG, Dorrestein PC. Best practices for analysing microbiomes. Nat Rev Microbiol. 2018;16(7):410-22. Epub 2018/05/26.

75. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15(3):R46. Epub 2014/03/04.

76. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. 2016;26(12):1721-9. Epub 2016/11/18.

77. Segata N, Börnigen D, Morgan XC, Huttenhower C.PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat Commun. 2013;4:2304. Epub 2013/08/15.

78. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 2007;35(21):7188-96. Epub 2007/10/20.

79. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 2014;42(Database issue):D633-42. Epub 2013/11/30.

80. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. 2012;6(3):610-8. Epub 2011/12/03.

81. 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. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335-6. Epub 2010/04/13.

82. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems. 2017;2(2). Epub 2017/03/16.

83. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72(7):5069-72. Epub 2006/07/06.

84. Kõljalg U, Larsson KH, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, Erland S, Høiland K, Kjøller R, Larsson E, Pennanen T, Sen R, Taylor AF, Tedersoo L, Vrålstad T, Ursing BM. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. New Phytol. 2005;166(3):1063-8. Epub 2005/05/05.

85. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, Ruscheweyh HJ, Tappu R.MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLoS Comput Biol. 2016;12(6):e1004957. Epub 2016/06/22.

86. Abubucker S, Segata N, Goll J, Schubert AM, Izard J, Cantarel BL, Rodriguez-Mueller B, Zucker J, Thiagarajan M, Henrissat B, White O, Kelley ST, Methé B, Schloss PD, Gevers D, Mitreva M, Huttenhower C. Metabolic reconstruction for metagenomic data and its application to the human microbiome. PLoS Comput Biol. 2012;8(6):e1002358. Epub 2012/06/22.

87. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G,

Alekseyev MA, Pevzner PA. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77. Epub 2012/04/18.

88. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31(10):1674-6. Epub 2015/01/23.

89. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ. 2015;3:e1319. Epub 2015/10/27.

90. Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, Bushman FD, Knight R, Kelley ST.Bayesian community-wide culture-independent microbial source tracking. Nat Methods. 2011;8(9):761-3. Epub 2011/07/19.

91. Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM, Gibbons SM, Larsen P, Shogan BD, Weiss S, Metcalf JL, Ursell LK, Vázquez-Baeza Y, Van Treuren W, Hasan NA, Gibson MK, Colwell R, Dantas G, Knight R, Gilbert JA. Longitudinal analysis of microbial interaction between humans and the indoor environment. Science. 2014;345(6200):1048-52. Epub 2014/08/30.

92. McIver LJ, Abu-Ali G, Franzosa EA, Schwager R, Morgan XC, Waldron L, Segata N, Huttenhower C.bioBakery: a meta'omic analysis environment. Bioinformatics. 2018;34(7):1235-7. Epub 2017/12/02.

93. McMurdie PJ, Holmes S. Phyloseq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data. Pac Symp Biocomput. 2012:235-46. Epub 2011/12/17.

94. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, Gilbert JA, Karsch-Mizrachi I, Johnston A, Cochrane G, Vaughan R, Hunter C, Park J, Morrison N, Rocca-Serra P, Sterk P, Arumugam M, Bailey M, Baumgartner L, Birren BW, Blaser MJ, Bonazzi V, Booth T, Bork P, Bushman FD, Buttigieg PL, Chain PS, Charlson E, Costello EK, Huot-Creasy H, Dawyndt P, DeSantis T, Fierer N, Fuhrman JA, Gallery RE, Gevers D, Gibbs RA, San Gil I, Gonzalez A, Gordon JI, Guralnick R, Hankeln W, Highlander S, Hugenholtz P, Jansson J, Kau AL, Kelley ST, Kennedy J, Knights D, Koren O, Kuczynski J, Kyrpides N, Larsen R, Lauber CL, Legg T, Ley RE, Lozupone CA, Ludwig W, Lyons D, Maguire E, Methé BA, Meyer F, Muegge B, Nakielny S, Nelson KE, Nemergut D, Neufeld JD, Newbold LK, Oliver AE, Pace NR, Palanisamy G, Peplies J, Petrosino J, Proctor L, Pruesse E, Quast C, Raes J, Ratnasingham S, Ravel J, Relman DA, Assunta-Sansone S, Schloss PD, Schriml L, Sinha R, Smith MI, Sodergren E, Spo A, Stombaugh J, Tiedje JM, Ward DV, Weinstock GM, Wendel D, White O, Whiteley A, Wilke A, Wortman JR, Yatsunenko T, Glöckner FO. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat Biotechnol. 2011;29(5):415-20. Epub 2011/05/10.

95. Gonzalez A, Navas-Molina JA, Kosciolek T, McDonald D, Vázquez-Baeza Y, Ackermann G, DeReus J, Janssen S, Swafford AD, Orchanian SB, Sanders JG, Shorenstein J, Holste H, Petrus S, Robbins-Pianka A, Brislawn CJ, Wang M, Rideout JR, Bolyen E, Dillon M,

Caporaso JG, Dorrestein PC, Knight R. Qiita: rapid, web-enabled microbiome meta-analysis. Nat Methods. 2018;15(10):796-8. Epub 2018/10/03.

96. Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, Schwager E, Crabtree J, Ma S; Microbiome Quality Control Project Consortium, Abnet CC, Knight R, White O, Huttenhower C. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat Biotechnol. 2017;35(11):1077-86. Epub 2017/10/03.

97. Pribram HF, Hass AC, Nishioka H. Radiographic localization of a spontaneous cerebrospinal fluid fistula. Case report. J Neurosurg. 1966;24(6):1031-3. Epub 1966/06/01.

98. Lozupone CA, Knight R. Global patterns in bacterial diversity. Proc Natl Acad Sci U S A. 2007;104(27):11436-40. Epub 2007/06/27.

99. Walters WA, Xu Z, Knight R. Meta-analyses of human gut microbes associated with obesity and IBD. FEBS Lett. 2014;588(22):4223-33. Epub 2014/10/14.

100. Zhou Y, Xu ZZ, He Y, Yang Y, Liu L, Lin Q, Nie Y, Li M, Zhi F, Liu S, Amir A, González A, Tripathi A, Chen M, Wu GD, Knight R, Zhou H, Chen Y. Gut Microbiota Offers Universal Biomarkers across Ethnicity in Inflammatory Bowel Disease Diagnosis and Infliximab Response Prediction. mSystems. 2018;3(1). Epub 2018/02/07.

101. Vázquez-Baeza Y, Hyde ER, Suchodolski JS, Knight R. Dog and human inflammatory bowel disease rely on overlapping yet distinct dysbiosis networks. Nat Microbiol. 2016;1:16177. Epub 2016/10/04.

102. Sze MA, Schloss PD. Looking for a Signal in the Noise: Revisiting Obesity and the Microbiome. MBio. 2016;7(4). Epub 2016/08/25.

103. Finucane MM, Sharpton TJ, Laurent TJ, Pollard KS. A taxonomic signature of obesity in the microbiome? Getting to the guts of the matter. PLoS One. 2014;9(1):e84689. Epub 2014/01/15.

104. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van Treuren W, Knight R, Bell JT, Spector TD, Clark AG, Ley RE. Human genetics shape the gut microbiome. Cell. 2014;159(4):789-99. Epub 2014/11/25.

Chapter 2. Intermittent Hypoxia and Hypercapnia Alter Diurnal Rhythms of Luminal Gut Microbiome and Metabolome

Abstract

Rationale

Obstructive sleep apnea (OSA), characterized by intermittent hypoxia and hypercapnia (IHC), affects the composition of the gut microbiome and metabolome. The gut microbiome has circadian oscillations that play a crucial role in regulating circadian and overall metabolic homeostasis. Thus, we hypothesized that IHC adversely alters gut luminal dynamics of key microbial families and metabolites.

Objective

To determine the diurnal dynamics of the fecal microbiome and metabolome of *Apoe-/-* mice after a week of IHC exposure.

Methods and Results

Individually-housed, 10-week-old *Apoe-/-* mice on an atherogenic diet were split into two groups. One group was exposed to daily IHC conditions for 10 hours (ZT2 to ZT12) while the other was maintained in room air. Six days after initiation of the IHC condition, fecal samples were collected every 4 hours for 24 hours (6 timepoints). We performed 16S rRNA gene amplicon sequencing and untargeted LC/MS to assess changes in the microbiome and metabolome. IHC induced global changes in the cyclical dynamics of the gut microbiome and metabolome. Ruminococcaceae, Lachnospiraceae, S24-7, and Verrucomicrobiaceae had the greatest shifts in their diurnal oscillations. In the metabolome, bile acids, glycerolipids (phosphocholines, phosphoethanolamines), and acylcarnitines were greatly affected. Multi-omic analysis of these results demonstrated that Ruminococcaceae and T**β**MCA co-occur and are associated with IHC

conditions, and Coriobacteriaceae and CDCA co-occur and are associated with control conditions.

Conclusions

IHC significantly changes the diurnal dynamics of the fecal microbiome and metabolome, increasing members and metabolites that are pro-inflammatory and pro-atherogenic, while decreasing protective ones.

Importance Statement

People with obstructive sleep apnea, identified by loud snoring and breathing irregularly while sleeping, are at a higher risk of high blood pressure, type 2 diabetes, cardiac arrhythmias, stroke, and sudden cardiac death. We wanted to understand whether the gut microbiome changes induced by obstructive sleep apnea could potentially explain some of these medical problems. By collecting stool from a mouse model of this disease at multiple time points during the day, we studied how obstructive sleep apnea changed the day-night patterns of microbes and metabolites of the gut. Since the oscillations of the gut microbiome play a crucial role in regulating metabolism, changes in these oscillations can explain why these patients can develop so many metabolic problems. We found changes in microbial families and metabolites that regulate many metabolic pathways contributing to the increased risk for heart disease seen in patients with obstructive sleep apnea.

Indexing Terms (3-5 words): Microbiome, Metabolome, Atherosclerosis, Circadian Rhythms

Table 2.1 Non-standard Abbreviations and Acronyms.
The full names of microbes, metabolites, receptors, and other scientific terms are listed in the first column with its corresponding abbreviation used in Chapter 2 in the second column.

NAME	ABBREVIATION
sub-Operational Taxonomic Unit	sOTU
Obstructive Sleep Apnea	OSA
Cardiovascular Disease	CVD
Intermittent Hypoxia and Hypercapnia	IHC
Principal Coordinates Analysis	PCoA
Zeitgeber time	ZT
Cholic acid	CA
Chenodeoxycholic acid	CDCA
Deoxycholic acid	DCA
Hyodeoxycholic acid	HDCA
Tauro-beta-muricholic acid	Tβ MCA
Ursodeoxycholic acid	UDCA
1-(9Z-Octadecenoyl)-sn-glycero-3-phosphoethanolamine	LysoPE (18:1(9Z)/0:0)
1-Stearoyl-2-myristoyl-sn-glycero-3-phosphocholine	PC(18:0/14:0)
Farnesoid X receptor	FXR

2.1 Introduction

In the second part of this thesis, we demonstrate how an atherogenic mouse model of obstructive sleep apnea (OSA) results in altered diurnal microbiome and metabolome patterns. Mice under OSA conditions have increased stress and disrupted behavioral patterns that lead to widespread effects on host physiology. We examined altered diurnal oscillations of the gut microbiome and metabolome in mice under OSA conditions compared to controls. For the microbiome, Ruminococcaceae, Lachnospiraceae, Muribaculaceae (formerly S24-7), and Verrucomicrobiaceae were most affected. Bile acids, glycerolipids, and acylcarnitines had the greatest shifts in diurnal oscillations in the metabolome. Both the microbiome and metabolome had shift that were more proinflammatory and proatherogenic in nature. Multi-omic analysis showed that Ruminococcaceae and tauro-β-muricholic acid (TβMCA) tend to co-occur and elevated levels of both are associated with OSA conditions. In this study, we demonstrated how a timed external environmental host stimulus results in an altered diurnal pattern in the gut microenvironment in a detrimental way.

Obstructive sleep apnea (OSA) is a major risk factor for cardiovascular disease (CVD), including metabolic syndrome, insulin resistance, cardiac arrhythmias, and atherosclerosis.[1] The mechanism(s) of how OSA, or its characteristic components, intermittent hypoxia and hypercapnia (IHC), increases CVD risks is poorly understood, but disruption of circadian rhythms has long been suspected.[2–4] The dyssynchrony between central and peripheral clock machinery could explain why IHC is able to disrupt so many different physiological systems simultaneously.[5–7] OSA can affect the central circadian clock through sleep fragmentation and increased sympathetic tone. However, how OSA affects the agents that entrain peripheral circadian clocks is poorly understood.

Hepatic and intestinal circadian rhythms are entrained by feeding/fasting cycles and the gut microbiome.[7,8] The gut microbiome is necessary for the maintenance of ileal and hepatic

circadian clocks and their synchrony with central circadian rhythms.[9,10] Moreover, the gut microbiome itself has cyclical fluctuations that are necessary for metabolic homeostasis.[9–15] Microbially-produced compounds, such as short-chain fatty acids (SCFAs)[16] as well as deconjugated and secondary bile acids[17], link the luminal environment with host hepatic and ileal circadian rhythms[8]. Given that peripheral circadian rhythms regulate circulating lipids[18], hematopoietic stem cells[19], vascular smooth muscle function, sympathetic tone, and blood pressure[20,21], OSA-induced changes to the gut microbiome could aggravate multiple physiological systems that promote atherosclerosis through their disruption. Moreover, disrupted circadian luminal dynamics can affect gut micro-niches and promote the growth of bacteria that are pro-inflammatory (e.g. Ruminococcaceae[22]) and hinder those that may be protective against CVD (e.g., *Akkermansia*[23]). Disruption in microbiome rhythms can also increase systemic inflammation through disruption of the gut barrier function.[24] Thus, the effect of IHC on luminal circadian dynamics can improve our understanding of how OSA increases CVD risks.

Prolonged IHC exposure in atherosclerotic mouse models results in faster and increased extent of atherosclerotic lesions, making them the preferred animal models of OSA. Interestingly, IHC exposure also alters the composition of the gut microbiome and fecal metabolome in both apolipoprotein E (*Apoe-/-*) and LDL-receptor (*Ldlr-/-*) knockout mice on atherogenic diets[25,26], findings that go beyond what is observed in these mice alone.[27,28] Moreover, the fecal metabolomic changes observed in these mice included metabolites known to affect atherosclerosis, including trimethylamine (TMA), deconjugated and secondary bile acids, fatty acids, and phytoestrogens.[25,26,29] However, some of these luminal metabolites have diurnal fluctuations and are differentially absorbed based on the enterocyte circadian clock.[30,31] Characterizing the time resolution of these changes would further our understanding of how luminal content could contribute to dysmetabolism and atherosclerosis.

To focus solely on the effects of IHC condition on the diurnal dynamics of the gut microbiome and metabolome, we maintained consistent genotype (*Apoe-/-*) and dietary conditions (atherogenic diet) between groups. We wanted to evaluate whether atherogenic gut luminal changes persist through a 24-hour period or only during certain windows of time. Overall, this study tests the hypothesis that IHC disrupts the diurnal rhythms of the gut microbiome and metabolome which may promote a pro-inflammatory luminal environment.

2.2 Methods and Materials

Animal Model and Description of IHC Induction

Individually-housed, 10-week-old *Apoe*$^{-/-}$ mice (littermates from Jackson Labs) were fed an Irradiated regular chow diet (Envigo Teklad LB-485 diet #7912) prior to experiment start. Upon starting the experiment, the mice were fed an atherogenic diet consisting of 1.25% cholesterol and 21% milk fat (4.5 Kcal/g; TD.96121; Envigo-Teklad Madison, WI) *ad libitum*. There were 8 mice total with 4 mice pseudo-randomly assigned to each of 2 groups (based on weight and acclimation cage). Vivarium was maintained in a 12:12 light-dark cycle room at 68-72ºF and below 40% humidity during the study. IHC exposure occurred as previously described.[25] In brief, mice were exposed to 10 hours of IHC conditions during the light period (ZT2-12). IHC exposure consisted of 4 minutes of synchronized reduction of O_2 from 21% to 8% with synchronized elevation of CO_2 from 0.5% to 8%, followed by alternating periods of 4 minutes of normoxia and normocapnia with one to two minute ramp intervals. Control mice were kept at room air (21% O_2 and 0.5% CO_2) for the duration of the experiment. Introduction of atherogenic diet and IHC conditions both commenced on Day 0. One week was chosen so that the mice would have several days adjustment to the IHC chamber**.** We wanted to focus on early environmental-induced changes, which drives the phenotype, rather than later in phenotype development where the effects of the dysmetabolic state could affect gut microbiome composition. All animal experiments

were conducted in accordance with the guidelines of the IACUC of the University of California, San Diego.

After one week of maintaining the mice in their relative environmental conditions, fecal samples were collected every 4 hours (ZT 2, 6, 10, 14, 18 and 22) over a 24-hour period. After collection, fecal samples were immediately stored at -80ºC until the end of the study. The microbiome was characterized by 16S rRNA amplicon sequencing and the metabolome was characterized by untargeted liquid chromatography-tandem mass spectrometry (LC-MS/MS) in a manner consistent with previous studies.[25,26]

Characterization of Microbiome

DNA extraction and 16S rRNA amplicon sequencing were done using the standard protocol for the Earth Microbiome Project (http://www.earthmicrobiome.org/protocols-and-standards/16s).[32] In brief, DNA was extracted using the Qiagen PowerSoil DNA extraction kit (Qiagen, Carlsbad, CA). The resulting DNA library was prepared for 16S rRNA amplicon sequencing as described previously.[32] These pooled samples were purified using Qiagen UltraClean PCR cleanup kit (Qiagen, Carlsbad, CA) and then sequenced on the Illumina HiSeq 2500 sequencing platform. The V4 region of 16S rRNA gene was sequenced using the primer pair 515F-806R with Golay error-correcting barcodes on the reverse primer.[33]

Raw sequence data were uploaded to Qiita[34] (https://qiita.ucsd.edu/) and processed using the Deblur[35] workflow with default parameters. There were 2,490,504 reads for 94 samples with an average of 28,519 reads per sample. This process generated a BIOM[36] format table that was rarefied to a depth of 12,000 sequences/sample to control for sequencing effort. This process removed two samples from analysis, one from Air (ZT2) and one from IHC (ZT14) that had read counts similar to blanks. A rooted phylogenetic tree was inferred using SATé-enabled phylogenetic placement plug-in[37] using QIIME 2[38] version 2019.10, which was used to insert 16S Deblur sub-Operational Taxonomic Units (sOTUs) into Greengenes[39] 13_8 at 99%. Because it takes into account both phylogeny and abundance, Weighted UniFrac[40] distances were used for microbiome principal coordinates analysis (PCoA) plots. Overall group differences were tested using PERMANOVA.[41] All sOTUs were collapsed to the phylum and family level for analysis and comparison because critical sOTUs that distinguish IHC and Air from each other did not identify past family level. MetaCycle using JTK method and correction for multiple hypotheses with Fisher's method was applied to determine circadian oscillatory patterns.[42] Rhythmicity was tested using LimoRhyde.[43] Mann-Whitney-Wilcoxon test was used to compare groups presented in the

boxplots. Data visualized using custom python scripts - GitHub link for Python code [https://github.com/knightlab-analyses/circadian-ihc/].

Characterization of Metabolome

Untargeted liquid chromatography-tandem mass spectrometry (LC-MS/MS) was performed on the stool for metabolomics as previously described.[25] In brief, samples were prepared by adding 500 µl of 50:50 methanol-H_2O to all fecal samples (30 to 50 mg approximately) followed by homogenization and centrifugation. Then, 450µL of the resulting supernatant was transferred to a 96-well deep-well plate and dried using centrifugal evaporation (CentriVap centrifugal vacuum concentrator; Labconco, Kansas City, MO). Next, samples were resuspended and samples were analyzed on a Vanquish ultrahigh-performance liquid chromatography (UPLC) system coupled to a Q Exactive orbital ion trap (Thermo Fisher Scientific, Bremen, Germany). For chromatographic separation, a C18 core shell column (Kinetex column, 50 by 2 mm, 1.7-µm particle size, 100-Å pore size; Phenomenex, Torrance, CA) with a flow rate of 0.5 ml/min (solvent A, H2O-0.1% formic acid [FA]; solvent B, acetonitrile-0.1% FA) was used. Flow-through parameters were set, run, and data were detected as previously described.[25]

The raw data resulting from the method described above were converted to m/z extensible markup language (mzXML) in centroid mode using MSConvert (part of ProteoWizard).[44] The mzXML files were cropped using an m/z range of 75.00 to 1,000.00 Da for further sample processing. Using a signal threshold of 2.0e5 and a 0.3-s minimum peak width to remove low quality spectra, MZmine2[45] (http://mzmine.sourceforge.net/) was used for feature extraction. The local minimum search algorithm was used with a minimum relative peak height of 1% for chromatographic deconvolution. Minimum retention time range was set to 0.6 seconds. Maximum peak width was set to 1 min. After that, alignment of the peak lists of all samples was performed with retention time deviation of 10s and mass tolerance of 10ppm was set for features. Next, MZmine2[45] (version 2.37) was used to create a feature matrix file that could then be linked

to metadata. The signal intensities of the features were normalized for subsequent analysis. Identification of molecular features was performed using MS1-based feature detection and MS2-based molecular networking using the GNPS[46] workflow (https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp). The actual GNPS jobs can be found at the following URL: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=9d8ee19ec2654d46a065080a0ff2290a. Using a decoy database approach in GNPS, we determined that the false-discovery rate (FDR) was less than 1% above a 0.6 cosine similarity score.[47] Thus, we used a cosine score of 0.65 to determine annotations. The MS/MS spectral annotations were determined by using MS/MS-based spectral library matches for GNPS level 2 or 3 annotation for all molecules.[48] Bile acid standards were purchased from Cayman Chemical (Ann Arbor, MI) and analyzed using the same LC-MS/MS method described above to attain level 1 identification as defined by the 2007 metabolomics standards initiative.[49] The annotated frequency table was analyzed using QIIME 2[38] version 2019.10. Canberra distances[50], which are more sensitive to rare features than Bray-Curtis, were used for metabolomic PCoA plots and significance was tested using PERMANOVA[41]. MetaCycle[42] utilizing the JTK method was used to determine circadian rhythmicity. Mann-Whitney-Wilcoxon test was used to compare groups presented in the boxplots. Data was visualized using custom python scripts - GitHub link for Python code [https://github.com/knightlab-analyses/circadian-ihc/].

Differential Abundance and Multi-Omic Analysis

Differential abundance analysis was performed with Songbird which accounts for the compositional nature of microbial data and uses a multinomial regression model to estimate differential ranks.[51] Optimized model parameters for the covariate of exposure treatment and the interaction of time in hours were determined for the microbiome and metabolomics data

(differential prior = 1.5 and learning rate = 0.001) and compared to a baseline model of 1 on the same parameters. Model fit was compared by the Q-squared (1 − model CV/baseline CV) where a Q-squared greater than zero ensures a good model fit. A Q-squared of 0.14 and 0.22 was obtained for the microbiome and metabolite data respectively. The resulting differentials were explored and verified through log-ratios through Qurro.[52] Multi-omic analysis of microbiome and metabolomics data was performed through mmvec (microbe–metabolite vectors), a neural network method for producing log conditional probabilities of co-occurrence between microbial and metabolite features visualized as heatmaps and paired latent representation in few dimensions which can be visualized in scatter or biplot ordinations.[53] A high conditional probability of close spatial similarity in the ordination indicates high co-occurrence between a microbe and metabolite pair while a negative conditional probability or high spatial distance in the ordination indicates low co-occurrence. The mmvec model parameters were optimized (batch-size = 5, learning-rate = 1e−3) to minimize the low cross-validation error and model likelihood. Differential co-occurrence patterns (mmvec microbe-metabolite interactions) in relation to the exposure treatment was evaluated by correlating MMVEC PC1 loadings with respect to the songbird log fold change differential with respect to the exposure treatment, which was significantly correlated (Metabolites; Spearman rho=0.26, P=1.23e-07; see results).

Data availability

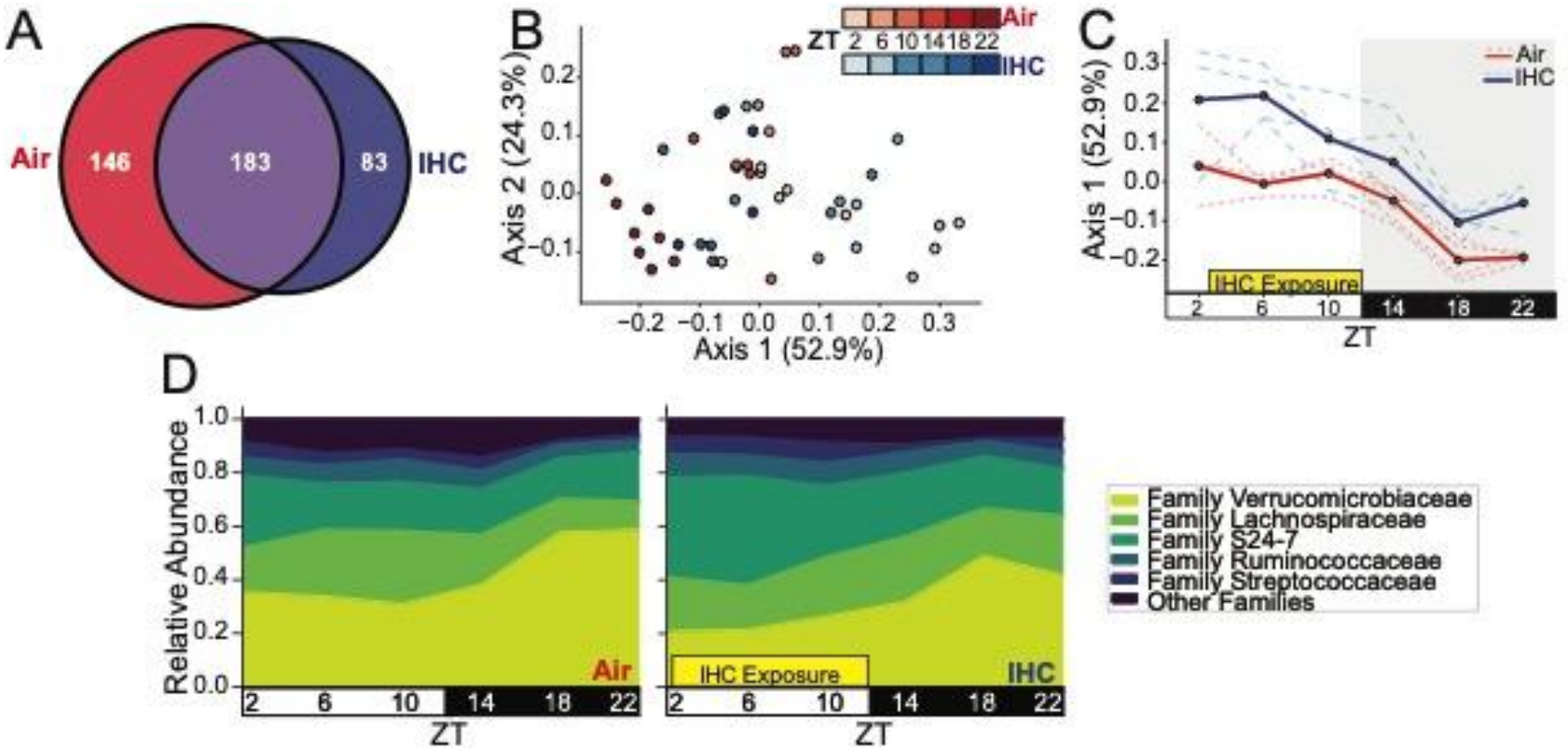
EBI accession for microbiome: ERP110592; MassIVE ID for metabolome is MSV000084847, link: https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=0434de6d06f5424d8bd567808d069d4e ; GitHub link for Python code [https://github.com/knightlab-analyses/circadian-ihc/].

2.3 Results

IHC Changes the Composition of the Gut Microbiome

To determine the early effects of OSA on the gut microbiome, we used the atherosclerosis-prone *Apoe-/-* murine model and exposed half of the cohort to IHC conditions, while control mice were exposed to room air. Since mice were individually housed for the experiment, cage effects were not a confounding variable. Our analysis revealed a large overlap in the sub-operational taxonomic units (sOTUs) (Fig. 2.1A), with 183 sOTUs in common between the two groups. These shared sOTUs comprised 56% of the total Air sOTUs and 69% of the total IHC sOTUs. The Air control group tended to have more sOTUs than the IHC group (Fig. 2.S1A, Mann-Whitney U test, p=0.053). In addition, mice in the Air condition had overall higher Faith's phylogenetic **α**-diversity compared to mice under the IHC condition (Fig. 2.S1B, Kruskal-Wallis, p=0.040). This finding is especially true for ZT-18, where **α**-diversity values diverged the most between the two groups (Mann-Whitney U, p=0.015). Gut microbiome biodiversity, such as measured by Faith's phylogenetic diversity, has previously been used as a surrogate measure of microbial community health.[54]

Figure 2.1. IHC Affects the Cyclical Dynamics of the Gut Microbiome. A) Venn diagram of unique non-zero detected sOTU in each cohort overall. Purple indicates sOTUs in common. B) Weighted UniFrac (beta diversity) PCoA of samples. Shading represents different time points as indicated. C) Weighted UniFrac PCoA of only Axis 1 over time. Significance tested with paired Wilcoxon test. Solid line indicates average for group, dotted line indicates individual mice. D) Proportional abundance representation of the top 5 microbial families. Notation: Control samples only exposure to normal Air conditions is red (n=4, 5-6 time points per mouse); Experimental samples exposed to IHC conditions are blue (n=4, 5-6 time points per mouse).


A
Air
146
183
83
IHC
B
Air
ZT 2 6 10 14 18 22
IHC
Axis 2 (24.3%)
Axis 1 (52.9%)
C
Air
IHC
Axis 1 (52.9%)
IHC Exposure
ZT
D
Relative Abundance
Air
IHC Exposure
IHC
ZT
Family Verrucomicrobiaceae
Family Lachnospiraceae
Family S24-7
Family Ruminococcaceae
Family Streptococcaceae
Other Families

Weighted UniFrac distances, a measure of **β**-diversity, were significantly different between the two groups across all time points (Fig. 2.1B, S1C) (pseudo-F 6.776, p=0.002 by PERMANOVA[41]). Analysis of these **β**-diversity distances demonstrates that the microbiome composition of the IHC in the dark period became more similar to that of the Air controls during the light period (Fig 2.1B, 1C). Compositional analysis showed Verrucomicrobiaceae, Lachnospiraceae, and S24-7 were the top three most prevalent taxa of the microbiome in both conditions across all time points (Fig 2.1D). Family S24-7 ("Homeothermaceae" is one of the current proposed names) is a relatively new family in the phylum Bacteroidetes, thought to be involved in carbohydrate metabolism, among other functions.[55] To determine whether IHC disrupted the diurnal dynamics of the luminal environment, we determined the proportion of bacterial families that had circadian oscillation in their abundance ($p<0.05$).[42] Compared to control mice, the mice under the IHC condition had about two-thirds as many bacterial families (Air 13%, IHC 8%) that had circadian oscillations (Fig. S1D**)**, and these families accounted for fewer reads (Fig. S1E). A full breakdown of changes in oscillation at the sOTU level is found in Supplementary Table 2.3. Thus, IHC caused a significant change in luminal dynamics over the course of 24-hours. This shift in luminal dynamics is characterized by a decrease in overall cycling, wherein the dark period microbial composition in IHC is more similar to that of the light period in control mice. We did not see significant changes in rhythmicity.[43]

IHC Exposure Results in Circadian Disruption of the Gut Microbiome

To examine the effects of IHC on the diurnal dynamics of the microbiome in more detail, we examined individual taxa over time. IHC induced dynamic changes in the composition of the microbiome that was easily detectable at the phylum level (Fig 2.S2A-C). There was no detectable circadian fluctuation in the Bacteroidetes phylum (Fig. 2.S2A). In the control condition, Firmicutes had circadian fluctuations with a peak in the late light period (Fig. 2.S2B), and Verrucomicrobia had circadian fluctuations with a peak in the late dark period (Fig. 2.S2C). Phylum

Verrucomicrobia contains only the sOTU for *Akkermansia muciniphila*. However, under IHC conditions, Firmicutes lost, and Bacteroidetes gained, circadian oscillation (Fig. 2.S2A-B). Circadian oscillation was maintained in Verrucomicrobia under the IHC condition (Fig. 2.S2C). However, relative abundances of this phylum were significantly higher in the Air controls during both light and dark periods (Fig. 2.S2C).

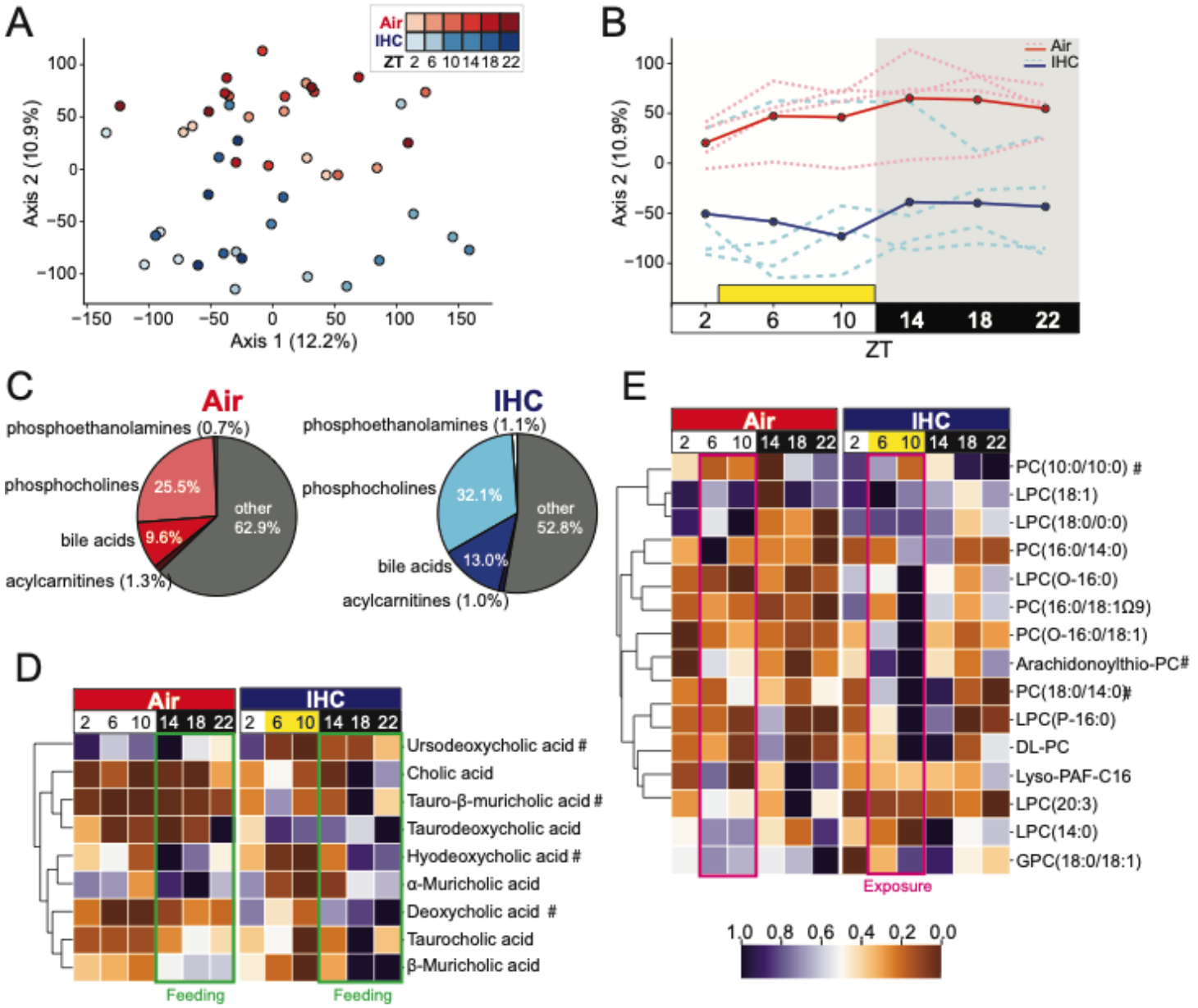
Then, we analyzed the cyclical dynamics of Ruminococcaceae and Lachnospiraceae, two bacterial families in the Firmicutes phylum that have been associated with atherosclerosis formation.[22] Both Ruminococcaceae (Fig. 2.S2D) and Lachnospiraceae (Fig. 2.S2E) have circadian oscillations in the control mice that are perturbed under the IHC condition. S24-7, a bacterial family in the Bacteroidetes phylum, was the only one that had more robust cycling under the IHC condition than it did under the control conditions (Fig. 2.S2F). In particular, S24-7 abundances increased significantly during the time of IHC exposure, where they were 2- to 3-fold higher than what was measured under the control condition. In contrast, peak differences in the relative abundance of Lachnospiraceae occurred during the dark period, where abundances were approximately 2-fold higher under the IHC condition. Overall, IHC perturbed the cyclical dynamics of the bacterial families that can affect the atherosclerosis phenotype.

IHC Changes Composition of the Fecal Metabolome

An overview of the fecal metabolome, analyzed using untargeted LC-MS/MS, shows significant separation between groups [pseudo-F 5.410, $p<0.001$ by PERMANOVA[41]] (Fig. 2.2A, 2.S3A. Axis 2 (10.9%) which represents a nearly similar amount of the variability in the data as Axis 1 (12.2%), shows a clearer separation between groups over time Fig. 2.2B. Overall, the IHC condition led to an increase in the relative amount of fecal bile acids (~35%) as well as glycerolipids, such as phosphoethanolamines (~57%) and phosphocholines (~26%) (Fig 2.2C). The cyclical dynamics of these individual metabolites, which have been implicated in

atherosclerosis, were significantly affected by IHC exposure (Fig. 2.2D-E, 2.S3B. A list of all annotated metabolites can be found in Supplemental Table 2.2.

Figure 2.2. IHC Affects the Cyclical Dynamics of the Fecal Metabolome. Please see Supplemental Table A.A.2.S1 for a list of the full annotation and abbreviation of metabolites displayed. A) Canberra PCoA of metabolomics samples, shading represents different time points. Significance determined by PERMANOVA. B) Canberra PCoA of Axis 2 over time. Solid line indicates mean for group, dotted line indicates individual mice. Yellow box indicates time under IHC exposure for the treatment group. C) Pie charts of key groups of metabolites, separated by condition. D) Heatmap of level 1 bile acids, organized using hierarchical clustering based on controls. Yellow indicates time under IHC exposure for the treatment group. For other level 3 bile acids, see Figure S3C. E) Heatmap of selected phosphocholines, organized using hierarchical clustering based on controls. The value of each square of the heatmap represents the mean relative abundance value for all mice in that condition for that time point. The heatmaps are also row normalized across both conditions and placed on a standard scale referenced in the center to allow easier comparison. Notation: # indicates a metabolite that is also shown in Figure 4 or Figure 2.S4. Air is red (n=4, 5-6 time points per mouse); IHC is blue (n=4, 5-6 time points per mouse).


A
Air
IHC
ZT 2 6 10 14 18 22
Axis 2 (10.9%)
Axis 1 (12.2%)
B
Air
IHC
Axis 2 (10.9%)
ZT
C
Air
phosphoethanolamines (0.7%)
phosphocholines
25.5%
other
62.9%
bile acids
9.6%
acylcarnitines (1.3%)
IHC
phosphoethanolamines (1.1%)
phosphocholines
32.1%
other
52.8%
bile acids
13.0%
acylcarnitines (1.0%)
D
Air
IHC
Ursodeoxycholic acid #
Cholic acid
Tauro-β-muricholic acid #
Taurodeoxycholic acid
Hyodeoxycholic acid #
α-Muricholic acid
Deoxycholic acid #
Taurocholic acid
β-Muricholic acid
Feeding
Feeding
E
Air
IHC
PC(10:0/10:0) #
LPC(18:1)
LPC(18:0/0:0)
PC(16:0/14:0)
LPC(O-16:0)
PC(16:0/18:1Ω9)
PC(O-16:0/18:1)
Arachidonoylthio-PC#
PC(18:0/14:0)#
LPC(P-16:0)
DL-PC
Lyso-PAF-C16
LPC(20:3)
LPC(14:0)
GPC(18:0/18:1)
Exposure

Sub-class analysis of the fecal metabolites demonstrated that time was an important factor in metabolomic differences. For example, the biggest differences in bile acids between IHC-conditioned mice and Air controls occurred in the dark period (Fig. 2.2D), except for UDCA which demonstrated the greatest changes during the light period. Phosphocholines, phosphoethanolamines, and acylcarnitines, metabolites that play an important role in atherosclerosis, are also altered under the IHC condition.[56–61] IHC resulted in comparatively higher levels of phosphocholines (Fig. 2.2C, E) and phosphoethanolamines (Fig. 2.2C, 2.S3B) during IHC exposure (i.e., light period) (Fig. 2.2E, 2.S3B). In addition, IHC mice also had higher levels of heme breakdown derivatives - stercobilin and urobilin - during the dark period (Fig. 2.S3B), which may be an indication of altered liver metabolism.[62] Overall, fecal IHC induced global changes in fecal metabolomics, particularly in secondary metabolites that are known to contribute to inflammation and atherosclerosis.

Diurnal Dynamics of the Fecal Metabolome is Altered Under IHC Conditions

We performed a more detailed analysis of the effects of IHC on the diurnal dynamics of the luminal metabolites that are presumed to either exacerbate or protect against atherosclerosis, affect circadian rhythms, or influence metabolic homeostasis. Bile acids are one of the key metabolites that can influence peripheral circadian rhythms, host metabolism, and atherosclerosis.[63] Deoxycholic acid (DCA), a pro-inflammatory secondary bile acid[64], had circadian oscillations under the IHC condition but not under the control condition (Air $p = 1.000$, IHC $p = 0.002$) (Fig. 2.S4A). The relative abundance of DCA was higher under IHC condition compared to control conditions at all time points, but especially so during the dark period ($p = 0.001$) (Fig. 2.S4A). Tauro-**β**-muricholic acid (T**β**MCA), known to contribute to the development of atherosclerosis through farnesoid X receptor (FXR) antagonism[65–67], had overall higher relative abundance levels in mice under IHC conditions (Fig.2.S4B). Significantly high levels of T**β**MCA were observed in IHC-conditioned mice during the light period ($p = 0.019$), with dark period

differences only approaching statistical significance (p = 0.063). T**β**MCA appears to have 12hr oscillations rather than 24hr oscillations. While pro-atherosclerotic T**β**MCA did not have a circadian oscillation under Air nor IHC conditions (Air p = 0.270, IHC p = 1.000) (Fig. 2.S4B), anti-atherosclerotic hyodeoxycholic acid (HDCA)[68,69] displayed circadian oscillations under IHC but not under control conditions (Air p = 1.000, IHC p = 0.002) (Fig. 2.S4C). Levels of anti-inflammatory ursodeoxycholic acid (UDCA)[68,70,71] trended toward being lower in mice under IHC condition, however this did not approach significance (Fig. 2.S4D). Overall, bile acid differences between IHC and Air conditions were far more pronounced during the dark period (Fig. 2.2D). Furthermore, with increased levels of pro-inflammatory and pro-atherosclerotic bile acids and a reduction in anti-inflammatory bile acids, IHC conditions appear to shift the metabolome in a detrimental direction for the host.

IHC had a significant effect on the cyclical dynamics of other metabolites that are important for atherosclerosis, such as acylcarnitines and glycerolipids, including phosphoethanolamines and phosphocholines (Fig. 2.S5). High serum levels of long-chain acylcarnitines, such as palmitoylcarnitine (Fig. 2.S5A) and oleoyl L-carnitine (Fig. 2.S5B) can promote inflammation and atherosclerosis.[72,73] However, fecal palmitoylcarnitine levels are significantly lower in IHC-conditioned mice during both periods of the circadian cycles (light period p = 0.013, dark period p = 0.017) (Fig. 2.S5A). Overall levels of oleoyl L-carnitine, another acylcarnitine, are also reduced in IHC-conditioned mice, particularly during the light period of the circadian rhythms (light period p=0.053, dark period p=0.761) (Fig. 2.S5B). Phosphoethanolamines and phosphocholines are the main pro-inflammatory glycerolipid derivatives that are absorbed into the bloodstream.[74] Fecal levels of phosphoethanolamines were generally increased in IHC-conditioned mice, especially during the light period (Fig. 2.2). Levels of one such phosphoethanolamine, LysoPE (18:1(9Z)/0:0) (1-(9Z-Octadecenoyl)-sn-glycero-3-phosphoethanolamine), had robust circadian oscillations (p < 0.001) under IHC conditions with significantly higher levels during the light period (p < 0.001) (Fig.

2.S5C). In addition, overall levels of phosphocholines were also increased in IHC-conditioned mice, especially during the light period (Fig. 2.2). Levels of one phosphocholine in our study, PC(18:0/14:0) (1-Stearoyl-2-myristoyl-sn-glycero-3-phosphocholine), had robust cycling under the IHC condition, but not under the air condition (Air p = 1.000, IHC p = 0.002), and was especially elevated during the light period (p = 0.013) (Fig 2.S5D). Similar trends are seen in other phosphocholines (Fig. 2.S5E, 2.S5F). Taken together, our results demonstrate that IHC induces a rapid shift in the gut luminal metabolite profile.

Trans-omic analysis of the microbiome and metabolome reveals key relationship between Ruminococcaceae and TbMCA

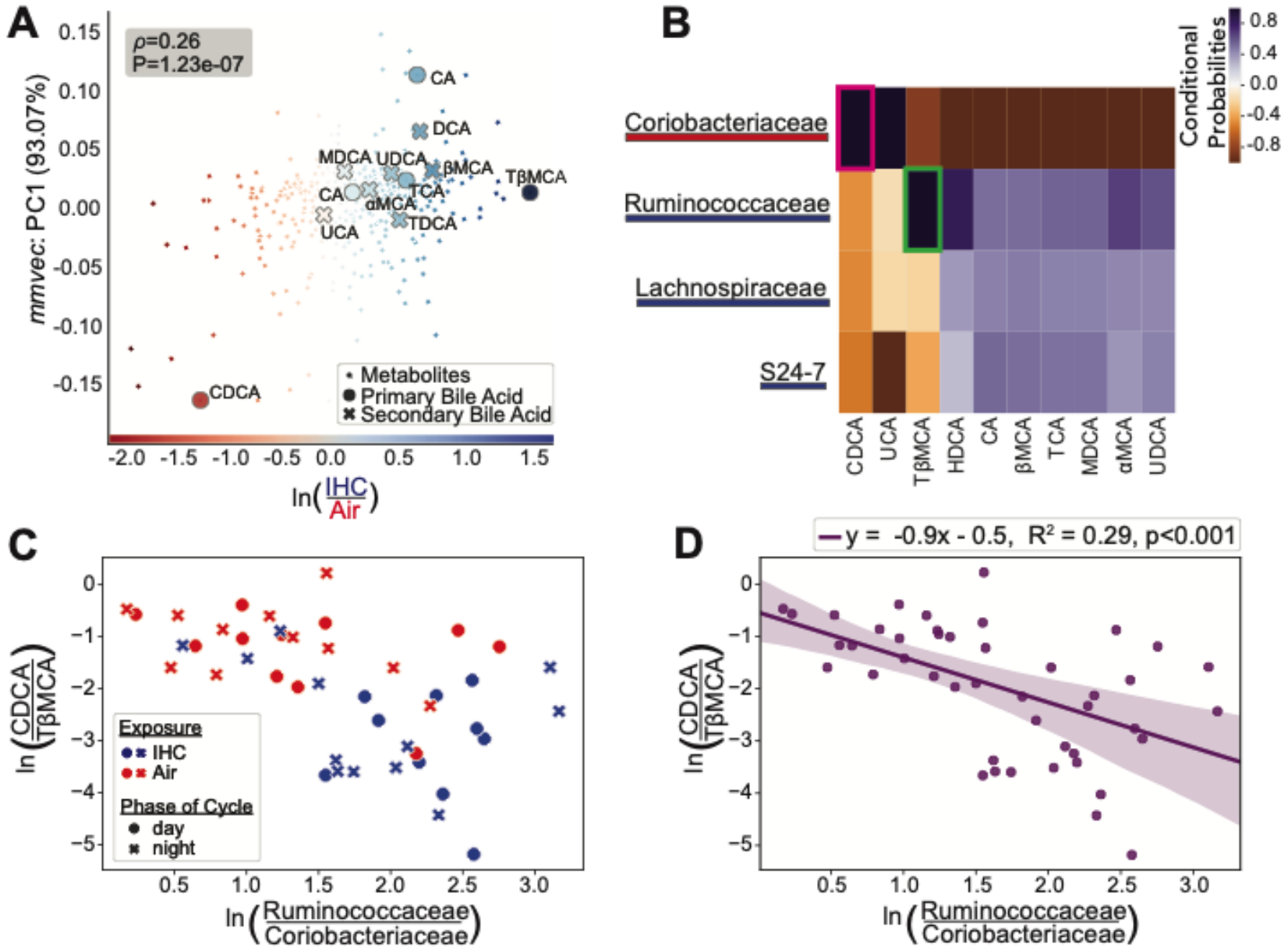
Next, we assessed whether there are specific relationships between microbial families and metabolites that are individually implicated in worsening CVD. One particular challenge in performing multi-omic analysis with microbiome data is that 16S amplicon sequencing yields sum constraint normalized data (i.e., relative abundances). Thus, this increases the probability of type I errors in the analysis and makes measurements of false discovery rates difficult.[51] Relative abundance values can fluctuate significantly from study to study due to artifactual differences in total number of microbial reads (i.e. total feature load). For example, when the relative abundance of a specific bacterial family is increased, we cannot determine if this is due to an increase in the number of bacteria within that family, or if it is due to a decrease in the number of other bacteria in other families. By using log-ratios for these analyses, we remove the biases created by total feature load and can calculate false discovery rates using already established methods.[51,75] Thus, these log ratio-based methods are more likely to result in repeatable trends in independently performed studies.

We used a machine learning neural network to predict the probability of microbe-metabolite interactions (*mmvec*[53]) as well as a multinomial logarithmic regression differential ranking analysis (*songbird*[51]). This analysis created ranked log-based conditional probabilities of

microbe and metabolite co-occurrence and identified relationships between the microbiome and metabolome data, which were then validated using the individual log-ratios. This analysis not only revealed which microbes and metabolites co-occur, but also whether they are correlated with differences between the IHC and control conditions.

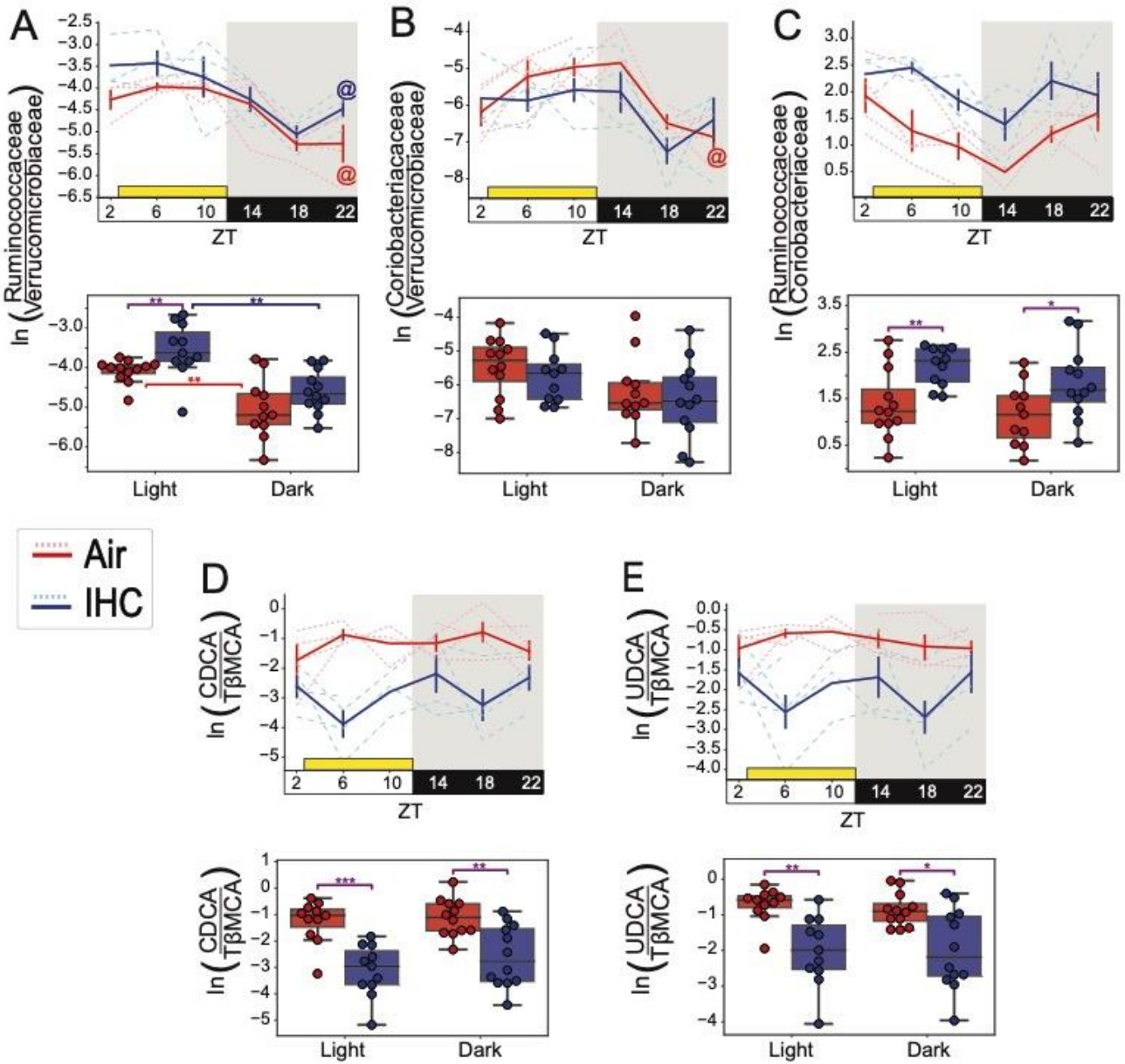
Based on this multi-omic analysis, CDCA and T**β**MCA are the two bile acids most differentially abundant between the IHC and control conditions (Fig. 2.3A). CDCA, which is associated with the Air condition, co-occurred with Coriobacteriaceae (Fig. 2. S2G), and T**β**MCA, which is associated with the IHC condition, co-occurred with Ruminococcaceae (Fig 2.A, B). Plotting the microbes and metabolites identified by *mmvec* validated its co-occurrence status (Fig. 2.A, 2.3B). The log-ratios of the top differentially co-occurring microbes and metabolites result in separation of the two conditions. However, further studies are necessary to determine whether these relationships are causal. We found a significant correlation (Pearson's correlation coefficient, $r = -0.539$, $R^2 = 0.29$, $p < 0.001$) between the log-ratio of the *mmvec* identified metabolites (CDCA in the control condition and T**β**MCA in the IHC condition) and the *mmvec* identified microbes (Coriobacteriaceae in the control condition and Ruminococcaceae in the IHC condition) (Fig. 2.3C, D). Hence, after one week of IHC exposure, the cyclical dynamics of the luminal environment show significant shifts in the fecal microbiome and metabolome. The shift towards FXR antagonism via T**β**MCA and correlated to Ruminococcaceae is correlated to the pathophysiology of OSA-related CVD.

Figure 2.3. Microbes and metabolites with linked expression levels as determined by mmvec analysis. A) Mmvec53 co-occurrence analysis (y-axis) based on songbird51 multinomial regression differential ranking analysis (x-axis). Bile acids are generally level 1 identification, except for one of the CA, CDCA, and MDCA (level 2 annotation). B) Log conditional probability heatmap, organized using hierarchical clustering, with the top 4 differentially abundant microbial families and the top differentially abundant bile acids. Pink and green boxes highlight the top 2 points with highest correlation values. C) log-ratios of the top correlated microbes (x-axis) and metabolites (y-axis) identified in B. Microbial logratio of all reads from sOTU that belong to family Ruminococcaceae divided by all reads from sOTU that belong to family Coriobacteriaceae. Metabolite logratio of raw values from CDCA divided by raw values of T**β**MCA. D) Linear regression plot using the same log-ratios as in C with best fit line and shaded areas representing 95% confidence interval. Log ratios based on natural log (ln). Control samples only exposure to normal Air conditions are red (n=4, 5-6 time points per mouse); Experimental samples exposed to IHC conditions are blue (n=4, 5-6 time points per mouse). Complete metadata can be found in Supplemental Table A.A.2.S4.


A
ρ=0.26
P=1.23e-07
mmvec: PC1 (93.07%)
0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15
CA
DCA
MDCA
UDCA
βMCA
TβMCA
CA
TCA
αMCA
TDCA
UCA
CDCA
Metabolites
Primary Bile Acid
Secondary Bile Acid
-2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5
ln(IHC/Air)
B
Conditional Probabilities
0.8
0.4
0.0
-0.4
-0.8
Coriobacteriaceae
Ruminococcaceae
Lachnospiraceae
S24-7
CDCA
UCA
TβMCA
HDCA
CA
βMCA
TCA
MDCA
αMCA
UDCA
C
ln(CDCA/TβMCA)
0
−1
−2
−3
−4
−5
Exposure
IHC
Air
Phase of Cycle
day
night
0.5 1.0 1.5 2.0 2.5 3.0
ln(Ruminococcaceae/Coriobacteriaceae)
D
y = -0.9x - 0.5, R² = 0.29, p<0.001
ln(CDCA/TβMCA)
0
−1
−2
−3
−4
−5
0.5 1.0 1.5 2.0 2.5 3.0
ln(Ruminococcaceae/Coriobacteriaceae)

However, relative abundance analyses did not show a significant difference between IHC and control conditions in these bacterial families (Fig. 2.S2D, G). To determine whether total feature load differences could have biased these results, we repeated our assessment of these two bacterial families using log-ratios. For the denominator of these log-ratios, we used Verrucomicrobiaceae due to its ubiquity and high abundance. As predicted by the differential rankings analysis (Fig. 2.3B), mice under the IHC condition had higher log-ratios of Ruminococcaceae to Verrucomicrobiaceae especially during the light period (Fig. 2.4A). Conversely, mice under the Air condition had higher log-ratios of Coriobacteriaceae to Verrucomicrobiaceae during the light/inactive period (Fig. 2.4B). Log-ratios of Ruminococcaceae to Coriobacteriaceae shows persistently higher levels of Ruminococcaceae in mice under IHC conditions, whereas mice under control conditions have higher relative levels of Coriobacteriaceae (Fig. 2.4C).

Figure 2.4. Cyclical Dynamics of log-ratios of Key Microbes and Metabolites. Additional selected log ratios (natural log, ln) and their cyclical dynamics over time (double line plot, upper) and their relative abundances grouped by cycle phase (box plots, lower). A) Log ratio of all reads from sOTU's that belong to family Ruminococcaceae divided by all reads from sOTU's that belong to family Verrucomicrobiaceae. B) log-ratio of all reads from sOTU's that belong to family Coriobacteriaceae divided by all reads from sOTU's that belong to family Verrucomicrobiaceae. C) log-ratio of all reads from sOTU's that belong to family Ruminococcaceae divided by all reads from sOTU's that belong to family Coriobacteriaceae. D) log-ratio of raw values from CDCA divided by raw values of T**β**MCA, the two most differentially abundant bile acids identified in Figure 2.5B. E) log-ratio of raw values of UDCA divided by the raw values of T**β**MCA. Solid line represents the mean, error bars indicate standard error of the mean. Individual mice indicated by dashed line. Shading indicates when room lights are off (i.e. active/feeding time for the mice). Yellow square indicates the 10 hours of the day where mice under the IHC condition would be exposed to experimental conditions [ZT2 (after collection), until ZT12]. MetaCycle with JTK method was used to determine cyclicity. Notation: * = $p<0.05$; ** = $p<0.01$, *** = $p<0.001$, ****=$p<0.0001$ by Mann-Whitney-Wilcoxon test. @ indicates circadian oscillations as determined by MetaCycle (JTK) = $p<0.05$. Control samples only exposure to normal Air conditions are red (N=4, 5-6 time points per mouse); Experimental samples exposed to IHC conditions are blue (N=4, 5-6 time points per mouse). Error bars were not placed for time points where there were fewer than 3 log-ratios available.


A
B
C
D
E
ln (Ruminococcaceae / Verrucomicrobiaceae)
ln (Coriobacteriaceae / Verrucomicrobiaceae)
ln (Ruminococcaceae / Coriobacteriaceae)
ln (CDCA / TβMCA)
ln (UDCA / TβMCA)
ZT
2 6 10 14 18 22
Light
Dark
Air
IHC

Log-ratio analysis of two most different bile acids (i.e., CDCA and T**β**MCA) also showed clear differences between IHC and control mice. As predicted by the differential rankings analysis (Fig. 2.3A, B), mice under the IHC condition have far more fecal T**β**MCA relative to CDCA compared to mice under the control condition (Fig. 2.4D). Interestingly, CDCA is a FXR agonist, while T**β**MCA is an FXR antagonist. FXR is a bile sensor that regulates bile acid uptake, metabolism, and excretion that can affect atherosclerosis.[66] Though the *mmvec* analysis did not identify UDCA, a secondary bile acid with anti-inflammatory properties [68,70,71], as being associated with any specific bacterial family or with any condition, we repeated the log-ratio analysis with this secondary bile acid. In this case, mice under the IHC condition have far more of the pro-atherosclerotic bile acid T**β**MCA relative to UDCA compared to mice under the control condition (Fig. 2.4E).

2.4 Discussion

In this study, we demonstrate that the compositional and diurnal dynamics of the microbiome and metabolome of atherogenic *Apoe-/-* mice are significantly impacted by IHC conditions. Despite the low number of mice used in this study (n=4), we have found significant differences in both microbial families and metabolites at different time points. IHC exposure leads to significant changes in diurnal oscillations of secondary metabolites that are key contributors to the pathogenesis of atherosclerosis. It is currently unclear if gain or loss of oscillations is more important for disease outcomes based on the literature. Taken together, IHC results in circadian dyssynchrony of the gut microbiome and metabolome, which promotes a pro-inflammatory luminal environment through which atherosclerosis is exacerbated. Moreover, it suggests that bile acid signaling and disturbed peripheral circadian rhythms likely contribute to IHC-induced increase in cardiovascular risk.

Several studies, both in mice and humans, demonstrate the metabolic implications of disruptions to the circadian homeostasis of the microbiome.[7,8] In humans, jet lag-induced disruption of microbiome diurnal dynamics leads to increased adiposity in gnotobiotic mice.[10] More recently, a large-scale epidemiological study demonstrated that circadian dyssynchrony of microbial oscillations was associated with type 2 diabetes. Moreover, these arrhythmic microbial risk signatures were highly predictive of metabolic disease.[76] Disruption of the circadian dynamics of the microbiome is a hallmark of animal models of obesity and dysmetabolism.[9–11] In this study, we show that circadian disruption of the gut microbiome is also a hallmark of preclinical models of OSA. Since the gut microbiome is necessary to entrain peripheral circadian rhythms, OSA could contribute to dysmetabolism by inducing circadian dyssynchrony. This finding could explain why OSA can increase CVD risk across so many different physiological systems (e.g. hypertension, insulin resistance).

Our previous work in preclinical models of OSA demonstrates that IHC induces reproducible microbiome and metabolome changes across two different mouse models of atherosclerosis, *Apoe-/-* and *Ldlr-/-* mice.[29] These luminal changes were predictive of IHC exposure, and could be used potentially to highlight atherosclerotic risk.[26] Despite measuring the microbiome and metabolome composition only one week after the initiation of daily IHC intervention, we found similar changes in this study. Moreover, since the study was solely focused on OSA and the impact of IHC, we used the most well-defined model (*Apoe-/-* mice on atherogenic diet) and determined the impact of a single factor (environmental changes in gases) that exacerbates atherosclerosis and whether it impacts the dynamics of the gut microbiome. Since genetics and diet were adequately controlled for in our experimental and control condition, there is no need to disentangle their effects from the observed effects of IHC. Though the circadian impact of genetics and diet have not yet been investigated in context of atherosclerosis, several studies have shown that they do impact circadian dynamics of the gut microbiome.[25–28]

Experiments were performed using best practice guidelines to minimize maternal, founder, and cage effects.[77–79] In addition, all mice were from the same source vivarium, room, and maternal line (to minimize maternal effect), were acclimated in shared holding cages in the vivarium (to control for founder effects), and then were pseudo-randomized into individually-housed cages (to control for cage effect).

IHC imposed microbiome compositional changes that are often observed in dysmetabolic states, including changes in diversity and in abundances of specific bacterial families. IHC-induced reduction of microbial diversity and richness occurred within one week of exposure. This observation suggests that IHC makes the lumen uninhabitable for many commensal microbes, likely by changing environmental micro-niches. Moreover, these changes are not restricted to only the time of exposure to IHC (i.e., light period); there are global shifts in the gut microbiome even during the times of the day when the animal is not being exposed to IHC. Interestingly, the dark period microbiome of the IHC is more similar to the light period microbiome composition of the control mice. Importantly, changes to the microbiome oscillations occur almost immediately, within one week of IHC exposure, confirming that this change results from IHC, rather than the atherosclerotic phenotype, which can take 10 weeks to develop.[25,26,29] Importantly, these changes can create a lasting and profound impact on metabolic health of the host.

Previous studies investigating diurnal cycling of the gut microbiome have used samples collected under 24h[9,11,13] and 48h conditions[10,12,80]. Though in general, it is preferable to use 48h data for circadian studies since it reduces Type I errors, recent advances in bioinformatic tools, such as Metacycle[42], allow for rigorous analysis of 24hr circadian data. This allows investigators to determine circadian cycling from more limited data while still reducing Type I error. Though these tools were created for transcriptional data, as opposed to microbiome data, since our work ultimately replicates and expands upon the results of previous microbiome IHC studies[81] and are consistent with circadian microbiome studies, these potential issues do not significantly impact

interpretation. Nevertheless, a more thorough study of comparing 24h and 48h diurnal microbiome data is warranted to determine if MetaCycle is as robust for this type of data as it is for transcriptional data.

IHC significantly affected the circadian dynamics of the gut microbiome. Verrucomicrobia, Firmicutes and Bacteroidetes - phyla that composed more than 90% of the gut microbiome - were all affected by the IHC condition. The relative abundances of families implicated in atherosclerosis, Ruminococcaceae and Lachnospiraceae[22,82], had circadian oscillations in control mice but lost these oscillations in mice exposed to IHC conditions. In particular, Lachnospiraceae, a bacterial family that has been associated with increase in pro-atherosclerotic TMAO and an increase in a thrombotic phenotype, was significantly elevated in IHC mice during the dark period.[83] IHC also led to the relative reduction of *Akkermansia muciniphila*, the only microbe in phylum Verrucomicrobia found in the gut luminal environment of mammals. *A. muciniphila* is crucial for gut barrier integrity, which helps prevent a pro-inflammatory state by impeding translocation of lumial compounds into the portal system.[84] Replenishing these species in the microbiome of Western diet-fed *Apoe-/-* mice resulted in a decrease in atherosclerotic lesions.[85] Further, IHC induced a gain of oscillation in S24-7, a family in the Bacteroidetes phylum. This family had significantly increased relative abundance levels during the light period. Some members of S24-7 contain a SpeB homolog, a cysteine protease[55], which helps these bacteria avoid detection from the immune system[55,86] and can potentially degrade the protective biofilm present on the surface of the mucosal layer[87]. Together the increase in S24-7 and decrease in *A. muciniphila* suggest a disruption of host mucosal layer and a breakdown of the gut barrier function. Whether these microbiome changes are the sole cause of IHC-induced increase in CVD risks or exacerbators of the dysmetabolic phenotype warrant further investigation.

Along with changes to the gut microbiome, IHC-conditioned metabolome was also altered with increased levels of pro-inflammatory metabolites particularly in bile acids. A comparative

analysis of dark period versus light period fecal samples revealed higher levels of pro-inflammatory and pro-atherosclerotic bile acids, such as DCA and TβMCA, in IHC mice regardless of time of collection. Moreover, atherosclerosis studies that collect fecal pellets during the light period for metabolomics analysis may underestimate their potential role in CVD risk associated with IHC. DCA gains circadian oscillation in IHC mice with the most significant changes from controls occurring during the dark period. Elevated fecal DCA boosts overall systemic inflammation[88–90], which in IHC may be especially problematic in context of a potentially disrupted intestinal barrier. TβMCA, a naturally occurring FXR antagonist[65], was also significantly elevated in IHC, especially during IHC exposure. There is conflicting evidence on the role of FXR in atherosclerosis development and progression, and thus it is possible that these pro-inflammatory changes are protective as opposed to pathogenic. However, deletion of FXR in *Apoe-/-* mice results in lesion exacerbation[67], suggesting that FXR antagonism with excessive TβMCA contributes to worse CVD under the IHC condition in our preclinical model of OSA. In addition, anti-inflammatory and anti-atherosclerotic bile acids (e.g., UDCA) were decreased under IHC conditions.

Overall differences in bile acid levels between IHC and control groups were more pronounced during the dark period. Moreover, our trans-omic analysis (mmvec plus ranked multinomial regression), which divulges relationships between luminal microbial composition and luminal metabolites that are related to our experimental condition, resulted in conditional probabilities with two key findings. First, the abundance of Ruminococcaceae co-occurs with TβMCA. Many members of Ruminococcaceae have 7α-dehydroxylation and 7β-dehydrogenation genes that help it perform bile acid biotransformations.[91, 92] In fact, Ruminococcaceae is positively correlated with fecal DCA levels[93], which it likely helps create with 7α-dehydroxylation. Increased levels of both Ruminococcaceae and TβMCA were associated with IHC-exposure, further highlighting the presence of a pro-inflammatory, pro-atherosclerotic environment under these

conditions. Second, our trans-omic analysis found that the abundance of Coriobacteriaceae co-occurred with CDCA. Increased levels of both Coriobacteriaceae and CDCA were associated with mice under the control condition. FXR agonism with CDCA protects against dyslipidemia and atherosclerosis in *Apoe-/-* mice.[94] Further investigation is required to understand how CDCA and Coriobacteriaceae may be linked. Coriobacteriaceae has been thought to have beneficial metabolic effects[95], including resistance to obesity and liver pathologies[96].

Bile acids are not the only atherosclerosis-related metabolites that were affected by IHC. There were also significant increases in pro-atherolscerotic glycerolipids, including phosphocholines and phosphoethanolamines, under the IHC condition. Phosphocholines are known to be components of LDL ("bad") cholesterol and interact with C-reactive protein in a pro-inflammatory and pro-atherosclerotic manner.[97–99] Increased excretion bilirubin breakdown products, stercobilin and urobilin, under IHC conditions may be an early biomarker of liver dysmetabolism.[62] Interestingly, IHC mice had significantly decreased levels of acylcarnitines during both phases of the day compared to control mice. This is the only pro-atherosclerotic metabolite we measured that was decreased under the IHC condition. Since changes in the microbiome suggest increased gut permeability, decreased levels of acylcarnitines in stool may indicate increased absorption into the serum, where they promote inflammation and are associated with increased risk of myocardial infarction.[100] Or this may indicate that IHC-induced atherosclerosis is not driven by acylcarnitines. Future studies will need to determine the relationship between fecal and serum acycarinities to help determine if this is the case.

One of the limitations of this study is that, due to methodological constraints, we were not able to assess the levels and diurnal oscillations of known pro-atherosclerotic and anti-inflammatory small molecules, such as TMA (which is then converted to TMA N-Oxide [TMAO] in the liver) and SCFAs, respectively. Cyclical fluctuations in SCFAs have been documented in control mice, as well as those with dysmetabolic phenotypes, and SCFAs can affect hepatic

peripheral circadian rhythms.[8] Given the extent of the IHC-induced circadian changes in luminal environment, there is little doubt that these metabolites likely fluctuate with OSA as well. Taken together, changes in the gut microbiome oscillations and the different types of metabolites have complex pathological implications that require further investigation.

Since IHC had a remarkable influence on the oscillations of the microbiome and metabolome in the luminal gut, and since IHC is composed of the synchronous changes in O_2 and CO_2 in opposite directions, our results in this work would indicate that each specific gas change may have an influence on the cyclicity of the gut microbiome. On the other hand, perturbed luminal dynamics suggests altered nutrient availability. IHC may disrupt normal feeding patterns, likely due to increased stress or lack of sleep in these mice. Normal feeding patterns are essential for the maintenance of the peripheral circadian clock. Time-restricted feeding (TRF), where a normal feeding pattern is maintained by consolidating access food only to the active time period, enforces central and peripheral circadian clock synchrony and prevents dysmetabolic phenotypes in a number of nutritional and non-nutritional challenges to metabolic homeostasis.[7,101–103] Whether TRF can protect against IHC-induced CVD risks is yet to be determined. Moreover, understanding how bile acid modifications performed by the gut microbiome modulates host metabolic mechanisms will provide valuable insight into the pathophysiology of IHC-induced atherosclerosis. Although there is sufficient evidence that bile acid signaling plays an important role in atherosclerosis, a better understanding of bacterial bile acid biotransformations and their contribution to IHC-induced pathogenesis will yield novel therapeutic targets. Since the changes in bile acids found in this study are centered around FXR receptor expression, future experiments should further investigate the role of bile acids receptors (i.e., FXR, TGR5) in mediating the effects of IHC on host atherosclerosis. Importantly, our study clearly demonstrates the importance of considering time of sample collection in the experimental

design, as many of the differences observed in our study would have gone unnoticed if samples were collected at a single time point or only during the light period.

2.5 Acknowledgments

We would like to express our gratitude to Michael Meehan for his contribution to metabolomics sample processing.

Chapter 2, in full, is a reprint of the material as it appears in, “Intermittent Hypoxia and Hypercapnia Alter Diurnal Rhythms of Luminal Gut Microbiome and Metabolome.” Celeste Allaband, Amulya Lingaraju, Cameron Martino, Baylee Russell, Anupriya Tripathi, Orit Poulsen, Ana Carolina Dantas Machado, Dan Zhou, Jin Xue, Emmanuel Elijah, Atul Malhotra, Pieter C Dorrestein, Rob Knight, Gabriel G Haddad, and Amir Zarrinpar. mSystems, e0011621. The dissertation author was the primary investigator and first author of this paper.

2.6 Sources of Funding

CA is supported by T32 OD017863. ACDM is funded by R01 HL148801-02S1. GGH is supported by NIH R01 HL146530 and R21 NS111270. AZ is supported by K08 DK102902, R03 DK114536, R21 MH117780, R01 HL148801. AM is funded by R01 HL085188; T32 HL134632; R01 HL 142114; R01 AG063925; R01 HL148436; K24 HL132105

All authors receive institutional support from P30 DK120515, P30 DK063491, and UL1 TR001442. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

2.7 Disclosures

AM reports income related to medical education from Livanova, Equillium and Corvus. ResMed provided a philanthropic donation to UC San Diego. AZ is a founder and equity-holder in Tortuga Biosciences. PD is advisor to Sirenas, Galileo and Cybele and is an advisor and co-founder of Enveda and Ometa that has been approved by UC-San Diego.

2.8 References

1. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després J-P, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Liu S, Mackey RH, Matchar DB, McGuire DK, Mohler ER, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Nichol G, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner MB, American Heart Association Statistics Committee and Stroke Statistics Subcommittee. 2015. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation 131:e29-322.

2. Guilleminault C, Connolly S, Winkle R, Melvin K, Tilkian A. 1984. Cyclical variation of the heart rate in sleep apnoea syndrome. Mechanisms, and usefulness of 24 h electrocardiography as a screening technique. Lancet 1:126–131.

3. Reutrakul S, Van Cauter E. 2014. Interactions between sleep, circadian function, and glucose metabolism: implications for risk and severity of diabetes. Ann N Y Acad Sci 1311:151–173.

4. Lemmer B, Scholtze J, Schmitt J. 2016. Circadian rhythms in blood pressure, heart rate, hormones, and on polysomnographic parameters in severe obstructive sleep apnea syndrome patients: effect of continuous positive airway pressure. Blood Press Monit 21:136–143.

5. Chaix A, Zarrinpar A, Panda S. 2016. The circadian coordination of cell biology. J Cell Biol 215:15–25.

6. Stenvers DJ, Scheer FAJL, Schrauwen P, la Fleur SE, Kalsbeek A. 2019. Circadian clocks and insulin resistance. Nat Rev Endocrinol 15:75–89.

7. Zarrinpar A, Chaix A, Panda S. 2016. Daily Eating Patterns and Their Impact on Health and Disease. Trends Endocrinol Metab 27:69–83.

8. Frazier K, Chang EB. 2020. Intersection of the Gut Microbiome and Circadian Rhythms in Metabolism. Trends Endocrinol Metab 31:25–36.

9. Leone V, Gibbons SM, Martinez K, Hutchison AL, Huang EY, Cham CM, Pierre JF, Heneghan AF, Nadimpalli A, Hubert N, Zale E, Wang Y, Huang Y, Theriault B, Dinner AR, Musch MW, Kudsk KA, Prendergast BJ, Gilbert JA, Chang EB. 2015. Effects of diurnal

variation of gut microbes and high-fat feeding on host circadian clock function and metabolism. Cell Host Microbe 17:681–689.

10. Thaiss CA, Zeevi D, Levy M, Zilberman-Schapira G, Suez J, Tengeler AC, Abramson L, Katz MN, Korem T, Zmora N, Kuperman Y, Biton I, Gilad S, Harmelin A, Shapiro H, Halpern Z, Segal E, Elinav E. 2014. Transkingdom control of microbiota diurnal oscillations promotes metabolic homeostasis. Cell 159:514–529.

11. Zarrinpar A, Chaix A, Yooseph S, Panda S. 2014. Diet and Feeding Pattern Affect the Diurnal Dynamics of the Gut Microbiome. Cell Metabolism 20:1006–1017.

12. Thaiss CA, Levy M, Korem T, Dohnalová L, Shapiro H, Jaitin DA, David E, Winter DR, Gury-BenAri M, Tatirovsky E, Tuganbaev T, Federici S, Zmora N, Zeevi D, Dori-Bachash M, Pevsner-Fischer M, Kartvelishvily E, Brandis A, Harmelin A, Shibolet O, Halpern Z, Honda K, Amit I, Segal E, Elinav E. 2016. Microbiota Diurnal Rhythmicity Programs Host Transcriptome Oscillations. Cell 167:1495-1510.e12.

13. Liang X, Bushman FD, FitzGerald GA. 2015. Rhythmicity of the intestinal microbiota is regulated by gender and the host circadian clock. Proc Natl Acad Sci USA 112:10479–10484.

14. Mukherji A, Bailey SM, Staels B, Baumert TF. 2019. The circadian clock and liver function in health and disease. J Hepatol 71:200–211.

15. Wang Y, Kuang Z, Yu X, Ruhn KA, Kubo M, Hooper LV. 2017. The intestinal microbiota regulates body composition through NFIL3 and the circadian clock. Science 357:912–916.

16. Segers A, Desmet L, Thijs T, Verbeke K, Tack J, Depoortere I. 2019. The circadian clock regulates the diurnal levels of microbial short-chain fatty acids and their rhythmic effects on colon contractility in mice. Acta Physiol (Oxf) 225:e13193.

17. Govindarajan K, MacSharry J, Casey PG, Shanahan F, Joyce SA, Gahan CGM. 2016. Unconjugated Bile Acids Influence Expression of Circadian Genes: A Potential Mechanism for Microbe-Host Crosstalk. PLoS ONE 11:e0167319.

18. Hussain MM, Pan X. 2009. Clock genes, intestinal transport and plasma lipid homeostasis. Trends Endocrinol Metab 20:177–185.

19. Méndez-Ferrer S, Lucas D, Battista M, Frenette PS. 2008. Haematopoietic stem cell release is regulated by circadian oscillations. Nature 452:442–447.

20. Panza JA, Epstein SE, Quyyumi AA. 1991. Circadian variation in vascular tone and its relation to alpha-sympathetic vasoconstrictor activity. N Engl J Med 325:986–990.

21. Muller JE, Tofler GH, Willich SN, Stone PH. 1987. Circadian variation of cardiovascular disease and sympathetic activity. J Cardiovasc Pharmacol 10 Suppl 2:S104-109; discussion S110-111.

22. Liu B, Zhang Y, Wang R, An Y, Gao W, Bai L, Li Y, Zhao S, Fan J, Liu E. 2018. Western diet feeding influences gut microbiota profiles in apoE knockout mice. Lipids Health Dis 17:159.

23. Chelakkot C, Choi Y, Kim D-K, Park HT, Ghim J, Kwon Y, Jeon J, Kim M-S, Jee Y-K, Gho YS, Park H-S, Kim Y-K, Ryu SH. 2018. Akkermansia muciniphila-derived extracellular vesicles influence gut permeability through the regulation of tight junctions. Exp Mol Med 50:e450.

24. Summa KC, Voigt RM, Forsyth CB, Shaikh M, Cavanaugh K, Tang Y, Vitaterna MH, Song S, Turek FW, Keshavarzian A. 2013. Disruption of the Circadian Clock in Mice Increases Intestinal Permeability and Promotes Alcohol-Induced Hepatic Pathology and Inflammation. PLoS ONE 8:e67102.

25. Tripathi A, Melnik AV, Xue J, Poulsen O, Meehan MJ, Humphrey G, Jiang L, Ackermann G, McDonald D, Zhou D, Knight R, Dorrestein PC, Haddad GG. 2018. Intermittent Hypoxia and Hypercapnia, a Hallmark of Obstructive Sleep Apnea, Alters the Gut Microbiome and Metabolome. mSystems 3.

26. Tripathi A, Xu ZZ, Xue J, Poulsen O, Gonzalez A, Humphrey G, Meehan MJ, Melnik AV, Ackermann G, Zhou D, Malhotra A, Haddad GG, Dorrestein PC, Knight R. 2019. Intermittent Hypoxia and Hypercapnia Reproducibly Change the Gut Microbiome and Metabolome across Rodent Model Systems. mSystems 4.

27. Ryan PM, London LEE, Bjorndahl TC, Mandal R, Murphy K, Fitzgerald GF, Shanahan F, Ross RP, Wishart DS, Caplice NM, Stanton C. 2017. Microbiome and metabolome modifying effects of several cardiovascular disease interventions in apo-E-/- mice. Microbiome 5:30.

28. Kiouptsi K, Jäckel S, Pontarollo G, Grill A, Kuijpers MJE, Wilms E, Weber C, Sommer F, Nagy M, Neideck C, Jansen Y, Ascher S, Formes H, Karwot C, Bayer F, Kollar B, Subramaniam S, Molitor M, Wenzel P, Rosenstiel P, Todorov H, Gerber S, Walter U, Jurk K, Heemskerk JWM, van der Vorst EPC, Döring Y, Reinhardt C. 2019. The Microbiota Promotes Arterial Thrombosis in Low-Density Lipoprotein Receptor-Deficient Mice. mBio 10.

29. Xue J, Zhou D, Poulsen O, Imamura T, Hsiao Y-H, Smith TH, Malhotra A, Dorrestein P, Knight R, Haddad GG. 2017. Intermittent Hypoxia and Hypercapnia Accelerate Atherosclerosis, Partially via Trimethylamine-Oxide. Am J Respir Cell Mol Biol 57:581–588.

30. Zhang Y-KJ, Guo GL, Klaassen CD. 2011. Diurnal variations of mouse plasma and hepatic bile acid concentrations as well as expression of biosynthetic enzymes and transporters. PLoS ONE 6:e16683.

31. Pan X, Hussain MM. 2007. Diurnal regulation of microsomal triglyceride transfer protein and plasma lipid levels. J Biol Chem 282:24707–24719.

32. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6:1621–1624.

33. Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, Gilbert JA, Jansson JK, Caporaso JG, Fuhrman JA, Apprill A, Knight R. 2016. Improved Bacterial

16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. mSystems 1.

34. Gonzalez A, Navas-Molina JA, Kosciolek T, McDonald D, Vázquez-Baeza Y, Ackermann G, DeReus J, Janssen S, Swafford AD, Orchanian SB, Sanders JG, Shorenstein J, Holste H, Petrus S, Robbins-Pianka A, Brislawn CJ, Wang M, Rideout JR, Bolyen E, Dillon M, Caporaso JG, Dorrestein PC, Knight R. 2018. Qiita: rapid, web-enabled microbiome meta-analysis. Nat Methods 15:796–798.

35. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. 2017. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems 2:e00191-16.

36. McDonald D, Clemente JC, Kuczynski J, Rideout JR, Stombaugh J, Wendel D, Wilke A, Huse S, Hufnagle J, Meyer F, Knight R, Caporaso JG. 2012. The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. GigaSci 1:7.

37. Warnow T. 2013. SATe-Enabled Phylogenetic Placement, p. 1–4. In Nelson, KE (ed.), Encyclopedia of Metagenomics. Springer New York, New York, NY.

38. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–857.

39. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and environmental microbiology 72:5069–72.

40. Chang Q, Luan Y, Sun F. 2011. Variance adjusted weighted UniFrac: a powerful beta diversity measure for comparing communities based on phylogeny. BMC Bioinformatics 12:118.

41. Anderson MJ. 2017. Permutational Multivariate Analysis of Variance (PERMANOVA), p. 1–15. In Balakrishnan, N, Colton, T, Everitt, B, Piegorsch, W, Ruggeri, F, Teugels, JL

(eds.), Wiley StatsRef: Statistics Reference Online. John Wiley & Sons, Ltd, Chichester, UK.

42. Wu G, Anafi RC, Hughes ME, Kornacker K, Hogenesch JB. 2016. MetaCycle: an integrated R package to evaluate periodicity in large scale data. Bioinformatics 32:3351–3353.

43. Singer JM, Hughey JJ. 2019. LimoRhyde: A Flexible Approach for Differential Analysis of Rhythmic Transcriptome Data. J Biol Rhythms 34:5–18.

44. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, Hoff K, Kessner D, Tasman N, Shulman N, Frewen B, Baker TA, Brusniak M-Y, Paulse C, Creasy D, Flashner L, Kani K, Moulding C, Seymour SL, Nuwaysir LM, Lefebvre B, Kuhlmann F, Roark J, Rainer P, Detlev S, Hemenway T, Huhmer A, Langridge J, Connolly B, Chadick T, Holly K, Eckels J, Deutsch EW, Moritz RL, Katz JE, Agus DB, MacCoss M, Tabb DL, Mallick P. 2012. A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol 30:918–920.

45. Pluskal T, Castillo S, Villar-Briones A, Orešič M. 2010. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics 11:395.

46. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu W-T, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu C-C, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw C-C, Yang Y-L, Humpf H-U, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, P CAB, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard P-M, Phapale P, Nothias L-F, Alexandrov T, Litaudon M, Wolfender J-L, Kyle JE, Metz TO, Peryea T, Nguyen D-T, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BO, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol 34:828–837.

47. Scheubert K, Hufsky F, Petras D, Wang M, Nothias L-F, Dührkop K, Bandeira N, Dorrestein PC, Böcker S. 2017. Significance estimation for large scale metabolomics annotations by spectral matching. Nat Commun 8:1494.

48. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW-M, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reily MD, Thaden JJ, Viant MR. 2007. Proposed

minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics 3:211–221.

49. MSI Board Members: 2007. The Metabolomics Standards Initiative. Nat Biotechnol 25:846–848.

50. Lance GN. 1967. A general theory of classificatory sorting strategies: II. Clustering systems. The Computer Journal 10:271–277.

51. Morton JT, Marotz C, Washburne A, Silverman J, Zaramela LS, Edlund A, Zengler K, Knight R. 2019. Establishing microbial composition measurement standards with reference frames. Nat Commun 10:2719.

52. Fedarko MW, Martino C, Morton JT, González A, Rahman G, Marotz CA, Minich JJ, Allen EE, Knight R. 2020. Visualizing 'omic feature rankings and log-ratios using Qurro. NAR Genom Bioinform 2:lqaa023.

53. Morton JT, Aksenov AA, Nothias LF, Foulds JR, Quinn RA, Badri MH, Swenson TL, Van Goethem MW, Northen TR, Vazquez-Baeza Y, Wang M, Bokulich NA, Watters A, Song SJ, Bonneau R, Dorrestein PC, Knight R. 2019. Learning representations of microbe-metabolite interactions. Nat Methods 16:1306–1314.

54. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto J-M, Kennedy S, Leonard P, Li J, Burgdorf K, Grarup N, Jørgensen T, Brandslund I, Nielsen HB, Juncker AS, Bertalan M, Levenez F, Pons N, Rasmussen S, Sunagawa S, Tap J, Tims S, Zoetendal EG, Brunak S, Clément K, Doré J, Kleerebezem M, Kristiansen K, Renault P, Sicheritz-Ponten T, de Vos WM, Zucker J-D, Raes J, Hansen T, MetaHIT consortium, Bork P, Wang J, Ehrlich SD, Pedersen O. 2013. Richness of human gut microbiome correlates with metabolic markers. Nature 500:541–546.

55. Ormerod KL, Wood DLA, Lachner N, Gellatly SL, Daly JN, Parsons JD, Dal'Molin CGO, Palfreyman RW, Nielsen LK, Cooper MA, Morrison M, Hansbro PM, Hugenholtz P. 2016. Genomic characterization of the uncultured Bacteroidales family S24-7 inhabiting the guts of homeothermic animals. Microbiome 4:36.

56. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, DuGar B, Feldstein AE, Britt EB, Fu X, Chung Y-M, Wu Y, Schauer P, Smith JD, Allayee H, Tang WHW, DiDonato JA, Lusis AJ, Hazen SL. 2011. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature 472:57–63.

57. Que X, Hung M-Y, Yeang C, Gonen A, Prohaska TA, Sun X, Diehl C, Määttä A, Gaddis DE, Bowden K, Pattison J, MacDonald JG, Ylä-Herttuala S, Mellon PL, Hedrick CC, Ley K, Miller YI, Glass CK, Peterson KL, Binder CJ, Tsimikas S, Witztum JL. 2018. Oxidized phospholipids are proinflammatory and proatherogenic in hypercholesterolaemic mice. Nature 558:301–306.

58. Chaurasia B, Summers SA. 2015. Ceramides - Lipotoxic Inducers of Metabolic Disorders. Trends Endocrinol Metab 26:538–550.

59. Vorkas PA, Shalhoub J, Isaac G, Want EJ, Nicholson JK, Holmes E, Davies AH. 2015. Metabolic phenotyping of atherosclerotic plaques reveals latent associations between free cholesterol and ceramide metabolism in atherogenesis. J Proteome Res 14:1389–1399.

60. Blair HC, Sepulveda J, Papachristou DJ. 2016. Nature and nurture in atherosclerosis: The roles of acylcarnitine and cell membrane-fatty acid intermediates. Vascul Pharmacol 78:17–23.

61. Huynh K, Barlow CK, Jayawardana KS, Weir JM, Mellett NA, Cinel M, Magliano DJ, Shaw JE, Drew BG, Meikle PJ. 2019. High-Throughput Plasma Lipidomics: Detailed Mapping of the Associations with Cardiometabolic Risk Factors. Cell Chem Biol 26:71-84.e4.

62. Hamoud A-R, Weaver L, Stec DE, Hinds TD. 2018. Bilirubin in the Liver-Gut Signaling Axis. Trends Endocrinol Metab 29:140–150.

63. Wilson A, McLean C, Kim RB. 2016. Trimethylamine-N-oxide: a link between the gut microbiome, bile acid metabolism, and atherosclerosis. Current Opinion in Lipidology 27:148–154.

64. Traub RJ, Tang B, Ji Y, Pandya S, Yfantis H, Sun Y. 2008. A Rat Model of Chronic Postinflammatory Visceral Pain Induced by Deoxycholic Acid. Gastroenterology 135:2075–2083.

65. Sayin SI, Wahlström A, Felin J, Jäntti S, Marschall H-U, Bamberg K, Angelin B, Hyötyläinen T, Orešič M, Bäckhed F. 2013. Gut Microbiota Regulates Bile Acid Metabolism by Reducing the Levels of Tauro-beta-muricholic Acid, a Naturally Occurring FXR Antagonist. Cell Metabolism 17:225–235.

66. Hageman J, Herrema H, Groen AK, Kuipers F. 2010. A Role of the Bile Salt Receptor FXR in Atherosclerosis. ATVB 30:1519–1528.

67. Hanniman EA, Lambert G, McCarthy TC, Sinal CJ. 2005. Loss of functional farnesoid X receptor increases atherosclerotic lesions in apolipoprotein E-deficient mice. J Lipid Res 46:2595–2604.

68. Sehayek E, Ono JG, Duncan EM, Batta AK, Salen G, Shefer S, Neguyen LB, Yang K, Lipkin M, Breslow JL. 2001. Hyodeoxycholic acid efficiently suppresses atherosclerosis formation and plasma cholesterol levels in mice. J Lipid Res 42:1250–1256.

69. Shih DM, Shaposhnik Z, Meng Y, Rosales M, Wang X, Wu J, Ratiner B, Zadini F, Zadini G, Lusis AJ. 2013. Hyodeoxycholic acid improves HDL function and inhibits atherosclerotic lesion formation in LDLR-knockout mice. FASEB J 27:3805–3817.

70. Bellentani S. 2005. Immunomodulating and anti-apoptotic action of ursodeoxycholic acid: where are we and where should we go?: European Journal of Gastroenterology & Hepatology 17:137–140.

71. Vang S, Longley K, Steer CJ, Low WC. 2014. The Unexpected Uses of Urso- and Tauroursodeoxycholic Acid in the Treatment of Non-liver Diseases. Glob Adv Health Med 3:58–69.

72. Goñi FM, Requero MA, Alonso A. 1996. Palmitoylcarnitine, a surface-active metabolite. FEBS Letters 390:1–5.

73. Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT, Britt EB, Fu X, Wu Y, Li L, Smith JD, DiDonato JA, Chen J, Li H, Wu GD, Lewis JD, Warrier M, Brown JM, Krauss RM, Tang WHW, Bushman FD, Lusis AJ, Hazen SL. 2013. Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis. Nat Med 19:576–585.

74. Quehenberger O, Armando AM, Brown AH, Milne SB, Myers DS, Merrill AH, Bandyopadhyay S, Jones KN, Kelly S, Shaner RL, Sullards CM, Wang E, Murphy RC, Barkley RM, Leiker TJ, Raetz CRH, Guan Z, Laird GM, Six DA, Russell DW, McDonald JG, Subramaniam S, Fahy E, Dennis EA. 2010. Lipidomics reveals a remarkable diversity of lipids in human plasma. J Lipid Res 51:3299–3305.

75. Bisanz JE, Upadhyay V, Turnbaugh JA, Ly K, Turnbaugh PJ. 2019. Meta-Analysis Reveals Reproducible Gut Microbiome Alterations in Response to a High-Fat Diet. Cell Host & Microbe 26:265-272.e4.

76. Reitmeier S, Kiessling S, Clavel T, List M, Almeida EL, Ghosh TS, Neuhaus K, Grallert H, Linseisen J, Skurk T, Brandl B, Breuninger TA, Troll M, Rathmann W, Linkohr B, Hauner H, Laudes M, Franke A, Le Roy CI, Bell JT, Spector T, Baumbach J, O'Toole PW, Peters A, Haller D. 2020. Arrhythmic Gut Microbiome Signatures Predict Risk of Type 2 Diabetes. Cell Host Microbe 28:258-272.e6.

77. Allaband C, McDonald D, Vázquez-Baeza Y, Minich JJ, Tripathi A, Brenner DA, Loomba R, Smarr L, Sandborn WJ, Schnabl B, Dorrestein P, Zarrinpar A, Knight R. 2019. Microbiome 101: Studying, Analyzing, and Interpreting Gut Microbiome Data for Clinicians. Clinical Gastroenterology and Hepatology 17:218–230.

78. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciolek T, McCall L-I, McDonald D, Melnik AV, Morton JT, Navas J, Quinn RA, Sanders JG, Swafford AD, Thompson LR, Tripathi A, Xu ZZ, Zaneveld JR, Zhu Q, Caporaso JG, Dorrestein PC. 2018. Best practices for analysing microbiomes. Nature Reviews Microbiology 16:410–422.

79. Goodrich JK, Di Rienzi SC, Poole AC, Koren O, Walters WA, Caporaso JG, Knight R, Ley RE. 2014. Conducting a Microbiome Study. Cell 158:250–262.

80. Tuganbaev T, Mor U, Bashiardes S, Liwinski T, Nobs SP, Leshem A, Dori-Bachash M, Thaiss CA, Pinker EY, Ratiner K, Adlung L, Federici S, Kleimeyer C, Moresi C, Yamada T, Cohen Y, Zhang X, Massalha H, Massasa E, Kuperman Y, Koni PA, Harmelin A, Gao N, Itzkovitz S, Honda K, Shapiro H, Elinav E. 2020. Diet Diurnally Regulates Small Intestinal Microbiome-Epithelial-Immune Homeostasis and Enteritis. Cell 182:1441-1459.e21.

81. Xue J, Allaband C, Zhou D, Poulsen O, Martino C, Jiang L, Tripathi A, Elijah E, Dorrestein PC, Knight R, Zarrinpar A, Haddad GG. 2021. Influence of Intermittent Hypoxia/Hypercapnia on Atherosclerosis, Gut Microbiome, and Metabolome. Front Physiol 12:663950.

82. Kurilshikov A, van den Munckhof ICL, Chen L, Bonder MJ, Schraa K, Rutten JHW, Riksen NP, de Graaf J, Oosting M, Sanna S, Joosten LAB, van der Graaf M, Brand T, Koonen DPY, van Faassen M, LifeLines DEEP Cohort Study, BBMRI Metabolomics Consortium, Slagboom PE, Xavier RJ, Kuipers F, Hofker MH, Wijmenga C, Netea MG, Zhernakova A, Fu J. 2019. Gut Microbial Associations to Plasma Metabolites Linked to Cardiovascular Phenotypes and Risk. Circ Res 124:1808–1820.

83. Zhu W, Gregory JC, Org E, Buffa JA, Gupta N, Wang Z, Li L, Fu X, Wu Y, Mehrabian M, Sartor RB, McIntyre TM, Silverstein RL, Tang WHW, DiDonato JA, Brown JM, Lusis AJ, Hazen SL. 2016. Gut Microbial Metabolite TMAO Enhances Platelet Hyperreactivity and Thrombosis Risk. Cell 165:111–124.

84. van der Lugt B, van Beek AA, Aalvink S, Meijer B, Sovran B, Vermeij WP, Brandt RMC, de Vos WM, Savelkoul HFJ, Steegenga WT, Belzer C. 2019. Akkermansia muciniphila ameliorates the age-related decline in colonic mucus thickness and attenuates immune activation in accelerated aging Ercc1 -/Δ7 mice. Immun Ageing 16:6.

85. Li J, Lin S, Vanhoutte PM, Woo CW, Xu A. 2016. Akkermansia Muciniphila Protects Against Atherosclerosis by Preventing Metabolic Endotoxemia-Induced Inflammation in Apoe-/- Mice. Circulation 133:2434–2446.

86. Nelson DC, Garbe J, Collin M. 2011. Cysteine proteinase SpeB from Streptococcus pyogenes - a potent modifier of immunologically important host and bacterial proteins. Biol Chem 392:1077–1088.

87. Roberts AL, Holder RC, Reid SD. 2010. Allelic replacement of the streptococcal cysteine protease SpeB in a Δsrv mutant background restores biofilm formation. BMC Res Notes 3:281.

88. Sheng L, Jena PK, Hu Y, Liu H-X, Nagar N, Kalanetra KM, French SW, French SW, Mills DA, Wan Y-JY. 2017. Hepatic inflammation caused by dysregulated bile acid synthesis is reversible by butyrate supplementation. J Pathol 243:431–441.

89. Zhao S, Gong Z, Zhou J, Tian C, Gao Y, Xu C, Chen Y, Cai W, Wu J. 2016. Deoxycholic Acid Triggers NLRP3 Inflammasome Activation and Aggravates DSS-Induced Colitis in Mice. Front Immunol 7:536.

90. Xu M, Cen M, Shen Y, Zhu Y, Cheng F, Tang L, Hu W, Dai N. 2020. Deoxycholic Acid-Induced Gut Dysbiosis Disrupts Bile Acid Enterohepatic Circulation and Promotes Intestinal Inflammation. Dig Dis Sci https://doi.org/10.1007/s10620-020-06208-3.

91. Vital M, Rud T, Rath S, Pieper DH, Schlüter D. 2019. Diversity of Bacteria Exhibiting Bile Acid-inducible 7α-dehydroxylation Genes in the Human Gut. Computational and Structural Biotechnology Journal 17:1016–1019.

92. Ikegami T, Honda A. 2018. Reciprocal interactions between bile acids and gut microbiota in human liver diseases. Hepatol Res 48:15–27.

93. Kakiyama G, Pandak WM, Gillevet PM, Hylemon PB, Heuman DM, Daita K, Takei H, Muto A, Nittono H, Ridlon JM, White MB, Noble NA, Monteith P, Fuchs M, Thacker LR, Sikaroodi

M, Bajaj JS. 2013. Modulation of the fecal bile acid profile by gut microbiota in cirrhosis. J Hepatol 58:949–955.

94. Mencarelli A, Cipriani S, Renga B, Francisci D, Palladino G, Distrutti E, Baldelli F, Fiorucci S. 2010. The bile acid sensor FXR protects against dyslipidemia and aortic plaques development induced by the HIV protease inhibitor ritonavir in mice. PLoS ONE 5:e13238.

95. Liu H, Zhang H, Wang X, Yu X, Hu C, Zhang X. 2018. The family Coriobacteriaceae is a potential contributor to the beneficial effects of Roux-en-Y gastric bypass on type 2 diabetes. Surgery for Obesity and Related Diseases 14:584–593.

96. Clavel T, Desmarchelier C, Haller D, Gérard P, Rohn S, Lepage P, Daniel H. 2014. Intestinal microbiota in metabolic diseases: From bacterial community structure and functions to species of pathophysiological relevance. Gut Microbes 5:544–551.

97. Witztum JL, Berliner JA. 1998. Oxidized phospholipids and isoprostanes in atherosclerosis: Current Opinion in Lipidology 9:441–448.

98. Berliner J. 2001. Evidence for a Role of Phospholipid Oxidation Products in Atherogenesis. Trends in Cardiovascular Medicine 11:142–147.

99. Chang M-K, Binder CJ, Torzewski M, Witztum JL. 2002. C-reactive protein binds to both oxidized LDL and apoptotic cells through recognition of a common ligand: Phosphorylcholine of oxidized phospholipids. Proceedings of the National Academy of Sciences 99:13043–13048.

100. Strand E, Pedersen ER, Svingen GFT, Olsen T, Bjørndal B, Karlsson T, Dierkes J, Njølstad PR, Mellgren G, Tell GS, Berge RK, Svardal A, Nygård O. 2017. Serum Acylcarnitines and Risk of Cardiovascular Death and Acute Myocardial Infarction in Patients With Stable Angina Pectoris. J Am Heart Assoc 6.

101. Chaix A, Zarrinpar A, Miu P, Panda S. 2014. Time-restricted feeding is a preventative and therapeutic intervention against diverse nutritional challenges. Cell Metab 20:991–1005.

102. Wilkinson MJ, Manoogian ENC, Zadourian A, Lo H, Fakhouri S, Shoghi A, Wang X, Fleischer JG, Navlakha S, Panda S, Taub PR. 2020. Ten-Hour Time-Restricted Eating Reduces Weight, Blood Pressure, and Atherogenic Lipids in Patients with Metabolic Syndrome. Cell Metab 31:92-104.e5.

103. Saran AR, Dave S, Zarrinpar A. 2020. Circadian Rhythms in the Pathogenesis and Treatment of Fatty Liver Disease. Gastroenterology 158:1948-1966.e1.

Chapter 3. Intermittent Hypoxia and Hypercapnia Alter Gut Microbiome and Metabolome to Promote Atherosclerosis

Abstract

Obstructive sleep apnea (OSA), a common sleep disorder characterized by intermittent hypoxia and hypercapnia (IHC), increases atherosclerosis risk. However, the contribution of intermittent hypoxia (IH) or intermittent hypercapnia (IC) in promoting atherosclerosis remains unclear. Since gut microbiota and metabolites have been implicated in atherosclerosis, we examined whether IH or IC alters the microbiome and metabolome to induce a pro-atherosclerotic state. Apolipoprotein E deficient mice (ApoE-/-) mice, treated with IH or IC on a high-fat diet (HFD) for 10 weeks, were compared to Air controls. Atherosclerotic lesions were examined, gut microbiome was profiled using 16S rRNA gene amplicon sequencing and metabolome was assessed by untargeted mass spectrometry. In the aorta, IC-induced atherosclerosis was significantly greater than IH and Air controls (aorta, IC 11.1±0.7% vs IH 7.6±0.4%, $p<0.05$ and vs Air 8.1±0.8%, $p<0.05$). In the pulmonary artery (PA), however, IH, IC and Air were significantly different from each other in atherosclerotic formation with the largest lesion observed under IH (PA, IH 40.9±2.0% vs IC 20.1±2.6% vs Air 12.2±1.5%, $p<0.05$). The most differentially abundant microbial families ($p<0.001$) were Peptostreptococcaceae, Ruminococcaceae and Erysipelotrichaceae. The most differentially abundant metabolites ($p<0.001$) were tauro-β-muricholic acid, ursodeoxycholic acid, and lysophosphoethanolamine(18:0). We conclude that a) IH and IC affect atherosclerosis progression differently in distinct vascular sites and b) the changes in the microbiome and metabolome promote a pro-inflammatory and pro-atherosclerotic luminal gut environment that is more evident in IH than IC.

3.1 Introduction

In the third part of this thesis, we present data from a longitudinal study of OSA in the same atherogenic mouse model as part two. We broke OSA conditions into two distinct parts, intermittent hypoxia (IH) and intermittent hypercapnia (IC). While holding time of sample collection constant, we examined 16S microbiome and untargeted liquid chromatography tandem mass spectrometry metabolome trends under both conditions over time until atherosclerotic lesions developed in the heart. Histopathological staining revealed that IC appeared to cause more lesions in the aorta than IH, but IH caused more lesions in the pulmonary arteries than IC. These findings were supported by both the microbiome and metabolome data, where both conditions demonstrated several significant differences from controls but in differing patterns. Together, these results indicate that the timed environmental exposure to OSA conditions seems to promote atherosclerotic luminal gut environment that is more prominent in IH than IC.

Obstructive sleep apnea (OSA) is a common disorder characterized by repetitive episodes of complete or partial upper airway obstruction during sleep. These apneic episodes lead to intermittent hypoxia and hypercapnia (IHC), wide intrathoracic pressure swings, as well as sleep fragmentation. OSA affects approximately 9-38% of the general adult population with 13-33% in men and 6-19% in women (Senaratna et al., 2017). Advanced age, male gender, and higher body-mass index increase OSA prevalence (Senaratna et al., 2017). OSA is independently associated with elevated risk of myocardial infarction, stroke, and cardiovascular mortality, mainly through the promotion of severe atherosclerosis. The pathophysiological mechanisms underlying OSA associated atherosclerotic risk are not completely understood.

Chronic intermittent hypoxia (IH), generated during recurrent apneic episodes, is a major factor linking OSA to cardiovascular diseases including atherosclerosis (Drager et al., 2011). Intermittent hypercapnia (IC) also occurs in OSA but is usually not evaluated in most OSA-based translational studies. IC can potentially affect the formation of atheroma (Xue et al., 2017) as well.

In the present investigation, we examined the lesion formation under these different study conditions (i.e., IH, IC) to better define the role of each component in atherosclerosis.

Considerable evidence indicates that human gut microbiota contributes to cardiovascular diseases including atherosclerosis as well as metabolic disorders, e.g. obesity and type 2 diabetes, both of which are atherogenic. Gut microbiota likely affects atherosclerosis through several mechanisms: (1) bacterial infection activates the immune system and triggers a harmful inflammatory response that aggravates plaque progression and rupture, (2) cholesterol and lipid metabolism altered by gut microbiota affect the development of atherosclerosis, and (3) microbial metabolites have either beneficial or deleterious effects on atherosclerosis (Jonsson and Backhed, 2017; Komaroff, 2018). Our recent studies indicate that the latter of these three likely plays an important role in OSA-induced atherosclerosis (Xue et al., 2017; Tripathi et al., 2018). Although trimethyalamine-N-oxide is perhaps the most well-known example of an atherogenic bacterial metabolite, bile acids and phosphocholines could be intermediates or involved in atherogenesis. For example, tauro-β-muricholic acid (TβMCA) is an FXR antagonist (Sayin et al., 2013) and can contribute to atherosclerosis (Hanniman et al., 2005; Ding et al., 2018). Ursodeoxycholic acid (UDCA) has been demonstrated to have anti-inflammatory effects, which could alleviate the development of atherosclerosis (Ko et al., 2017). Lysophospholipids have also been implicated deleterious roles in atherosclerosis (Matsumoto et al., 2007; Li et al., 2016b). Interestingly, gut microbiota composition is altered in different mouse models of OSA (Moreno-Indias et al., 2015; Xue et al., 2017; Tripathi et al., 2018). In this work, we explored the changes in gut microbiota population and metabolites induced by IH and IC separately. The knowledge obtained will help us dissect out the individual impact of IH and IC on gut microbiota and metabolites as well as on atherosclerosis formation.

We hypothesized that IH or IC induces specific alterations in the gut microbiome and their metabolites which may promote atherosclerosis. The questions we sought to address in the

current study were (1) what is the particular role of IH or IC in inducing or promoting atherosclerosis, (2) what is the response of the vascular system (i.e., aorta vs pulmonary artery) to IH or IC in term of atherogenesis, (3) what is the signature of IH or IC on gut microbiome and metabolites, and (4) what is the potential impact of these changes of gut microbiome and metabolites on the development of atherosclerosis?

3.2 Materials and Methods

Animals with High Fat Diet

Atherosclerosis-prone ten-week old male *ApoE*$^{-/-}$ mice on C57BL/6J background (002052; The Jackson Laboratory, Bar Harbor, ME) were used (Piedrahita et al., 1992) and *ApoE* deficiencies were confirmed by PCR. The mice were given a high fat diet (HFD) containing 1.25% cholesterol and 21% milk fat (4.5 Kcal/g; TD.96121; Envigo-Teklad Madison, WI) for 10 weeks while being exposed to either IH, IC or room air. All animal protocols were approved by the Animal Care Committee of the University of California, San Diego and followed the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health.

Intermittent Hypoxia and Hypercapnia Exposure

Intermittent hypoxia (IH) or intermittent hypercapnia (IC) were administered in a computer-controlled atmosphere chamber (OxyCycler, Reming Bioinstruments, Redfield, NY) as previously described (Xue et al., 2017). Mice were exposed to 8% O_2 or 8% CO_2 for short periods (~4 min) separated by alternating periods (~4 min) of normoxia [O_2] = 21%) and normocapnia ([CO_2] = ~0.5%) with 1–2 min ramp intervals, 10 minutes per cycle, 10 hours per day during the light cycle, for 10 weeks. Control mice were on the same HFD but in room air (21% O_2 and 0.5% CO_2).

Quantification of Atherosclerotic Lesions

Atherosclerosis was quantified by computer-assisted image analysis (ImageJ, NIH Image)(Schneider et al., 2012) in Sudan Red-stained en face preparations of the aorta and pulmonary arteries as previously described (Xue et al., 2017). The extent of lesion was presented by the percentage of Sudan IV-stained area to the total area of the tissue examined. Images of the aortic arch were cropped from the rest of the aorta by measuring the same distance from the bifurcation to the aortic body using photo-editing software (Adobe Photoshop CS6, Adobe Systems Inc., San Jose, CA). All the measurements were done by blinded investigators. Data were presented as means ± SEM. Mann-Whitney U test was employed and $p<0.05$ was considered statistically significant.

Microbiome

Fecal samples were collected consistently between 9AM and 11AM (ZT3-ZT5) on collection days and stored at -80°C until the end of the study. We chose to collect samples at ZT3-ZT5 due to a concomitant circadian study from our group indicating it was the time of greatest microbiome composition differences between IHC and Air (Unpublished data). Then, samples were prepared for sequencing and analysis in a manner consistent with the Earth Microbiome Project standard protocols (http://www.earthmicrobiome.org/protocols-and-standards/16s) (Caporaso et al., 2012). The V4 region of 16S rRNA gene was sequenced using the primer pair 515f to 806r with Golay error-correcting barcodes on the reverse primer. After processing, raw sequence data was uploaded to Qiita (Gonzalez et al., 2018) (QIITA #11829) and processed using the Deblur (Amir et al., 2017) workflow with default parameters into a BIOM format table. The BIOM table was processed through QIIME 2 (version 2019.10) (Bolyen et al., 2019). Dataset was rarified to 9,400 reads to control for sequencing effort. Weighted UniFrac (Chang et al., 2011) distances were used for microbiome PCoA plots and significance was tested using PERMANOVA

(Anderson). The assigned sequence variants (ASVs) were collapsed to the Family taxonomic level. Differential abundance screening was performed using a permutation-based test with FDR correction in Calour (Xu et al., 2019), http://biocore.github.io/calour/index.html# (Jiang et al., 2017). The bacterial families were selected based on significance present under IC and IH conditions as well as previously known influence on phenotype. Data was visualized using EMPeror (Vazquez-Baeza et al., 2013) and custom python scripts (https://github.com/knightlab-analyses/longitudinal-osa).

Metabolome

Each fecal sample was examined by untargeted liquid chromatography-tandem mass spectrometry (LC-MS/MS) in the same manner as described previously (Tripathi et al., 2018). In brief, the samples were homogenized and then transferred to a standard 96-well plate and analyzed on a Vanquish ultrahigh-performance liquid chromatography (UPLC) system coupled to a Q Exactive orbital ion trap (Thermo Fisher Scientific, Bremen, Germany). A C18 core shell column (Kinetex column, 50 by 2 mm, 1.7-µm particle size, 100-Å pore size; Phenomenex, Torrance, CA) was used for chromatographic separation. Raw spectra were converted to *m/z* extensible markup language (mzXML) in centroid mode using MSConvert (part of ProteoWizard) (Chambers et al., 2012). After isotope peak removal and alignment of peaks, MZmine2 (Pluskal et al., 2010) was used to create a feature matrix containing the feature retention times and the exact mass and peak areas of the corresponding extracted ion chromatograms. Identification of molecular features was performed using MS1-based feature detection and MS2-based molecular networking using the GNPS workflow (https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp). The actual GNPS jobs can be found at the following URL: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=78acff728c48421497ebf59441e18ea4. We used authentic bile acid standards from Cayman Chemical (Ann Arbor, MI) for level 1 identification of metabolites as defined by the

2007 metabolomics standards initiative (Sumner et al., 2007). There was an annotation rate of 41% (556/1359). The unannotated frequency table was analyzed using QIIME2 (version 2019.10). Canberra distances (Lance and Williams, 1967) were used for metabolomic PCoA plots and significance was tested using PERMANOVA (Anderson). Data was visualized using EMPeror (Vazquez-Baeza et al., 2013). Differential abundance screening was performed using a permutation-based test with FDR correction in Calour (Jiang et al., 2017; Xu et al., 2019). Features were selected based on significance present under IC and IH conditions as well as previously known influence on phenotype. The MS/MS spectral annotations were determined by using MS/MS-based spectral library matches on GNPS for level 3 identification of all non-bile acid molecules represented (Sumner et al., 2007). Bile acid standards were run using the same method for level 1 identification for all represented bile acids. After initial processing, a single sample that clustered with blank controls was dropped from further analyses. Total sum normalization was used to calculate the relative abundances. Values were then plotted together using custom python scripts (https://github.com/knightlab-analyses/longitudinal-osa).

Longitudinal Data Statistical Analysis

For both the microbiome and metabolome, linear mixed effect modeling performed using q2-longitudinal (Bokulich et al., 2018) to determine if the highlighted feature was significantly different between exposure conditions (Air/IH/IC) with respect to time. The formula used was: Feature (microbiome or metabolome) ~ host_age * exposure_type.

Data Availability

Microbiome: EBI accession ERP110592

Metabolome: IH/IC MSV000082973, MassIVE link for IH/IC: https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=2ee8ad1a1c764aaf96677d480617c040.

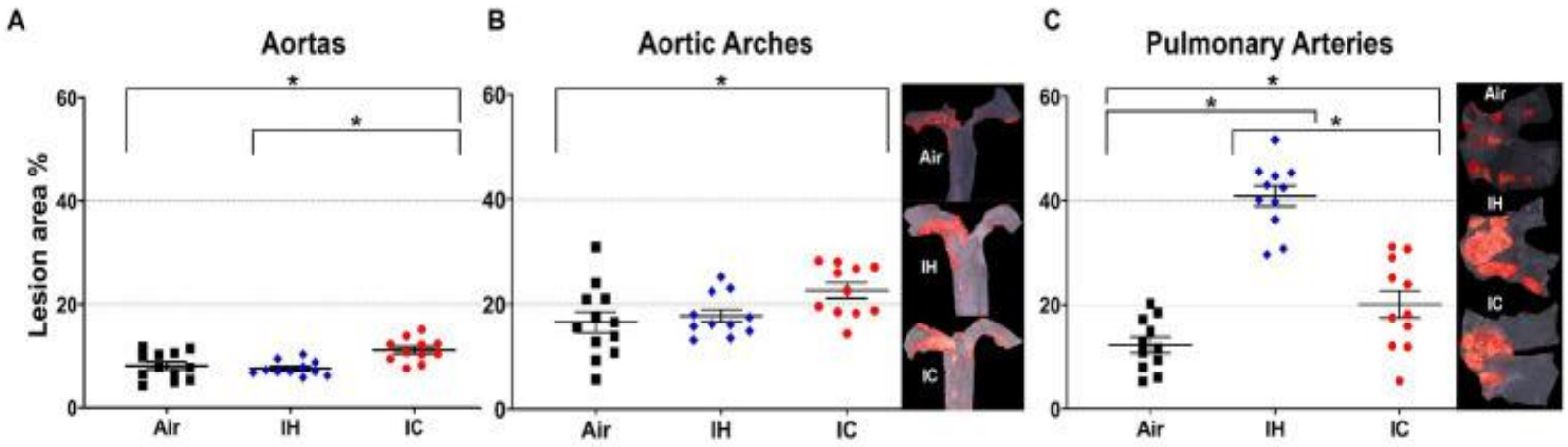
GitHub link (https://github.com/knightlab-analyses/longitudinal-osa).

3.3 Results

Different Impacts of IH and IC on the Progression of Atherosclerosis in the Aorta, Aortic Arch and Pulmonary Artery

OSA patients suffer from oscillatory gas changes (IH and IC). To find out the effect of these gas changes on atherosclerosis, we examined the atherosclerotic lesions in the aorta, aortic arch, and pulmonary artery (PA) after 10 weeks of IH or IC exposure in the presence of HFD. Our data showed that there were significant increases in lesion formation in the aorta and aortic arch in IC exposed mice as compared with Air controls (aorta, Air 8.1±0.8% vs IC 11.1±0.7%, $p<0.05$; aortic arch, Air 16.6±2.0% vs IC 22.6±1.5%, $p<0.05$)(Fig. 3.1A and 3.1B). However, IH-treated mice showed a similar degree of lesions as Air controls without a significant difference (aorta, Air 8.1±0.8% vs IH 7.6±0.4%, $p>0.05$; aortic arch, Air 16.6±2.0% vs IH 17.8±1.2%, $p>0.05$) (Fig. 3.1A and 3.1B). In the PA, the larger atherosclerotic formation was detected in the IH group, followed by IC, both of them showing significantly greater lesions than air controls (PA, IH 40.9±2.0% vs IC 20.1±2.6% vs Air 12.2±1.5%, $p<0.05$)(Fig. 3.1C). Our data suggest that (1) IH promotes atherosclerosis much more in the PA than in the aorta and (2) IC contributes to atherosclerosis in the aorta, aortic arch and PA.

Figure 3.1.. Atherosclerotic lesions in the (A) aortas, (B) aortic arches and (C) pulmonary arteries of ApoE-/- mice after 10-wk treatments. Mice were exposed to either Air, IH or IC for 10 weeks. The en-face lesions were quantified as the percentage of lesion area in the total area of the blood vessel examined. Side panels: representative Sudan IV-stained images of aortic arch (B) and PA (C). Statistical significance (Mann-Whitney U test), * $p<0.05$.


A
Aortas
B
Aortic Arches
C
Pulmonary Arteries
Lesion area %
60
40
20
0
Air
IH
IC
*
*
Air
IH
IC

Different Impacts of IH and IC on the Gut Microbiome and Metabolome

Examination of 16S rRNA gene amplicon sequencing data of IH/IC/Air resulted in the selection of the top most differentially abundant microbial families (Table S1). The assigned sequence variants (ASVs) were grouped together at the Family level for comparisons. Differential abundance testing was calculated by permutation test with discrete FDR correction in Calour (Xu et al., 2019). The top six differentially abundant families were Peptostreptococcaceae, Ruminococcaceae, Erysipelotrichaceae, Verrucomicrobiaceae, Coriobacteriaceae, and Lachnospiraceae. The significance of each selected family with respect to time and experimental condition was determined by linear mixed effect modeling.

Data from untargeted liquid chromatography with tandem mass spectrometry revealed 116 unique metabolites in common (Table S1). Metabolites were divided into five different categories: phosphocholines, acylcarnitines, phosphoethanolamines, bile acids, and other. The other category included amino acids, dipeptides, and other small molecules. The following figures displayed six representative metabolites from each of the different categories based on differential abundance testing calculated by permutation test with discrete FDR correction (Xu et al., 2019), the greatest differences noted in at least one condition, and relevance to atherosclerosis. The six highlighted metabolites were TβMCA [bile acid], UDCA [bile acid], 1-Stearoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine [LysoPE(18:0)] [phosphoethanolamines], 1-Hexadecyl-sn-glycero-3-phosphocholine [Lyso-PAF C-16] [phosphocholines], Oleoyl L-carnitine [acylcarnitines], and tryptophan [other].

A) IH versus Air

Mice under IH conditions diverged significantly from Air controls over time in the beta diversity distances of both the microbiome (Fig. 3.2A, 3.S1A, 3.S3A and C) (pseudo-F 39.775; $p<0.001$) and metabolome (Fig. 3.3A, 3.S2A, 3.S3B and D) (pseudo-F 5.963; $p=0.002$). The

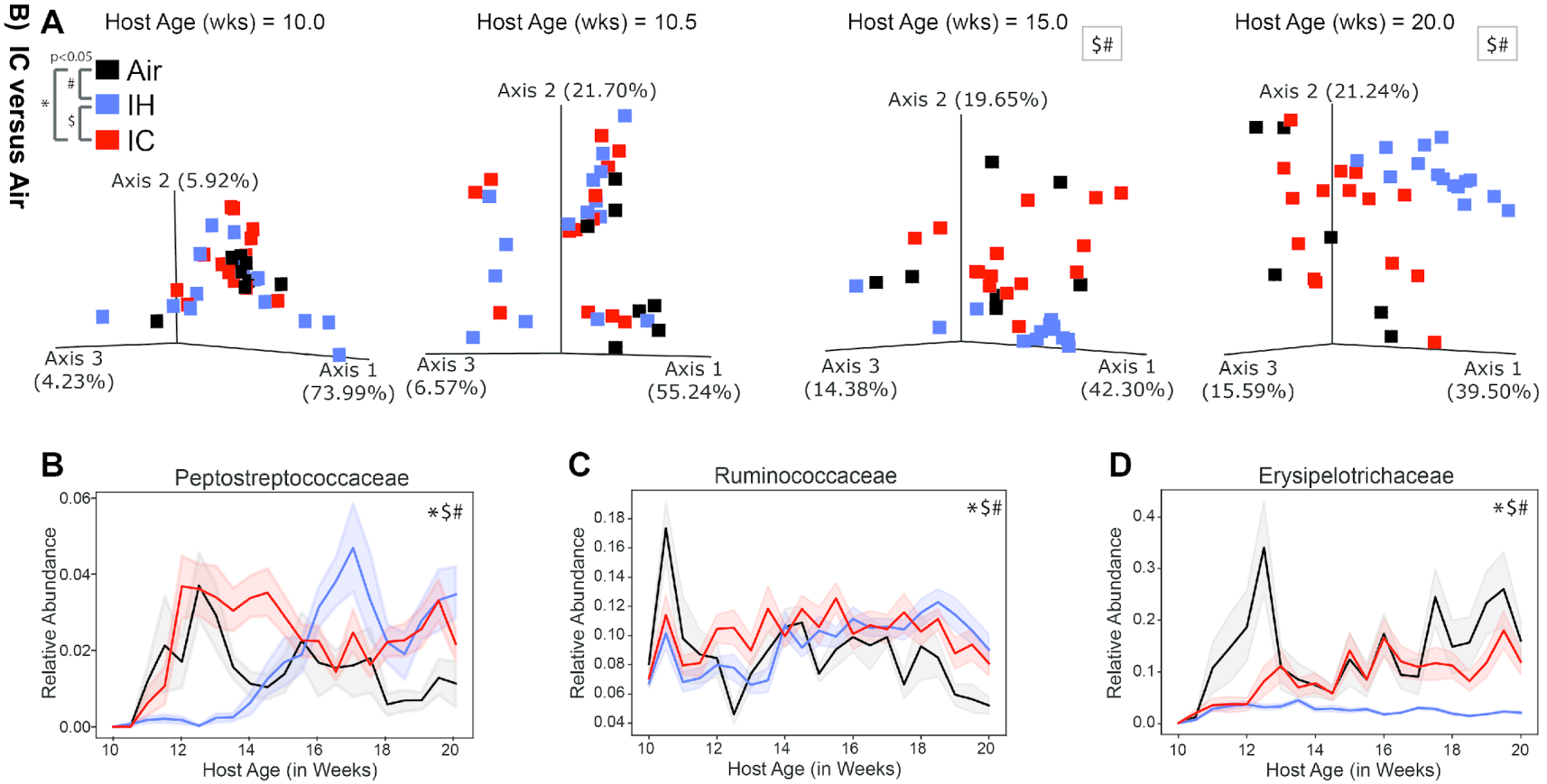
microbiome had a lower p-value and higher pseudo-F value than the metabolome, indicating greater separation between the two conditions for the microbiome.

All of the microbial families shown (Fig. 3.2B to 2D and 3.S1C to 3.S1E) had significant differences between IH-conditioned mice and Air controls. Fig. 3.S1B was a heatmap of the top six differential microbial families under Air, IH, IC conditions and presented a global overview of the changes in gut microbiota over the treatment time. Pro-atherosclerotic (Koeth et al., 2013) Peptostreptococcaceae ($p<0.001$) (Fig. 3.2B) initially showed reduced relative abundance compared to controls. But at week 14, two weeks before the phenotype was known to appear, the relative abundance started to dramatically increase in comparison to controls. The reason for the switch was unclear. Interestingly, anti-atherosclerotic (Cani and de Vos, 2017) Verrucomicrobiaceae ($p<0.001$) (Fig. 3.S1D) exhibited the opposite trend change to Peptostreptococcaceae, i.e. relative abundance was higher during the first 4 weeks of IH exposure then became significantly lower than controls afterwards. IH-conditioned mice had increased relative abundances in pro-atherosclerotic (Wang et al., 2015; Liu et al., 2018) Ruminococcaceae ($p<0.001$) (Fig. 3.2C), pro-atherosclerotic (Karlsson et al., 2012) Coriobacteriaceae ($p<0.001$) (Fig. 3.S1C), and pro-atherosclerotic (Wang et al., 2015; Liu et al., 2018) Lachnospiraceae ($p<0.001$) (Fig. 3.S1E) compared to Air control mice over the course of the experiment. Conversely, IH-conditioned mice had decreased relative abundances of Erysipelotrichaceae ($p=0.009$)(Fig. 3.2D). Erysipelotrichaeceae is known to be associated with cholesterol metabolism (PMID: 23124234).

Three of the six represented metabolites were significantly different between IH-conditioned mice and controls over the course of the study. Fig. 3.S2B was a heatmap of six differential metabolites under Air, IH, IC conditions and presented a global overview of their changes over the treatment time. Compared to control mice, IH-conditioned mice had decreased relative abundance of LysoPE(18:0) ($p=0.031$) (Fig. 3.3D) and increased relative abundances of

pro-inflammatory Lyso-PAF C-16 ($p<0.001$) (Fig. 3.S2C) and oleoyl L-carnitine ($p=0.002$) (Fig. 3.S2D). Of note, the relative abundance of pro-atherosclerotic bile acid TβMCA was increased as well as relative abundances of anti-atherogenic bile acid UDCA and anti-inflammatory tryptophan were decreased in IH-conditioned mice via permutation test with discrete FDR correction (TβMCA, $p<0.001$; UDCA, $p<0.001$; tryptophan, $p<0.001$) (Fig. 3.3B, 3.3C and 3.S2E). However, these differences didn't meet significant criteria of linear mixed effect modeling (relative_abundance~host_age*exposure_type, TβMCA, $p=0.053$; UDCA, $p=0.371$; tryptophan, $p=0.446$). Overall, the gut luminal environment appeared to be pro-inflammatory and pro-atherosclerotic under IH condition.

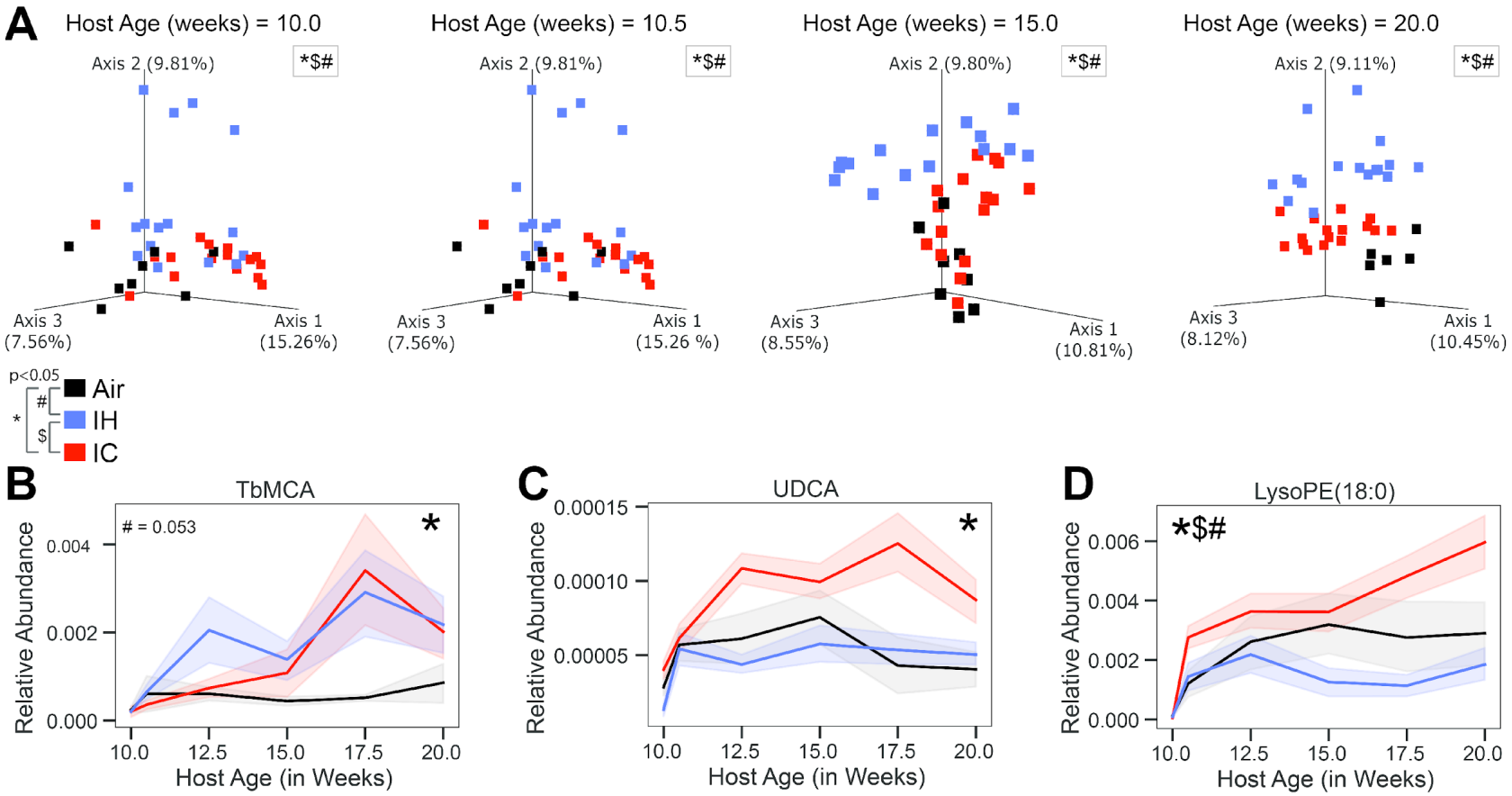
Figure 3.2.. 16S microbiome of ApoE-/- mice on HFD during chronic 10-wk treatment. (A) Weighted UniFrac PCoA of the microbiome during 4 key time points, each time point calculated in isolation. Longitudinal relative abundance values for: (B) family Peptostreptococcaceae; (C) family Ruminococcaceae; (D) family Erysipelotrichaceae. PERMANOVA used for statistical comparisons at the population level. Linear Mixed Effect (LME) modeling used for statistical comparisons of individual families over time. The shaded areas in parts B-D represent standard error of the mean. Air/controls are black (n=6), IH is blue (n=12), IC is red (n=12). Statistical significance $p<0.05$, * IC vs Air, # IH vs Air and $ IH vs IC.


B) IC versus Air
A
Host Age (wks) = 10.0
Host Age (wks) = 10.5
Host Age (wks) = 15.0
Host Age (wks) = 20.0
p<0.05
Air
IH
IC
Axis 2 (5.92%)
Axis 3 (4.23%)
Axis 1 (73.99%)
Axis 2 (21.70%)
Axis 3 (6.57%)
Axis 1 (55.24%)
Axis 2 (19.65%)
Axis 3 (14.38%)
Axis 1 (42.30%)
Axis 2 (21.24%)
Axis 3 (15.59%)
Axis 1 (39.50%)
B
Peptostreptococcaceae
C
Ruminococcaceae
D
Erysipelotrichaceae
Relative Abundance
Host Age (in Weeks)

Compared to Air-controls, IC-conditioned mice also had significant differences in the beta diversity distances of both the microbiome (Fig. 3.2A, 3.S1A, and 3.S3E and G) (pseudo-F 11.274; $p<0.001$) and metabolome (Fig. 3.3A, 3.S2A, 3.S3 F and H) (pseudo-F 3.046; $p=0.004$). The magnitude of the pseudo-F value associated with the comparison of the IC microbiome to controls was substantially less than seen for IH, suggesting differences compared to the control group were not as robust. Also, the separation between IC and control samples was less obvious by the end of the study for both the microbiome and metabolome than it was for IH.

Three of the six microbial families were significantly different by linear mixed effect modeling in IC-conditioned mice as compared to controls, i.e. Peptostreptococcaceae ($p=0.039$) (Fig. 3.2B), Ruminococcaceae ($p=0.003$) (Fig. 3.2C), and Erysipelotrichaceae ($p=0.04$) (Fig. 3.2D). In addition, four metabolites demonstrated overall significant differences. The relative abundances were increased for pro-atherosclerotic TβMCA ($p=0.013$) (Fig. 3.3B), LysoPE(18:0) ($p=0.041$) (Fig. 3.3D) and Lyso-PAF C-16 ($p=0.002$) (Fig. 3.S2C) as well as anti-atherogenic UDCA ($p=0.022$) (Fig. 3.3C) in IC-conditioned mice as compared to controls. While increased serum levels of acylcarnitines have been associated with adverse cardiovascular events (Strand et al., 2017), the observed changes of oleoyl L-carnitine ($p=0.147$) (Fig. 3.S2D) under IC were less impressive than those seen in IH exposure compared to their respective controls. Collectively, IC showed the unique microbiome and metabolomic signatures that distinguish from those changes in Air and IH.

Figure 3.3.. Untargeted LC-MS/MS metabolomics of ApoE-/- mice on HFD during chronic 10-wk treatment. (A) Canberra PCoA of the metabolome during 4 key time points, each time point calculated in isolation. Longitudinal relative abundance values for: (B) Tauro-β-muricholic acid (TβMCA) [Level 1 identification]; (C) Ursodeoxycholic acid (UDCA) [Level 1 identification]; (D) lysophosphoethanolamine 1-Stearoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine [LysoPE(18:0)] [Level 3 identification]. PERMANOVA used for statistical comparisons at the population level. Linear Mixed Effect (LME) modeling used for statistical comparisons of individual metabolites over time. The shaded areas in parts B-D represent standard error of the mean. Air/controls are black (n=6), IH is blue (n=12), IC is red (n=12). Statistical significance $p<0.05$, * IC vs Air, # IH vs Air and $ IH vs IC.


A
Host Age (weeks) = 10.0
Host Age (weeks) = 10.5
Host Age (weeks) = 15.0
Host Age (weeks) = 20.0
*$#
Axis 2 (9.81%)
Axis 2 (9.80%)
Axis 2 (9.11%)
Axis 3 (7.56%)
Axis 1 (15.26%)
Axis 3 (8.55%)
Axis 1 (10.81%)
Axis 3 (8.12%)
Axis 1 (10.45%)
p<0.05
#
*
$
Air
IH
IC
B
TbMCA
= 0.053
C
UDCA
D
LysoPE(18:0)
Relative Abundance
Host Age (in Weeks)

C) IH versus IC

Comparison of IC with IH revealed potential factors that may contribute to their different atherosclerotic phenotypes. This comparison showed significant differences in beta diversity distances of both the microbiome (Fig. 3.2A, 3.S1A, 3.S3I and K) (pseudo-F 50.286; $p<0.001$) and metabolome (Fig. 3.3A, 3.S2A, 3.S3J and L) (pseudo-F 6.411; $p=0.002$). The microbiome and metabolome pseudo-F values indicate that IH was as different from IC as it was from Air controls.

The relative abundances of the top six microbial families were significantly different between the IH and IC-conditioned mice. Similar to the differences between IH and Air, IH-conditioned mice had relatively decreased levels of Peptostreptococcaceae ($p<0.001$) (Fig. 3.2B), Erysipelotrichaceae ($p<0.001$) (Fig. 3.2D) and Verrucomicrobiaceae ($p<0.001$) (Fig. 3.S1D) compared to IC-conditioned mice. Additionally, Coriobacteriaceae ($p<0.001$) (Fig. 3.S1C) and Lachnospiraceae ($p<0.001$) (Fig. 3.S1E) were increased in IH-conditioned mice compared to IC-conditioned mice. However, the relative abundance of Ruminococcaceae was increased under IH compared to air controls but was decreased under IH compared to IC ($p<0.001$) (Fig. 3.2C).

Furthermore, the difference seen in LysoPE(18:0) (Fig. 3.3D) when comparing IH to IC were similar to the difference seen in the comparison between IH and air. However, unlike the comparison between IH and Air, there was no significant difference for Lyso-PAF C-16 ($p=0.502$) (Fig. 3.S2C) and oleoyl L-carnitine ($p=0.088$) (Fig. 3.S2D) between IH and IC.

Together, these findings demonstrated that though IC-conditioned mice present unique differences, IH-conditioned mice differed from controls far more than IC-conditioned mice (IH>IC>Air).

3.4 Discussion

Although OSA patients suffer from both episodic hypoxia and hypercapnia, only the impact of IH has been extensively studied from a cardiovascular viewpoint. However, IC is also a key player in the pathophysiology of human diseases (Wang et al., 2016). In this work, the respective roles of IH and IC in the progression of atherosclerosis were defined by examining them separately. We have made several important observations.

First, as compared to Air, both IH and IC accelerated the development of atherosclerosis, in an animal model of atherosclerosis. The main difference between each condition was the extent of atherosclerosis in the pulmonary artery vs. the aorta/aortic arch. More importantly, this study convincingly showed that IC alone can promote atherosclerosis.

Second, IH exacerbated atherosclerosis in the pulmonary artery far more than it did in the aorta. One possible explanation for this is the inherent response of the vascular bed to hypoxia. Depending on its extent and duration, hypoxia can cause vasodilation in most tissues of the body. By contrast, in the lungs, hypoxia induces vasoconstriction resulting in pulmonary arterial hypertension and damage to endothelial cells, both of which are atherogenic. Another possibility is the vascular response to the microbiome or the metabolome that results from the interaction between the host and gut microbiota. This is a still a major area of investigation and it is difficult at present to dissect the effect of the microbiota, inflammatory response and cytokines on the blood vessels themselves.

We have shown that atherosclerosis can be promoted by IHC in the pulmonary artery trunk and its proximal branches of both *ApoE*$^{-/-}$ and *Ldlr*$^{-/-}$ mice (Douglas et al., 2013; Xue et al., 2017; Imamura et al., 2019), demonstrating that the effect of IHC on the pulmonary artery is not genetic background-dependent. The current IH and IC data corroborate the notion that pulmonary artery atherosclerosis is promoted by these blood gas changes as well. Of note, IHC causes pulmonary hypertension (PH) and right ventricular strain (Douglas et al., 2013). The causal

relationship between pulmonary artery atherosclerosis and PH is not clear. Pulmonary artery atherosclerosis has been observed in patients with diseases that cause PH, such as atrial septal defect and shunt and chronic obstructive pulmonary disease (Russo et al., 1999; Nascimento et al., 2009) and was associated with hypertensive pulmonary vascular disease (Moore et al., 1982). As a matter of fact, OSA patients have elevated pulmonary artery pressure (Minai et al., 2009) and a higher incidence of acute pulmonary embolism (Alonso-Fernandez et al., 2013).

In the past, greater attention was given to OSA-related hypoxia. IH alone with HFD exacerbated aortic atherosclerosis in *ApoE*$^{-/-}$ mice (Jun et al., 2010). However, IH did not seem to increase aortic lesions compared to Air controls in our study. This discrepancy between previous studies and ours is probably related to multiple factors, including the specific experimental protocol of IH exposure, exposure duration, and composition of HFD.

Third, current study demonstrated that IC alone facilitates the development of atherosclerosis in the aorta and pulmonary artery. Previous studies have shown that hypercapnia has both beneficial and deleterious effects (Shigemura et al., 2017). On the one hand, hypercapnia has been associated with improved outcome in patients with acute lung injury (Acute Respiratory Distress Syndrome Network et al., 2000), which is thought to be mediated by inhibition of the NF-κB inflammatory pathway (Contreras et al., 2015). In addition, hypercapnia inhibits hypoxia-induced pulmonary vascular remodeling (Ooi et al., 2000) and prevents hypoxia-induced pulmonary hypertension (Kantores et al., 2006). On the other hand, hypercapnia injures alveolar epithelial cells (Lang et al., 2000), impairs lung edema clearance (Vadasz et al., 2008) and reduces alveolar epithelial repair (O'Toole et al., 2009). It also modulates innate immunity and host defense that increase the susceptibility to and mortality of pulmonary infections (Gates et al., 2013).

Gut microbiota can have either protective or deleterious effects in the host. In terms of atherosclerosis, bacterial family Peptostreptococcaceae has been shown to be positively

associated with both an omnivorous diet and trimethylamine N-oxide (TMAO) production in humans (Koeth et al., 2013), suggesting its potential role in metabolism of dietary L-carnitine into trimethylamine (TMA). TMA is further oxidized to TMAO by hepatic flavin monooxygenases. TMAO enhances macrophage cholesterol accumulation, foam cell formation and atherosclerosis, all correlated with an increased risk of heart attack, stroke and death (Wang et al., 2011; Tang et al., 2013). Increased relative abundance of Peptostreptococcaceae was observed after 4-week IH and IC exposures, two weeks before the atherosclerosis became evident, indicating that Peptostreptococcaceae plays a role in IH- and IC-induced atherosclerosis.

Akkermansia muciniphila, the only genus of family Verrucomicrobiaceae in mammals, is believed to have health benefits in humans. Its abundance is known to be inversely correlated with obesity, diabetes, cardiometabolic diseases including atherosclerosis and low-grade inflammation (Cani and de Vos, 2017). The relative abundance of *A. muciniphila* was ultimately decreased under IH and presumed to have been more permissive to the development of atherosclerotic lesions. Based on recent findings (Li et al., 2016a), reduced *A. muciniphila* under IH may contribute to IH-induced atherosclerosis by causing a leaky gut which allows pro-inflammatory mediators to be more readily absorbed systemically. However, increased abundance of *A. muciniphila* was found under IC, potentially indicating an attempt to return to homeostasis.

Another interesting example linking atherosclerosis to microbiota is related to the bacterial families Ruminococcaceae, Lachnospiraceae and Coriobacteriaceae. Previous studies have shown that Ruminococcaceae and Lachnospiraceae were positively correlated with atherosclerotic lesion size in *ApoE*$^{-/-}$ mice (Wang et al., 2015; Liu et al., 2018) and Coriobacteriaceae was enriched in patients with symptomatic atherosclerosis (Karlsson et al., 2012). These findings suggest that these bacteria play an important role in atherogenesis in our mice. Although the underlying mechanism is not well understood, it is often attributed to their roles

in inflammatory pathways (Martinez-Medina et al., 2006; Andoh et al., 2007), lipid metabolism (Martinez et al., 2009; Liu et al., 2015) and trimethylamine N-oxide (TMAO) production (Wang et al., 2015; Qiu et al., 2018; Chen et al., 2019). Ruminococcaceae and Lachnospiraceae species hydrolyze starch and other sugars to produce butyrate and other short-chain fatty acids (SCFAs). SCFAs are ligands for the G protein-coupled receptors 41 and 43 and are hypothesized to reduce inflammatory pathways by activating these receptors (Biddle et al., 2013; Ohira et al., 2017). Notably, Ruminococcaceae and Lachnospiraceae contain bile acid inducible (*bai*) genes which encode enzymes involved in 7α-dehydroxylation and can convert host primary bile acids to secondary bile acids (Vital et al., 2019). The secondary bile acids act through farnesoid X receptor (FXR) and G-protein-coupled bile acid receptor (TGR5) and affect host metabolism and health. In fact, both FXR and TGR5 have been implicated in atherosclerosis (Hanniman et al., 2005; Miyazaki-Anzai et al., 2014; Miyazaki-Anzai et al., 2018).

One of the top differentially abundant metabolites is tauro-β-muricholic acid (TβMCA) which is a primary bile acid synthesized in mice. Evidence has shown that the level of TβMCA can be affected by gut microbiota (Sayin et al., 2013) and induced by TMAO, which is an important molecule in the development of atherosclerosis (Ding et al., 2018). Since a) TβMCA is a naturally occurring FXR antagonist (Sayin et al., 2013) and b) it has been reported that loss of functional FXR increases atherosclerotic lesions in *ApoE*$^{-/-}$ mice along with a more atherogenic plasma lipid and lipoprotein profile (Hanniman et al., 2005), it is likely that the elevated TβMCA under IH and IC facilitates atherosclerotic formation.

Increased abundance was detected for Lyso-PAF C-16 under IH and IC conditions. Lysophosphatidylcholine is known to promote atherosclerosis by various mechanisms including activating endothelial cells to produce reactive oxygen species, enhancing the release of pro-inflammatory cytokines, attracting immune cells to the vascular endothelial wall and mediating atherogenic activity of ox-LDL (Androulakis et al., 2005; Matsumoto et al., 2007; Li et al., 2016b).

This suggests that Lyso-PAF C-16 plays an important role in the progression of atherosclerosis induced by IH and IC.

3.5 Conclusion

Taken altogether, we are the first to show that IC, in addition to IH, contributes to OSA-related atherosclerosis. Moreover, we also demonstrate that intermittent hypoxia and hypercapnia, a hallmark of OSA, change the gut microbiota and metabolites. The changes in the gut luminal environment likely influence the development of atherosclerosis by modulating host gut permeability, inflammatory responses, microbial metabolites TMA/TMAO, as well as bile acid and lipid metabolism. The knowledge obtained in the current study paves the way for a better understanding of the mechanistic link between IH/IC, gut microbiome, and OSA-induced atherosclerosis.

Non-standard Abbreviations and Acronyms:

OSA	obstructive sleep apnea
IHC	intermittent hypoxia and hypercapnia
IH	intermittent hypoxia
IC	intermittent hypercapnia
ApoE	apolipoprotein E
Ldlr	low density lipoprotein receptor
HFD	high fat diet
PA	pulmonary artery
PH	pulmonary hypertension
ASVs	assigned sequence variants
LC-MS/MS	liquid chromatography-tandem mass spectrometry
mzXML	*m/z* extensible markup language
PCoA	principal coordinates analysis
LysoPE(18:0)	1-Stearoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine
Lyso-PAF C-16	1-Hexadecyl-sn-glycero-3-phosphocholine
UDCA	ursodeoxycholic acid
TβMCA	tauro-β-muricholic acid
TMAO	trimethylamine N-oxide
FXR	farnesoid X receptor
TGR5	G-protein-coupled bile acid receptor, Gpbar1

3.6 Acknowledgments

We would like to thank Gail Ackermann and Gregory Humphrey from Dr. Knight's laboratory for their technical assistance. We also thank Travis Smith from Dr. Haddad's laboratory at UCSD for animal care.

Chapter 3, in full, is a reprint of the material as it appears in "Influence of intermittent hypoxia/hypercapnia on atherosclerosis, gut microbiome, and metabolome." Jin Xue, Celeste Allaband, Dan Zhou, Orit Poulsen, Cameron Martino, Lingjing Jiang, Anupriya Tripathi, Emmanuel Elijah, Pieter C Dorrestein, Rob Knight, Amir Zarrinpar, and Gabriel G Haddad. Frontiers in Physiology, 12:663950. The dissertation author was a primary investigator and a first author of this paper.

3.7 Sources of Funding

CA is funded by NIH T32 OD017863. AZ is supported by American Heart Association Beginning Grant-in-Aid (16BGIA27760160), NIH K08 DK102902, R03 DK114536, R21 MH117780, R01 HL148801. GGH is supported by NIH R01 HL146530 and R21 NS111270. Authors receive institutional support from NIH P30 DK120515, P30 DK063491, P30 CA014195, and UL1 TR001442.

3.8 Disclosures

AZ is a co-founder and equity holder of Tortuga Biosciences.

3.9 References

1. Acute Respiratory Distress Syndrome Network, Brower RG, Matthay MA, Morris A, Schoenfeld D, Thompson BT, Wheeler A. (2000). Ventilation with lower tidal volumes as

compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. N Engl J Med 342, 1301-1308. doi: 10.1056/NEJM200005043421801

2. Alonso-Fernández A, de la Peña M, Romero D, Piérola J, Carrera M, Barceló A, Soriano JB, García Suquia A, Fernández-Capitán C, Lorenzo A, García-Río F. (2013). Association between obstructive sleep apnea and pulmonary embolism. Mayo Clin Proc 88, 579-587. doi: 10.1016/j.mayocp.2013.02.005

3. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. (2017). Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems 2. doi: 10.1128/mSystems.00191-16

4. Anderson, M.J. "Permutational Multivariate Analysis of Variance (PERMANOVA)," in Wiley StatsRef: Statistics Reference Online.), 1-15. doi: 10.1002/9781118445112.stat07841

5. Andoh, A., Sakata, S., Koizumi, Y., Mitsuyama, K., Fujiyama, Y., and Benno, Y. (2007). Terminal restriction fragment length polymorphism analysis of the diversity of fecal microbiota in patients with ulcerative colitis. Inflamm Bowel Dis 13, 955-962. doi: 10.1002/ibd.20151

6. Androulakis, N., Durand, H., Ninio, E., and Tsoukatos, D.C. (2005). Molecular and mechanistic characterization of platelet-activating factor-like bioactivity produced upon LDL oxidation. Journal of lipid research 46, 1923-1932. doi: 10.1194/jlr.m500074-jlr200

7. Biddle, A., Stewart, L., Blanchard, J., and Leschine, S. (2013). Untangling the Genetic Basis of Fibrolytic Specialization by Lachnospiraceae and Ruminococcaceae in Diverse Gut Communities. Diversity 5. doi: 10.3390/d5030627

8. Bokulich NA, Dillon MR, Zhang Y, Rideout JR, Bolyen E, Li H, Albert PS, Caporaso JG. (2018). q2-longitudinal: Longitudinal and Paired-Sample Analyses of Microbiome Data. mSystems 3, e00219-00218. doi: 10.1128/mSystems.00219-18

9. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu YX, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS 2nd, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology 37, 852-857. doi: 10.1038/s41587-019-0209-9

10. Cani, P.D., and De Vos, W.M. (2017). Next-Generation Beneficial Microbes: The Case of Akkermansia muciniphila. Front Microbiol 8, 1765. doi: 10.3389/fmicb.2017.01765

11. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal 6, 1621-1624. doi: 10.1038/ismej.2012.8

12. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, Hoff K, Kessner D, Tasman N, Shulman N, Frewen B, Baker TA, Brusniak MY, Paulse C, Creasy D, Flashner L, Kani K, Moulding C, Seymour SL, Nuwaysir LM, Lefebvre B, Kuhlmann F, Roark J, Rainer P, Detlev S, Hemenway T, Huhmer A, Langridge J, Connolly B, Chadick T, Holly K, Eckels J, Deutsch EW, Moritz RL, Katz JE, Agus DB, MacCoss M, Tabb DL, Mallick P (2012). A cross-platform toolkit for mass spectrometry and proteomics. Nature Biotechnology 30, 918-920. doi: 10.1038/nbt.2377

13. Chang, Q., Luan, Y., and Sun, F. (2011). Variance adjusted weighted UniFrac: a powerful beta diversity measure for comparing communities based on phylogeny. BMC Bioinformatics 12, 118. doi: 10.1186/1471-2105-12-118

14. Chen, C.P., Wu, W.-K., Panyod, S., Wu, M.-S., and Lee-Yan, S. (2019). Cardiovascular Disease Protective Effect of Allicin Through Gut Microbiota Modulation (FS07-08-19). Current Developments in Nutrition 3. doi: 10.1093/cdn/nzz040.FS07-08-19

15. Contreras, M., Masterson, C., and Laffey, J.G. (2015). Permissive hypercapnia: what to remember. Curr Opin Anaesthesiol 28, 26-37. doi: 10.1097/ACO.0000000000000151

16. Ding L, Chang M, Guo Y, Zhang L, Xue C, Yanagita T, Zhang T, Wang Y. (2018). Trimethylamine-N-oxide (TMAO)-induced atherosclerosis is associated with bile acid metabolism. Lipids Health Dis 17, 286. doi: 10.1186/s12944-018-0939-6

17. Douglas RM, Bowden K, Pattison J, Peterson AB, Juliano J, Dalton ND, Gu Y, Alvarez E, Imamura T, Peterson KL, Witztum JL, Haddad GG, Li AC (2013). Intermittent hypoxia and hypercapnia induce pulmonary artery atherosclerosis and ventricular dysfunction in low density lipoprotein receptor deficient mice. J Appl Physiol (1985) 115, 1694-1704. doi: 10.1152/japplphysiol.00442.2013

18. Drager, L.F., Polotsky, V.Y., and Lorenzi-Filho, G. (2011). Obstructive sleep apnea: an emerging risk factor for atherosclerosis. Chest 140, 534-542. doi: 10.1378/chest.10-2223

19. Gates KL, Howell HA, Nair A, Vohwinkel CU, Welch LC, Beitel GJ, Hauser AR, Sznajder JI, Sporn PH. (2013). Hypercapnia impairs lung neutrophil function and increases mortality in murine pseudomonas pneumonia. Am J Respir Cell Mol Biol 49, 821-828. doi: 10.1165/rcmb.2012-0487OC

20. Gonzalez A, Navas-Molina JA, Kosciolek T, McDonald D, Vázquez-Baeza Y, Ackermann G, DeReus J, Janssen S, Swafford AD, Orchanian SB, Sanders JG, Shorenstein J, Holste H, Petrus S, Robbins-Pianka A, Brislawn CJ, Wang M, Rideout JR, Bolyen E, Dillon M, Caporaso JG, Dorrestein PC, Knight R. (2018). Qiita: rapid, web-enabled microbiome meta-analysis. Nature Methods 15, 796-798. doi: 10.1038/s41592-018-0141-9

21. Hanniman, E.A., Lambert, G., Mccarthy, T.C., and Sinal, C.J. (2005). Loss of functional farnesoid X receptor increases atherosclerotic lesions in apolipoprotein E-deficient mice. J Lipid Res 46, 2595-2604. doi: 10.1194/jlr.M500390-JLR200

22. Imamura, T., Xue, J., Poulsen, O., Zhou, D., Karin, M., and Haddad, G.G. (2019). Intermittent hypoxia and hypercapnia induces inhibitor of nuclear factor-kappaB kinase subunit beta-dependent atherosclerosis in pulmonary arteries. Am J Physiol Regul Integr Comp Physiol 317, R763-R769. doi: 10.1152/ajpregu.00056.2019

23. Jiang, L., Amir, A., Morton, J.T., Heller, R., Arias-Castro, E., and Knight, R. (2017). Discrete False-Discovery Rate Improves Identification of Differentially Abundant Microbes. mSystems 2. doi: 10.1128/mSystems.00092-17

24. Jonsson, A.L., and Backhed, F. (2017). Role of gut microbiota in atherosclerosis. Nat Rev Cardiol 14, 79-87. doi: 10.1038/nrcardio.2016.183

25. Jun J, Reinke C, Bedja D, Berkowitz D, Bevans-Fonti S, Li J, Barouch LA, Gabrielson K, Polotsky VY. (2010). Effect of intermittent hypoxia on atherosclerosis in apolipoprotein E-deficient mice. Atherosclerosis 209, 381-386. doi: 10.1016/j.atherosclerosis.2009.10.017

26. Kantores C, McNamara PJ, Teixeira L, Engelberts D, Murthy P, Kavanagh BP, Jankov RP. (2006). Therapeutic hypercapnia prevents chronic hypoxia-induced pulmonary hypertension in the newborn rat. Am J Physiol Lung Cell Mol Physiol 291, L912-922. doi: 10.1152/ajplung.00480.2005

27. Karlsson FH, Fåk F, Nookaew I, Tremaroli V, Fagerberg B, Petranovic D, Bäckhed F, Nielsen J. (2012). Symptomatic atherosclerosis is associated with an altered gut metagenome. Nat Commun 3, 1245. doi: 10.1038/ncomms2266

28. Ko WK, Lee SH, Kim SJ, Jo MJ, Kumar H, Han IB, Sohn S. (2017). Anti-inflammatory effects of ursodeoxycholic acid by lipopolysaccharide-stimulated inflammatory responses in RAW 264.7 macrophages. PLoS One 12, e0180673. doi: 10.1371/journal.pone.0180673

29. Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT, Britt EB, Fu X, Wu Y, Li L, Smith JD, DiDonato JA, Chen J, Li H, Wu GD, Lewis JD, Warrier M, Brown JM, Krauss RM, Tang WH, Bushman FD, Lusis AJ, Hazen SL. (2013). Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. Nat Med 19, 576-585. doi: 10.1038/nm.3145

30. Komaroff, A.L. (2018). The Microbiome and Risk for Atherosclerosis. JAMA 319, 2381-2382. doi: 10.1001/jama.2018.5240

31. Lance, G.N., and Williams, W.T. (1967). A general theory of classificatory sorting strategies: II. Clustering systems. The Computer Journal 10, 271-277. doi: 10.1093/comjnl/10.3.271

32. Lang JD Jr, Chumley P, Eiserich JP, Estevez A, Bamberg T, Adhami A, Crow J, Freeman BA. (2000). Hypercapnia induces injury to alveolar epithelial cells via a nitric oxide-dependent pathway. Am J Physiol Lung Cell Mol Physiol 279, L994-1002. doi: 10.1152/ajplung.2000.279.5.L994

33. Li, J., Lin, S., Vanhoutte, P.M., Woo, C.W., and Xu, A. (2016a). Akkermansia Muciniphila Protects Against Atherosclerosis by Preventing Metabolic Endotoxemia-Induced Inflammation in Apoe-/- Mice. Circulation 133, 2434-2446. doi: 10.1161/CIRCULATIONAHA.115.019645

34. Li, Y. F., Li, R. S., Samuel, S. B., Cueto, R., Li, X. Y., Wang, H., & Yang, X. F. (2016b). Lysophospholipids and their G protein-coupled receptors in atherosclerosis. Front Biosci (Landmark Ed) 21, 70-88. doi: 10.2741/4377

35. Liu, B., Zhang, Y., Wang, R., An, Y., Gao, W., Bai, L., Li, Y., Zhao, S., Fan, J., & Liu, E. (2018). Western diet feeding influences gut microbiota profiles in apoE knockout mice. Lipids Health Dis 17, 159. doi: 10.1186/s12944-018-0811-8

36. Liu, S., Bennett, D. C., Tun, H. M., Kim, J. E., Cheng, K. M., Zhang, H., & Leung, F. C. (2015). The effect of diet and host genotype on ceca microbiota of Japanese quail fed a cholesterol enriched diet. Front Microbiol 6, 1092. doi: 10.3389/fmicb.2015.01092

37. Martinez-Medina, M., Aldeguer, X., Gonzalez-Huix, F., Acero, D., and Garcia-Gil, L.J. (2006). Abnormal microbiota composition in the ileocolonic mucosa of Crohn's disease patients as revealed by polymerase chain reaction-denaturing gradient gel electrophoresis. Inflamm Bowel Dis 12, 1136-1145. doi: 10.1097/01.mib.0000235828.09305.0c

38. Martínez, I., Wallace, G., Zhang, C., Legge, R., Benson, A. K., Carr, T. P., Moriyama, E. N., & Walter, J. (2009). Diet-induced metabolic improvements in a hamster model of hypercholesterolemia are strongly linked to alterations of the gut microbiota. Appl Environ Microbiol 75, 4175-4184. doi: 10.1128/AEM.00380-09

39. Matsumoto, T., Kobayashi, T., and Kamata, K. (2007). Role of lysophosphatidylcholine (LPC) in atherosclerosis. Curr Med Chem 14, 3209-3220. doi: 10.2174/092986707782793899

40. Minai, O. A., Ricaurte, B., Kaw, R., Hammel, J., Mansour, M., McCarthy, K., Golish, J. A., & Stoller, J. K. (2009). Frequency and impact of pulmonary hypertension in patients with obstructive sleep apnea syndrome. Am J Cardiol 104, 1300-1306. doi: 10.1016/j.amjcard.2009.06.048

41. Miyazaki-Anzai, S., Masuda, M., Kohno, S., Levi, M., Shiozaki, Y., Keenan, A. L., & Miyazaki, M. (2018). Simultaneous inhibition of FXR and TGR5 exacerbates atherosclerotic formation. J Lipid Res 59, 1709-1713. doi: 10.1194/jlr.M087239

42. Miyazaki-Anzai, S., Masuda, M., Levi, M., Keenan, A.L., and Miyazaki, M. (2014). Dual activation of the bile acid nuclear receptor FXR and G-protein-coupled receptor TGR5 protects mice against atherosclerosis. PLoS One 9, e108270. doi: 10.1371/journal.pone.0108270

43. Moore, G.W., Smith, R.R., and Hutchins, G.M. (1982). Pulmonary artery atherosclerosis: correlation with systemic atherosclerosis and hypertensive pulmonary vascular disease. Arch Pathol Lab Med 106, 378-380.

44. Moreno-Indias, I., Torres, M., Montserrat, J. M., Sanchez-Alcoholado, L., Cardona, F., Tinahones, F. J., Gozal, D., Poroyko, V. A., Navajas, D., Queipo-Ortuño, M. I., & Farré, R. (2015). Intermittent hypoxia alters gut microbiota diversity in a mouse model of sleep apnoea. Eur Respir J 45, 1055-1065. doi: 10.1183/09031936.00184314

45. Nascimento, D., Nunes, L., Oliveira, F., Vencio, E., Teixeira, V., and Reis, M. (2009). Pulmonary atherosclerosis associated with an atrial septal defect in old age: case report of an elderly autopsied patient. Pathol Res Pract 205, 137-141. doi: 10.1016/j.prp.2008.07.008

46. O'Toole, D., Hassett, P., Contreras, M., Higgins, B. D., McKeown, S. T., McAuley, D. F., O'Brien, T., & Laffey, J. G. (2009). Hypercapnic acidosis attenuates pulmonary epithelial wound repair by an NF-kappaB dependent mechanism. Thorax 64, 976-982. doi: 10.1136/thx.2008.110304

47. Ohira, H., Tsutsui, W., and Fujioka, Y. (2017). Are Short Chain Fatty Acids in Gut Microbiota Defensive Players for Inflammation and Atherosclerosis? Journal of Atherosclerosis and Thrombosis 24, 660-672. doi: 10.5551/jat.RV17006

48. Ooi, H., Cadogan, E., Sweeney, M., Howell, K., O'regan, R.G., and Mcloughlin, P. (2000). Chronic hypercapnia inhibits hypoxic pulmonary vascular remodeling. Am J Physiol Heart Circ Physiol 278, H331-338. doi: 10.1152/ajpheart.2000.278.2.H331

49. Piedrahita, J.A., Zhang, S.H., Hagaman, J.R., Oliver, P.M., and Maeda, N. (1992). Generation of mice carrying a mutant apolipoprotein E gene inactivated by gene targeting in embryonic stem cells. Proc Natl Acad Sci U S A 89, 4471-4475. doi: 10.1073/pnas.89.10.4471

50. Pluskal, T., Castillo, S., Villar-Briones, A., and Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics 11, 395. doi: 10.1186/1471-2105-11-395

51. Qiu, L., Tao, X., Xiong, H., Yu, J., and Wei, H. (2018). Lactobacillus plantarum ZDY04 exhibits a strain-specific property of lowering TMAO via the modulation of gut microbiota in mice. Food Funct 9, 4299-4309. doi: 10.1039/c8fo00349a

52. Russo, A., De Luca, M., Vigna, C., De Rito, V., Pacilli, M., Lombardo, A., Armillotta, M., Fanelli, R., & Loperfido, F. (1999). Central pulmonary artery lesions in chronic obstructive pulmonary disease: A transesophageal echocardiography study. Circulation 100, 1808-1815. doi: 10.1161/01.cir.100.17.1808

53. Sayin, S. I., Wahlström, A., Felin, J., Jäntti, S., Marschall, H. U., Bamberg, K., Angelin, B., Hyötyläinen, T., Orešič, M., & Bäckhed, F. (2013). Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. Cell Metab 17, 225-235. doi: 10.1016/j.cmet.2013.01.003

54. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9, 671-675. doi: 10.1038/nmeth.2089

55. Senaratna, C. V., Perret, J. L., Lodge, C. J., Lowe, A. J., Campbell, B. E., Matheson, M. C., Hamilton, G. S., & Dharmage, S. C. (2017). Prevalence of obstructive sleep apnea in

the general population: A systematic review. Sleep Med Rev 34, 70-81. doi: 10.1016/j.smrv.2016.07.002

56. Shigemura, M., Lecuona, E., and Sznajder, J.I. (2017). Effects of hypercapnia on the lung. J Physiol 595, 2431-2437. doi: 10.1113/JP273781

57. Strand, E., Pedersen, E. R., Svingen, G. F., Olsen, T., Bjørndal, B., Karlsson, T., Dierkes, J., Njølstad, P. R., Mellgren, G., Tell, G. S., Berge, R. K., Svardal, A., & Nygård, O. (2017). Serum Acylcarnitines and Risk of Cardiovascular Death and Acute Myocardial Infarction in Patients With Stable Angina Pectoris. J Am Heart Assoc 6. doi: 10.1161/JAHA.116.003620

58. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reily MD, Thaden JJ, Viant MR. (2007). Proposed minimum reporting standards for chemical analysis. Metabolomics 3, 211-221. doi: 10.1007/s11306-007-0082-2

59. Tang, W. H., Wang, Z., Levison, B. S., Koeth, R. A., Britt, E. B., Fu, X., Wu, Y., & Hazen, S. L.. (2013). Intestinal Microbial Metabolism of Phosphatidylcholine and Cardiovascular Risk. New England Journal of Medicine 368, 1575-1584. doi: 10.1056/NEJMoa1109400

60. Tripathi, A., Melnik, A. V., Xue, J., Poulsen, O., Meehan, M. J., Humphrey, G., Jiang, L., Ackermann, G., McDonald, D., Zhou, D., Knight, R., Dorrestein, P. C., & Haddad, G. G. (2018). Intermittent Hypoxia and Hypercapnia, a Hallmark of Obstructive Sleep Apnea, Alters the Gut Microbiome and Metabolome. mSystems 3. doi: 10.1128/mSystems.00020-18

61. Vadász, I., Dada, L. A., Briva, A., Trejo, H. E., Welch, L. C., Chen, J., Tóth, P. T., Lecuona, E., Witters, L. A., Schumacker, P. T., Chandel, N. S., Seeger, W., & Sznajder, J. I. (2008). AMP-activated protein kinase regulates CO2-induced alveolar epithelial dysfunction in rats and human cells by promoting Na,K-ATPase endocytosis. J Clin Invest 118, 752-762. doi: 10.1172/JCI29723

62. Vazquez-Baeza, Y., Pirrung, M., Gonzalez, A., and Knight, R. (2013). EMPeror: a tool for visualizing high-throughput microbial community data. Gigascience 2, 16. doi: 10.1186/2047-217X-2-16

63. Vital, M., Rud, T., Rath, S., Pieper, D.H., and Schluter, D. (2019). Diversity of Bacteria Exhibiting Bile Acid-inducible 7alpha-dehydroxylation Genes in the Human Gut. Comput Struct Biotechnol J 17, 1016-1019. doi: 10.1016/j.csbj.2019.07.012

64. Wang, D., Thomas, R.J., Yee, B.J., and Grunstein, R.R. (2016). Hypercapnia is more important than hypoxia in the neuro-outcomes of sleep-disordered breathing. J Appl Physiol (1985) 120, 1484. doi: 10.1152/japplphysiol.01008.2015

65. Wang, Z., Klipfell, E., Bennett, B. J., Koeth, R., Levison, B. S., Dugar, B., Feldstein, A. E., Britt, E. B., Fu, X., Chung, Y. M., Wu, Y., Schauer, P., Smith, J. D., Allayee, H., Tang, W. H., DiDonato, J. A., Lusis, A. J., & Hazen, S. L. (2011). Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature 472, 57-63. doi: 10.1038/nature09922

66. Wang, Z., Roberts, A. B., Buffa, J. A., Levison, B. S., Zhu, W., Org, E., Gu, X., Huang, Y., Zamanian-Daryoush, M., Culley, M. K., DiDonato, A. J., Fu, X., Hazen, J. E., Krajcik, D., DiDonato, J. A., Lusis, A. J., & Hazen, S. L. (2015). Non-lethal Inhibition of Gut Microbial Trimethylamine Production for the Treatment of Atherosclerosis. Cell 163, 1585-1595. doi: 10.1016/j.cell.2015.11.055

67. Xu, Z. Z., Amir, A., Sanders, J., Zhu, Q., Morton, J. T., Bletz, M. C., Tripathi, A., Huang, S., McDonald, D., Jiang, L., & Knight, R. (2019). Calour: an Interactive, Microbe-Centric Analysis Tool. mSystems 4. doi: 10.1128/mSystems.00269-18

68. Xue, J., Zhou, D., Poulsen, O., Imamura, T., Hsiao, Y. H., Smith, T. H., Malhotra, A., Dorrestein, P., Knight, R., & Haddad, G. G. (2017). Intermittent Hypoxia and Hypercapnia Accelerate Atherosclerosis, Partially via Trimethylamine-Oxide. Am J Respir Cell Mol Biol 57, 581-588. doi: 10.1165/rcmb.2017-0086OC

Chapter 4. Microbiome Diurnal Dynamics Dominate Phenotype Effects, Enabling Replicability When Controlled

Abstract

Although many aspects of microbiome studies have been standardized to improve experimental replicability, none account for how the daily diurnal fluctuations in the gut lumen cause dynamic changes in 16S sequencing and untargeted LC-MS/MS metabolomic results. Here we show that sample collection time affects the conclusions drawn from microbiome studies and are larger than the effect size of a daily experimental intervention or dietary changes. The timing of divergence of the microbiome composition between experimental and control groups are unique to each experiment, and are especially pronounced in 16S experiments. Sample collection times as short as only four hours apart lead to vastly different conclusions. Lack of consistency in the time of sample collection may explain poor cross-study replicability in microbiome research leading to incorrect understanding of the role of the gut microbiome in physiological homeostasis and pathophysiological studies.

4.1 Introduction

In the fourth part of this thesis, we argue that biological rhythms are an underappreciated and under-reported experimental variable that is actually critical to experiment replicability. We re-analyzed five diurnal 16S microbiome datasets plus a single additional new dataset to determine the impact of time of sample collection on the interpretation of results. Using a filtered Weighted UniFrac beta diversity distance matrix for only between condition distances, we examined microbiome composition changes between experimental and control groups. We found that composition dynamics are unique to each experiment with sample collection times as short as only four hours apart could lead to different conclusions. The compositional dynamics of untargeted LC-MS/MS metabolomes appeared less dynamic across the day or longitudinally. These results indicate that inconsistencies in the time of sample collection may at least partially explain the lack of replicability seen in microbiome studies - it is like trying to measure changes in sea level without knowing tides or waves exist. In addition, poor experimental timing may lead to erroneous understanding of the gut microbiome in disease. We recommend that sample collection time be recorded in the methods section of all manuscripts moving forward to increase replicability.

The lack of replicability of microbiome studies has been a barrier to understanding how host-microbe interactions contribute to physiological homeostasis and pathophysiological processes, including heart disease and cancer. (Schloss, 2018). As the field reaches a critical inflection point of moving from descriptive and associative research to mechanistic and interventional studies, the ability to rapidly and reproducibly characterize the microbiome is critical to the development of novel microbiome-mediated therapeutics and diagnostic biomarkers (Gilbert et al., 2018). In early studies, many confounding variables involving model systems, sample collection protocols, and pipeline processing were not routinely accounted for in study design, often resulting in irreproducible, noisy data (Knight et al., 2018; Schloss, 2018). The

investigation of these irreproducible and noisy data led to the discovery of important confounds that influence the results, such as the maternal effect (Ley et al., 2005), cage effect (Deloris Alexander et al., 2006), facility differences (Friswell et al., 2010), as well as laboratory and sample handling protocols (Sinha et al., 2017). However, despite the introduction of standardization of experimental protocols and analysis pipelines, unexplained variability and lack of replicability still plagues microbiome research.

One underexplored factor is that the microbiome is dynamic, and exhibits diurnal oscillations (Alvarez et al., 2020; Frazier and Chang, 2020). The diurnal fluctuation of gut microbiota remains one of the most robustly reproducible characteristics of the luminal environment in murine models of disease. Disruption of microbiome diurnal dynamics by changing the nutritional quality of the diet (Leone et al., 2015; Thaiss et al., 2014; Zarrinpar et al., 2014) or feeding pattern (Zarrinpar et al., 2014), using circadian clock transgenic mice (Liang et al., 2015; Thaiss et al., 2016), or crossing time zones to induce jet-lag (Thaiss et al., 2014) are all associated with metabolic syndrome spectrum diseases (e.g. insulin resistance, increased adiposity) (Alvarez et al., 2020). The gut microbiome is intimately linked to host peripheral circadian rhythms. Microbiome-depleted mice (i.e. antibiotic-induced depletion or germ-free mice) have dampened epithelial and hepatic circadian rhythms (Leone et al., 2015; Mukherji et al., 2013; Weger et al., 2019). Analysis of the microbiome from human stool samples collected from a multitude of time points (Kaczmarek et al., 2017; Skarke et al., 2017), as well as 24-hour salivary collections (Collado et al., 2018; Kohn et al., 2020; Takayasu et al., 2017), suggest that the human microbiome also has diurnal fluctuations. In addition, loss of diurnal dynamics of the gut microbiome was recognized as a risk factor for developing type 2 diabetes in a longitudinal study of a large patient cohort (Reitmeier et al., 2020).

Many labs that study the microbiome anecdotally report collecting their specimens for each experiment at a specific, single time point. However, it is not clear whether the collection

time is chosen rationally based on experimental design, convenience to the experimenters, or if this window of time is consistent between experimental replicates both within and outside of the laboratory. We hypothesize that if this effect is significant enough to affect results, there should be evidence in currently existing diurnal microbiome studies that sampling at different times leads to different conclusions.

Studies with samples collected in a circadian fashion are relatively rare. They are also only generally performed using mice, which defecate with enough frequency to perform the experiment. When searching for the keywords “circadian microbiome” AND “mice” in PubMed (https://pubmed.ncbi.nlm.nih.gov/) for articles published over an 8 year period (from 2014-2021), we found 79 articles that met our initial criteria. Only 66 of those were research articles, and of the remainder we found only 14 articles that contained 16S amplicon sequencing samples collected for more than 3 time points within a 24 or 48 hour period. Of these 14 studies, four had complete publicly available data on ENA/EBI. Of the remainder, four had incomplete datasets on ENA/EBI - (Thaiss et al., 2014) (Thaiss et al., 2016) (Wu et al., 2018) (Ren et al., 2021) - and the rest were not publicly available. We then contacted the authors of all studies with missing or incomplete data and got the following responses: (1) three were unable to locate the missing data (Thaiss et al., 2014) (Thaiss et al., 2016) (Liang et al., 2015), (2) three could not provide data in a format suitable for re-analysis (Beli et al., 2019; Ren et al., 2021; Wang et al., 2020), and three did not respond to repeated inquiries (Guo et al., 2019; Mistry et al., 2020; Shao et al., 2018). After extensive correspondence, we were able to obtain five datasets in a form suitable for re-analysis (Allaband et al., 2021; Leone et al., 2015; Tuganbaev et al., 2020; Wu et al., 2018; Zarrinpar et al., 2014). In addition, we included a recently published dataset from the same mice in one of our circadian studies (Zarrinpar et al., 2014; *manuscript under review*). We also generated a single new dataset for analysis in this paper that is unique in that it includes two circadian collections over the course of a single experiment. Two of the studies (Allaband et al.,

2021; Tripathi et al., 2019) containing matching untargeted LC-MS/MS metabolomics are also analyzed using the same methods, allowing preliminary comparisons to metabolomics. Overall, these studies show the importance of sample collection time in affecting the conclusion of microbiome studies.

4.2 Results

Time of sample collection is critical to microbiome study conclusions

To determine whether the time of sample collection was included in experimental methods of microbiome studies, we reviewed over 550 articles published in 2019 from major journals where new 16S or metagenomic datasets were generated (**Fig S1A, S1B**). Less than 10% of these articles reported any timing information about their sample collection and only 0.32% reported a specific time of sample collection (**Fig S1C**). Since microbiome studies do not commonly report time of sample collection in their methods, we investigated the effects of microbiome sample collection on the potential interpretation of a study. To do so, we used between-condition weighted UniFrac β-diversity (Lozupone and Knight, 2005) distances (BCD) to show how similar the microbiomes from the two conditions are to each other at any given time point. Thus, increasing or decreasing BCD allows us to assess microbiome compositional fluxes between experimental conditions over time. First, we wanted to investigate whether sampling time affects the conclusion of a study with a discrete daily intervention. We started by reanalyzing a previous dataset from our laboratory that had a discrete daily intervention external to the host (Allaband et al., 2021). This study used apolipoprotein E knock-out mice (*Apoe*$^{-/-}$) mice under intermittent hypoxia hypercapnia (IHC) conditions to mimic obstructive sleep apnea conditions (Summary of experiment in **Fig 1A**). In the IHC experiment, we noted that the BCD fluctuated greatly, nearly doubling within a 24hr period (**Fig 1B**), suggesting that compositional assessments from different times would yield radically different results. Since an overwhelming majority of microbiome studies

neither report the time of sample collection, nor confirm that samples are collected at a specific time, and microbiome study guidelines make no explicit recommendation that samples should be collected at the same time, we realized that we had no data supporting our assumption that samples are collected at the same time in a typical microbiome experiment.

In this experiment, BCD increased during IHC exposure, with maximal divergence of the two groups at ZT-6 (**Fig 1B**). Maximal convergence (similarity) occurred at ZT-18, a half day after the maximal divergence when both groups were experimentally similar. Subsequently, despite the lack of the presence of IHC to separate the groups at that time, distances increased during ZT-22 which suggests a potential microbiome response to host anticipatory stress. In addition, the BCD values themselves were found to conform to a diurnal pattern (MetaCycle, JTK method, $p<0.001$). Next, we used the distance matrix to create a heatmap of the average BCD between IHC and control mice for each time point combination to determine all potential outcomes of the study (**Fig 1C**). The highest BCD (greatest divergence) between the two groups was Air ZT-18 and IHC ZT-6, which are 12 hrs apart. The lowest BCD (greatest convergence) between the two groups was Air ZT-22 and IHC ZT-18, both of which occur during the dark phase and are only 4 hrs apart. The highest BCD is 2.8 times the lowest across all timepoints, while the within-condition distances for Air (4.9X) and intermittent-hypoxia-hypercapnia (4.3X) dynamic ranges were greater. The two groups had overall significantly different microbiome compositions (PERMANOVA, all Air vs all intermittent-hypoxia-hypercapnia, $p=0.005$), with ZT-6 driving differences (PERMANOVA, $p=0.035$). All other timepoints showed the two groups as being non-significantly different (PERMANOVA, $p>0.05$). Thus, the beta-diversity of the two conditions can differ 2.8-fold depending on the time of sample collection, potentially affecting the conclusions of the study.

To determine whether these different sampling times affect conclusions of the compositional analysis while accounting for bias caused by relative compositional bias and

unknown microbial loads for each sample (Morton, et al., 2019), we examined log-ratios of biologically relevant phyla and families at the time points corresponding to the highest and lowest BCD (**Fig 1D**). The impact of sample collection time was most obvious at the phylum level, where the relative proportion of Bacteroidetes to Verrucomicrobia shifted strikingly towards Bacteroidetes in mice under IHC conditions at the highest BCD, but the experimental groups were indistinguishable at the lowest BCD. These differences existed at the sub-phylum level as well. For example, the log-ratio balance of three metabolically important families (Ruminococcaceae and S24-7, in relation to Verrucomicrobiaceae) shifted significantly during maximal BCD, but the balance was similar between experimental groups at minimal BCD (**Fig 1E**). Overall, these results demonstrate that time of sample collection has a significant effect on the microbiome composition results that can affect the experimental conclusions if collection is only performed at a single time point. Additionally, these results let us pick optimal sampling times in order to maximize the effect size of the biological conclusions.

Diet and Feeding Pattern Influence Sample Collection Time Results

The experiment in the previous section demonstrates that the BCD increased during an environmental intervention (i.e. exposure to IHC) and decreased when this intervention was removed. We hypothesized that host environmental differences result in changes in stress levels and immune function that result in changes to gastrointestinal luminal microniches and cause divergence between the two conditions (BCD increase), converging again (BCD decrease) as the stress response fades. Since diet and feeding patterns are known to induce large and reproducible effects on the gastrointestinal environment and resulting microniches (Bisanz et al., 2019), we hypothesized that it would be less influenced by diurnal microbiome dynamics.

We pursued this hypothesis by analyzing the results from one of our previously published studies (Zarrinpar et al., 2014) that investigated the effect of diet and feeding patterns on murine host physiology and the diurnal dynamics of the cecal microbiome. In mice on the same diet but

with different feeding schedules, the BCD should change in response to differences in the feeding schedules of the experimental groups. In this experiment (**Fig 2A**), wild-type male C57Bl/6J mice were provided with either a normal chow diet (NCD) or a high-fat diet (HFD). Their access to food was either *ad libitum* or time-restricted (TRF). After 8 weeks on condition and after demonstration of a metabolic phenotype difference between HFD-*ad libitum* and HFD-TRF mice (Zarrinpar et al., 2014), we collected cecal content every 4 hours for 24hrs to examine dynamic changes in microbiome composition over the course of a day.

Since HFD mice are known to eat relatively consistently throughout the day and do not fast during the light phase, we would expect low BCD between HFD-ad libitum and HFD-TRF mice during ZT-17 and ZT-21, when both groups have access to food. We expect high BCD during ZT-1 to ZT-13, when one group has access to food and the other group is forced to fast. As expected, the HFD-*ad libitum* to HFD-TRF BCD was the highest at ZT-13 when the two groups should be the most divergent (**Fig 2B**). We also saw that the HFD-*ad libitum* to HFD-TRF BCD was significantly lower at ZT-17. However, BCD was significantly lower at ZT-5 than ZT-13, and indistinguishable from ZT-17, suggesting that the intestinal environment is not solely influenced by the presence of a consumed diet in the lumen. Furthermore, the heatmap comparing all the combinations of different collection times shows a nearly 2.5-fold difference in peak and nadir BCD. There is a trend of the highest values being in the lower left corner (**Fig 2C**), which indicates that light phase of HFD-TRF and dark phase HFD-*ad libitum* have the greatest divergence. Thus, while the feeding schedule does impact microbiome composition, there are also composition shifts not directly attributable to the experimental design.

Next, we looked at mice on different diets but with the same access to food, NCD-*ad libitum* to HFD-*ad libitum*. Since diet macronutrient profile is a large driver of microbiome differences between cohorts, we wanted to determine if oscillatory dynamics of the gut microbiome could influence microbiome results or even be detectable. We hypothesized that the

greatest differences between the two groups would be when they are eating different diets during the dark phase. Thus, we would expect the highest BCD to occur during ZT-13 to ZT-21 when one group is eating NCD and the other HFD. However, despite having radically different diets, the BCD from all of the dark phase time points are relatively low and are not significantly different from each other (**Fig 2D**). The biggest compositional shifts occurred at the transition from the light phase to the dark phase. The time point of greatest divergence is at ZT-9 when NCD mice are largely fasting, while HFD mice are likely eating at low to moderate levels. The heatmap, which shows a 2.7-fold difference in the peak to nadir BCD, also confirms that NCD-*ad libitum* ZT-9 as being different from all other HFD time points (**Fig 2E**). This same pattern is seen in a separate published dataset using a slightly different diet (Leone et al., 2015) comparing NCD-*ad libitum* and an *ad libitum* milk-fat diet, that also yielded high BCD (groups diverged) during the light phase and low BCD (groups were similar) during the dark phase (**Fig S2**). This indicates that the luminal environment differences caused by diet consumption alone do not drive differences between experimental groups and that dynamic oscillations of the luminal environment affects the interpretation of dietary changes

Then, we looked at a combination of both diet and feeding pattern differences, using NCD-*ad libitum* to HFD-TRF BCD. Knowing that diet has such a huge effect on the microbiome, we hypothesized that the greatest differences between NCD-*ad libitum* and HFD-TRF would be when they are both eating different diets during the dark phase since both groups would be fasting during the light phase. Thus, we would expect the highest BCD to occur during ZT-17 or ZT-21. Opposite to our hypothesis, we found that the highest BCD values were during the light phase, especially ZT-9 (**Fig 2F**). Despite one group fasting at ZT-13 (HFD-TRF) while the other group eating robustly (NCD-*ad libitum*), we still saw a significant decrease in BCD values when we would have expected them to diverge. Thus, neither feeding/fasting rhythms nor diet drive these temporal fluctuations. In addition, the diurnal pattern of NCD-*ad libitum* to HFD-TRF BCD

fluctuations most closely resembled the comparison between two different diets fed *ad libitum* (**Fig 2D**). The heatmap confirms a similar pattern of NCD-*ad libitum* to HFD-TRF value distribution across timepoints (**Fig 2G**) as NCD-*ad libitum* to HFD-*ad libitum* BCD (**Fig 2E**), with a 2.7-fold difference in peak to nadir BCD.

Thus, while the feeding pattern and diet do appear to significantly influence microbiome composition, their effects are not predictable on a timepoint-by-timepoint basis. If an experimental variable effect as large and reproducible as that imposed by diet is affected by time of day that the sample was collected, then experimental variables with smaller effects - such as medications, metabolites, and genotype - are likely to be even more variable with respect to time.

Gastrointestinal Region Influence Sample Collection Time Results

Though the microbiome of the large and small intestine are quite different (Baker, 1942), the diurnal dynamics of the latter has only recently been actively researched (Tuganbaev et al., 2020). We hypothesized that the dynamic response to changes in diet are not the same between gastrointestinal regions. We pursued this hypothesis by analyzing the results from a previously published study that investigated the diurnal dynamics between different GI regions (Leone et al., 2015). Leone, et al. compared a normal chow diet (NCD) to a high milk-fat diet (MFD) and examined the differences in the microbiome communities of both the cecum and ileum during a 24hr period (**Fig 3A**). The cecum and ileum had significantly different NCD-ad *libitum* to MFD-*ad libitum* BCD at ZT-6, in the middle of the light phase (**Fig 3B**). Thus, while microbial composition was generally similar between the two dietary conditions, there is at least one time point where time of sample collection would have made a difference when comparing dietary responses in the two organs. Heatmaps comparing BCD at different collection times for the ileal samples in this experiment show opposite trends in the timepoints of highest and lowest similarity compared to cecal samples (**Fig 3C, 3D**). While they had opposite trends in the timepoints that had the peak and trough values, the magnitude of change between these values was relatively similar with a

3.5-fold dynamic range in the cecum and 3.8-fold dynamic range in the ileum (**Fig 3C, 3D**). Thus, ileal and cecal diurnal dynamics are not always identical and can, at times, be significantly different. These differences could not have been predicted by experimental design.

Moreover, the study used for **Fig 2** also had ileal samples from the same mice sequenced separately in a recently published study (*manuscript under review*). Similar to the Leone, et al. study, these results generally revealed completely different daily patterns in the ileum (**Fig S3**) than those seen in the cecum (**Fig 2**). The dynamic range of values present in the heatmaps (highest BCD/lowest BCD) is approximately 3.0 in the ileum which is 15% higher than that in the cecum (**Fig 2C, E, G** vs **Fig S3C, E, G**), which was also similar between the two different studies. Thus these reproducible results show that the ileum responds differently over the course of the day than the cecum to the same conditions.

Finally, in a separate study, Wu, et al. investigated the effects of light exposure (i.e. 12h light:12h dark [LD] vs. 24hr dark [DD]) on the jejunal and ileal microbiome of Balb/c mice. Though the observed large BCD differences were particularly at ZT-22 compared to other time points in this experimental condition, the jejunal BCD was fairly consistent across all time points (**Fig S4**). Thus, though sampling time affects the outcomes studies on the ileal microbiome, it does not seem to affect the outcomes of studies in the jejunal samples. Thus, this study further confirms that different GI systems within the same mice have different microbiome dynamics (or perhaps none at all in the jejunum).

Micro-niche site has a critical time window

The previous section demonstrates that different gastrointestinal regions do indeed have unique temporal microbiome patterns based on multiple studies. In this section, we hypothesize that even specific micro-niche sites within a single gastrointestinal region can have unique temporal patterns. To test this hypothesis, we re-analyzed data from a previously published study investigating temporal differences between luminal and mucosal micro-niches in the small

intestine (**Fig 4A**) (Tuganbaev et al., 2020). Tuganbaev, et al. used 8-week-old C57BL/6 mice fed NCD *ad libitum* and collected small intestinal samples every 4 hours for 24hrs, except for ZT-8. The luminal to mucosal BCD was significantly elevated at ZT-20, in the middle of the dark phase. At this time point, BCD increased nearly two-fold compared to the other time points on average (**Fig 4B**). Heatmaps of luminal to mucosal BCD at different collection time combinations show peak divergence during the dark phase and convergence during the light phase with a 3.4-fold difference in BCD (**Fig 4C**). This difference was in the same range as what we reported for the ileum and cecum in the previous section. Furthermore, a log ratio of two common probiotic genera was significantly different from all other other timepoints at ZT20 for both mucosal and luminal samples ($p < 0.05$, paired-Wilcoxon-rank-sum test) (**Fig 4D**). Thus, micro-niche sites can also have unique temporal dynamics present that would not have been predicted based solely on experimental design.

Longitudinal data is also susceptible to the influence of time

Samples from the IHC experiment (**Fig 1**) were collected a week after the experiment started with the intent to characterize the microbiome induced by the environmental exposure, prior to the dysmetabolic phenotype affecting the gut microbiome. However, in the TRF study (**Fig 2**), samples were collected after the phenotype was present. Since many microbiome experiments do not report the rationale for the timing of their sample collection, we questioned whether the length of experimental exposure time affects BCD. We performed a new study to examine where BCD changes over the course of a long study. In this study (**Fig 5A**), the *Ldlr* knock-out ($Ldlr^{-/-}$) mice received either *ad libitum* (control group) or TRF (experimental group) access to the atherogenic diet. This model develops atherosclerotic lesions after 16 weeks. After 1 week ("early"; pre-phenotype development) and 20 weeks ("late"; post-phenotype development), we collected stool every 4 hours for 24hrs to examine dynamic changes in time point composition over the course of a long term experiment.

As shown in other studies, the time of sample collection during the day affects *ad libitum* to TRF BCD distances. During both the early and late phase of the experiment, maximum divergence *ad libitum* to TRF occurred during the dark period (highest mean BCD = ZT-20; **Fig 5B**). The BCD patterns conformed to a circadian-like pattern ($p < 0.05$, MetaCycle, JTK method) during both the early and late collection, with nearly identical amplitude and minor shifts in period and phase (**Fig S5**). Furthermore, the *ad libitum* to TRF BCD was not significantly different between the early and late part of the study at any time point (**Fig 5B**), demonstrating consistency within the study over time. The peak to trough ratios were also nearly identical between the early (**Fig 5C**) and late collection (**Fig 5D**) - 2.1-fold and 2.2-fold dynamic range, respectively. On the other hand, the specific time point combinations that resulted in these peak and trough values differed slightly, indicating potentially a subtle shift over time that might eventually become significant if a long enough study was to be performed. In general, these results demonstrate that longitudinal measures of BCD in a non-continuous intervention within a single experiment are relatively consistent over time.

To investigate the effects of longitudinal exposure to a daily discrete external intervention, we re-analyzed previously published data from our lab. In a previously published cohort of mice in an experiment investigating changes in the microbiome in response to intermittent-hypoxia-hypercapnia conditions (similar to **Fig 1A**) over several weeks until phenotype development (Tripathi et al., 2019) (**Fig 5E**). In this cohort, samples were collected once per day, during ZT-3 to ZT-5 (i.e. the time of greatest divergence), twice weekly over 10 weeks. All samples collected, stored, and processed in this study used the exact same analysis pipelines as **Fig 1**. The aim of this re-analysis is to determine whether the dynamic changes in composition during the course of a longitudinal experiment with time of sample collection held constant. While the control of IHC BCD fluctuated significantly during the course of the experiment, a general step-wise trend (increase, then plateauing, repeat) was seen (**Fig 5F**). The groups did diverge with significantly

increased BCD over time (week 10.5 compared to week 19.5, $p = 2.56x10^{-8}$, paired Wilcoxon rank sum, test statistic 1126) as the phenotype developed. By holding the time of collection constant, we observed a compositional shift that occurred over time as the phenotype developed that might not have been observed at a different time.

To determine if BCD is also relevant in longitudinal human studies, we re-analyzed a study that investigated the effects of a four day longitudinal dietary change (i.e. plant to animal based diet) in adult subjects on the speed and extent of shifts in the gut microbiome (David et al., 2014). When BCD was similarly calculated using weighted UniFrac, the plant-to-animal diet BCD demonstrated that the two groups did diverge the most on day 4 on condition (**Fig S6**). Interestingly, this study clearly demonstrates that BCD captures how time and diet overcame the known large effect of inter-individual variability in human subjects. However, humans defecate on average once a day, and, thus, with not enough frequency for analysis of changes to investigate dynamics shifts as we have done in mice. There have been attempts to reconstruct daily microbiome diurnal rhythms using several thousand human samples (Reitmeier et al., 2020), which have also shown diurnal pattern disruption in a disease state. Thus, time of sample collection is likely relevant in human samples as well.

The fecal metabolome BCD is not as dynamic as the fecal microbiome BCD

To determine whether cyclical fluctuations in the composition of the gut metabolome affect functional measures, we examined the untargeted liquid chromatography-tandem mass spectrometry (LC-MS/MS) from the same samples from the IHC experiment in atherosclerosis-prone mice discussed in **Fig 1A** (Piedrahita et al., 1992). We used non-phylogenetic **β**-diversity metric, canberra, for metabolomics to assess BCD as published in the source paper.

Unlike the 16S microbiome assessments (**Fig 1B, S7A-B**), there were minimal shifts in BCD between the IHC and control groups over the course of the day (**Fig 6A,Fig S7C-D**). The BCD heatmap matrix demonstrates a peak (Air ZT10, IHC ZT2) and a trough (Air ZT2, IHC ZT22)

that is very different (**Fig 6B**) from what we determined from 16S measures (**Fig 1C**). We examined selected metabolites that are potentially involved in atherosclerosis formation at these peak and trough time points (**Fig 6C**). Time point variation appears to not play as significant a role in fecal metabolite composition at the biologically relevant group (left) or subgroup (right) level. We also measured longitudinal untargeted LC-MS/MS metabolomics for the experiments discussed in **Fig 4E-F** for four time points. The longitudinal data revealed no significant change in BCD from early to late metabolomics measures (Wilcoxon signed-rank test; test statistic W=3685, p=0.11) (**Fig 6D**). Thus, though there may be diurnal oscillation of specific metabolites, global assessments of the metabolome appear less dynamic than the microbiome based on BCD, both within a day and over the course of an experiment. Moreover, based on this limited sample set, the effect of the diurnal rhythm does not appear to supersede experimental variables as seen in the 16S microbiome data. However, diurnal metabolomics data only existed for one study, and additional studies are needed to assess whether this is universally true across most experimental conditions.

4.3 Discussion

One of the greatest barriers to the development of microbiome-mediated therapeutic agents remains the lack of replicability in microbiome research. Though several confounding factors that influence 16S microbiome composition results have been described (e.g. cage and maternal effects, primers used for amplicon sequencing, sample collection techniques), there is still significant inter- and intra-individual variability in microbiome research (Gilbert et al., 2018). Many studies address this by increasing the number of subjects to improve the signal-to-noise ratio (Goodrich et al., 2014). However, increasing the number of subjects makes the cost of high-quality microbiome studies outside of the reach of most resource-limited labs, nor would it guarantee reproducibility in light of the findings presented in this study. Hence, investigating the

factors that affect microbiome variability and improve replicability remains an important area of research as the field moves toward a more mechanistic understanding of how the microbiome affects host physiology and disease progression.

Since 2014 there has been unequivocal and reproducible research from multiple labs demonstrating diurnal fluctuations in the composition of the gut microbiome (Alvarez et al., 2020; Frazier and Chang, 2020; Leone et al., 2015; Liang et al., 2015; Thaiss et al., 2014, 2016; Uhr et al., 2019; Voigt et al., 2014; Zarrinpar et al., 2014). Yet neither sample collection time nor the rationale for the selection of this time is reported outside of studies that are focused on diurnal fluctuations of the microbiome. Here, we show that the conclusions of a microbiome research study are greatly dependent on the time of sample collection, and that experimental and control groups undergo a cycle of diverging and converging microbiome composition depending on the nature and timing of experimental interventions. Although we hypothesized that BCD would decrease when the two conditions have the greatest similarity in environmental exposure, we found that this was inconsistent between experiments and highly dependent on the experimental exposure. Moreover, BCD, a measure of microbiome dissimilarity, was highly dependent on the gastrointestinal region sampled (**Fig 2**) and even luminal micro-niche (**Fig 3**), demonstrating that these factors need to be under consideration as well. Surprisingly, in some studies, samples collected just a few hours apart could completely change the conclusions of the study and lead to nearly diametrically opposite results (**Fig 2**). In some experiments, BCD continued to increase over weeks of exposure, whereas in others, longitudinal exposure did not affect these distance metrics. Interestingly, the BCD of the metabolome appears unaffected by sample time collection both within a day and longitudinally during an experiment. These results demonstrate that the conclusions of microbiome compositional studies are highly dependent on sample time collection, and thus experiments are difficult to replicate without this critical piece of information. More

functional information, such as metabolomics, may not be subject to the same change in dynamics observed with 16S sequencing.

Moreover, our findings suggest a fluidity of composition that is sensitive to a variety of host factors including environmental exposures, diet, gut region, and luminal micro-niche. Our BCD analysis confirms that, in some experiments, peak and trough distances can be as short as four hours apart (**Fig 2**). That is, shifting the collection of one condition by four hours could yield dramatically and potentially opposite conclusions on the similarity of the microbiome from experimental and control groups. This time scale may still be an overestimate; we did not collect stool samples at less than four-hour intervals. Thus, conflicting results from different laboratories may be due to differences in phase of the circadian cycle at the time of collection, timing relative to the experimental intervention, investigator chronotype (e.g. morning lark vs. night owl), or vivarium lighting setup. In studies with discrete daily interventions such as those described in this study, these differences can be quite pronounced.

Based on our literature review analysis, since the vast majority (>90%) of microbiome studies do not report when samples are collected, laboratories may unknowingly be collecting at suboptimal time points. Furthermore, although it is likely a good assumption, due to convention and best utilization of researcher time, the methods section of published papers does not confirm that the control and experimental conditions are collected at the same time or within a specific window. In addition, while **Fig 1C** and **Fig 1D** suggest that optimization of collection time points could be accomplished by sampling at the time of highest beta diversity for each group, caution should be taken not to artificially influence results. It would also be prudent to establish standard collection times for experiments in a field to ensure replication. To improve replicability, investigators should provide an explanation for the collection time of samples as it relates to their scientific hypothesis with the knowledge that anticipatory changes in the microbiome are quite pronounced.

The composition of the microbiome appears less tightly constrained and more dynamic than the metabolome. Though there is a pervasive assumption that compositional shifts in the microbiome imply a functional shift in their activity, our analysis of metabolomics across multiple studies did not find a divergence and convergence of BCD that parallels the 16S results. This finding is consistent with the conclusions from the Human Microbiome Project (The Human Microbiome Project Consortium, 2012) showing great inter- and intra-individual variability in 16S rRNA gene amplicon sequencing, but less variation among subjects with the metagenomic functional repertoire. Though many secondary metabolites, such as bile acids, have diurnal fluctuations (Zhang et al., 2011), these may not be pronounced enough to affect our metabolomics beta-diversity measures. Our findings that the metabolome, as a whole, is not as dynamic as the microbiome suggests the possibility that circadian clock entrainment might occur through microbiota or its components rather than through general changes in metabolites, thereby further emphasizing the importance of increased sampling frequency and time of sample collection. However, we cannot rule out that specific secondary metabolites, such as bile acids or short-chain fatty acids, may be responsible for entraining host peripheral rhythms.

Though there is good evidence that the human gut microbiome has diurnal fluctuations (Brooks and Hooper, 2020; Sailani et al., 2020), the relevance of these compositional changes in a subject that, unlike the murine models, usually only defecates once a day is not clear. The microbiome compositional shifts that occur within a 24-hour period may be more a reflection of the length of fecal storage in the rectal vault rather than a relationship with circadian machinery. Nevertheless, our secondary analysis of previously published work demonstrates that even in human studies BCD continues to increase over the length of experimental exposure (**Fig S6**). Mathematical modeling of the human microbiome provides preliminary evidence of temporal dynamics for most genera (Kenney et al., 2020). Microbiome oscillations, entrained by feeding as well as light-dark signals, also impact host metabolism and immunity (Brooks and Hooper, 2020).

Changes to the microbiome contribute to metabolic diseases such as obesity, diabetes, and NAFLD (Musso et al., 2011). Thus, the time of sample collection may also reveal important information about the role of microbiota in different types of diseases in patient cohorts as well.

Outside of studies focused on circadian rhythms, microbiome studies do not commonly report time of sample collection in their methods nor confirm that control and experimental samples are collected at the same time of day. While several of the studies suffer from a low sample number, the fact that findings are replicated in laboratories from several different institutions with related study designs indicates how understudied this phenomenon is. Because of the rarity of circadian rhythm studies, it is likely that additional studies will need to be performed to determine optimal sampling times to standardize each field that uses a host with a circadian rhythm. We also propose that sample collection time be reported in ZT notation in future studies going forward. Otherwise, if we are not controlling for host circadian rhythm time, it is like trying to measure sea level rise while not knowing that tides or waves exist.

4.4 Methods

Literature Review: (**Fig S1**) We used the advanced search option from the four main journal groups, including the American Society for Microbiology (ASM) (https://msystems.asm.org), Science (https://search.sciencemag.org), Nature (https://www.nature.com), and Cell Press (https://www.cell.com). Searching for the term "microbiome" in all search fields (abstract, title, main text) during the year 2019 (Jan 1, 2019, to Dec 31, 2019) resulted in 586 articlesfrom 9 journals; mSystems (ASM), Science Translational Medicine (Science), Science Signaling (Science), Science Advances (Science), Science Immunology (Science), Nature (Nature), Nature Microbiology (Nature), Nature Communications (Nature), Cell Host Microbe (Cell), Cell (Cell), Cell Reports (Cell), Cell Metabolism (Cell). Our collection sheet includes a total of 16 columns: journal group, journal, year, article title, DOI, PMID,

first author, last author, Microbiome (yes/no), vivarium (yes/no), vivarium setting, sample host, sample type, collection time, time note, and collection time reason. Notation of collection time was recorded as follows: explicitly stated (“yes”; 8AM, ZT4, etc.), implicitly stated (“relative”; “before surgery”, “in the morning”, etc.), or unstated (“not provided”; “daily”, “once a week”, etc.).

Microbiome: All of the data in this paper is a re-analysis of previously published 16S studies. Please refer to the respective source papers for detailed methods, including sample handling and preliminary processing. Raw data was procured from the respective data repositories as stated in the source paper, typically the European Nucleotide Archive (ENA). This data was then run through a standard QIIME2 pipeline (version 2021.8) (Bolyen et al., 2019) as follows: samples demultiplexed, denoised via deblur (Amir et al., 2017) into the amplicon sequence variant (ASV) table, feature table underwent rarefaction (as stated in source paper, see individual methods sections), representative sequences underwent fragment insertion on Greengenes_13_8 via SATé-enabled phylogenetic placement (Mirarab et al., 2011) to create the phylogenetic tree, and weighted UniFrac distances (Lozupone et al., 2011) were calculated. The resulting weighted unifrac distance matrix was filtered for only between-condition distances (BCD) as relevant to each study. Thus, using BCD values will show how similar the microbiomes from the two conditions are to each other at any given time point. Since BCD values are a subset of the Weighted UniFrac distance matrix values, both conditions (control and experimental) are taken into account with each distance value shown. Changes in BCD will demonstrate convergence (decreasing distance, increased similarity) or divergence (increasing distance, increased dissimilarity) of the microbiome composition between two groups. Circadian time notation is used throughout the paper to denote when samples were collected: Zeitgeber Time (ZT) were lights on = ZT-0 . Data was visualized using custom python scripts, which can be found at https://github.com/knightlab-analyses/dynamics.

Figure 4.1 Briefly, two groups of ten-week-old male *Apoe*$^{-/-}$ mice on C57BL/6J background (002052; The Jackson Laboratory, Bar Harbor, ME) were individually housed in a 12-hour light:12-hour dark (12:12 L:D) vivarium. All mice were given an atherosclerotic-promoting diet (1.25% cholesterol, 21% milk fat; 4.5 Kcal/g; TD.96121; Envigo-Teklad Madison, WI) starting at 10 weeks of age until the end of the study. Mice in the experimental group were exposed to intermittent hypoxia and hypercapnia (IHC) conditions that consisted of 4 min of synchronized O_2 reduction from 21% to 8% and synchronized elevation of CO_2 from 0.5% to 8%, followed by alternating periods of 4 min of normoxia and normocapnia with 1- to 2-min ramp intervals. IHC conditions were administered in a computer-controlled atmosphere chamber (OxyCycler, Reming Bioinstruments, Redfield, NY) for 10 hours per day during the lights on phase (ZT-2 to ZT-12) when mice are sleeping for 10 weeks. Mice in the control group were exposed to normal room air (21% O_2 and 0.5% CO_2) during that same time period. After 6 days, fecal samples were collected every 4 hours for 24hrs (n = 4/group). 16S amplicon sequencing was performed on the V4 region using standard protocols (http://www. earthmicrobiome.org/emp-standard-protocols/). Rarefaction was set at 12,000 reads to control for sequencing effort. Please see the source paper for additional details (Allaband et al., 2021; Piedrahita et al., 1992).

Figure 4.1. Microbiome Analysis of Apoe-/- Mice Exposed to IHC Show Vastly Different Outcomes Depending on Time Point of Sample Collection A) Experimental design. IHC=intermittent-hypoxia-hypercapnia. B) between-condition distances (BCD) weighted UniFrac β-diversity distances. Significance is determined using paired Wilcoxon rank-sum test. The BCD values in this experiment were oscillating in a diurnal fashion (MetaCycle, JTK method, $p<0.001$). C) BCD heatmap by time point. Highest highlighted in green, lowest highlighted in orange. At the peak and trough time points identified in C, (D) the logarithmic ratios of differentially abundant key phyla of interest and (E) the logarithmic ratios of differentially abundant key families of interest. Notation: ns = not significant, * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$

Figure 1

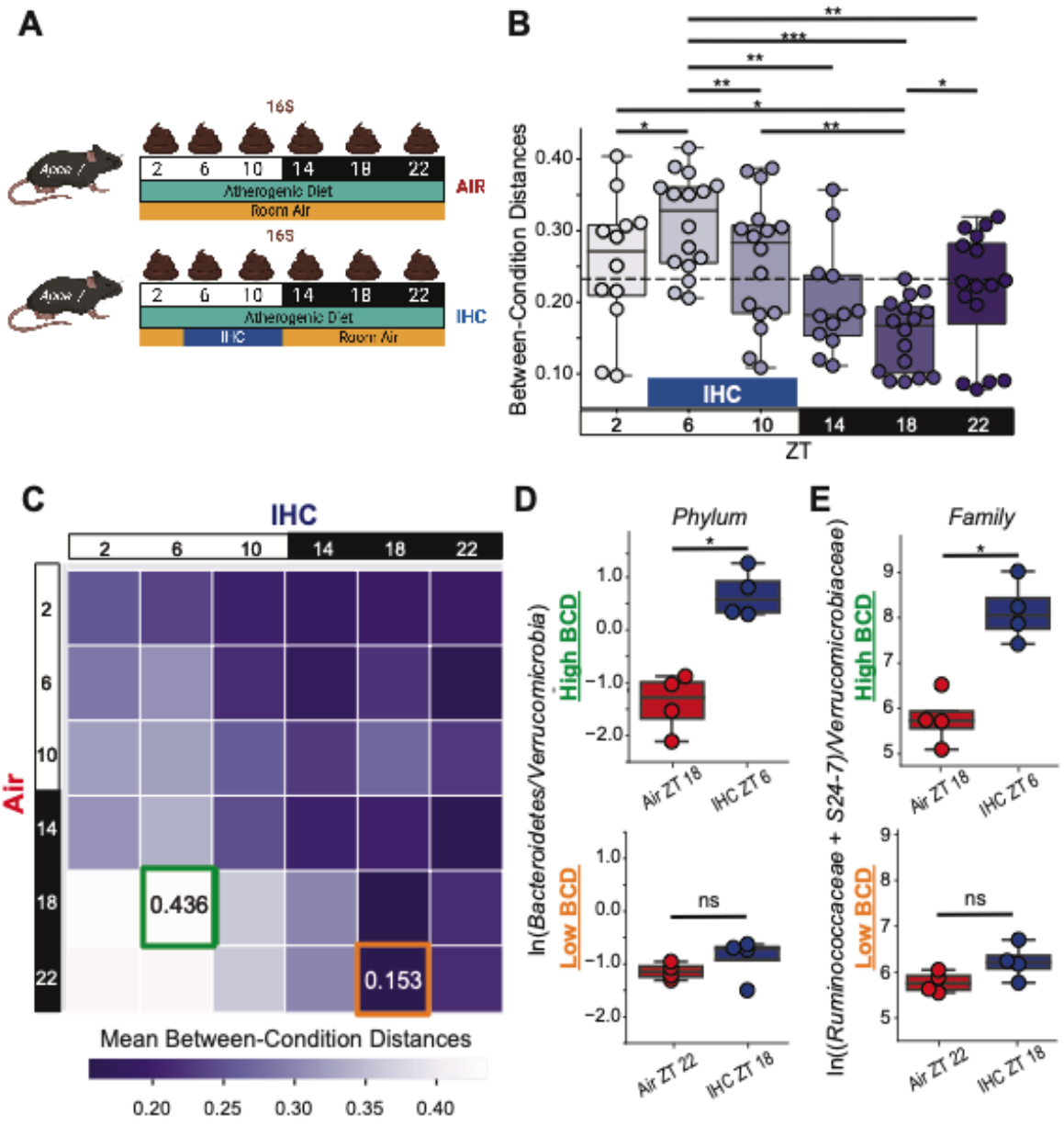

A
16S
2 6 10 14 18 22
Atherogenic Diet
Room Air
AIR
16S
2 6 10 14 18 22
Atherogenic Diet
IHC
Room Air
IHC
B
Between-Condition Distances
0.40
0.30
0.20
0.10
IHC
2 6 10 14 18 22
ZT
C
IHC
2 6 10 14 18 22
Air
2
6
10
14
18
22
0.436
0.153
Mean Between-Condition Distances
0.20 0.25 0.30 0.35 0.40
D
Phylum
ln(Bacteroidetes/Verrucomicrobia)
High BCD
Low BCD
Air ZT 18
IHC ZT 6
Air ZT 22
IHC ZT 18
ns
E
Family
ln((Ruminococcaceae + S24-7)/Verrucomicrobiaceae)
High BCD
Low BCD
Air ZT 18
IHC ZT 6
Air ZT 22
IHC ZT 18
ns

Figure 4.2, 4.S3 - In short, wild-type SPF C57Bl/6 group-housed male mice (3 mice per cage) were provided either normal chow diet (LabDiet 5001, 13.5% calories from fat, crude fiber 5.1%) or a high fat diet (61% fat, HFD) and were fed in either an *ad libitum* manner, with access to food at all times, or fed in time-restricted (TRF) manner. TRF mice were allowed unrestricted access to HFD from ZT-13 to ZT-21. Mice on an NCD *ad libitum* diet (controls) typically fast during the light phase and consume >80% of their diet during the dark phase (Hatori et al., 2012; Kohsaka et al., 2007). However, mice on a HFD *ad libitum* diet (diet-induced obesity) lose this diurnal feeding pattern and spread their caloric intake throughout both the dark and light phase (Hatori et al., 2012; Kohsaka et al., 2007). TRF of HFD consolidates feeding to the nocturnal period by providing access to food in a narrow time window, from ZT-13 to ZT-21 in this experiment, and is known to prevent the dysmetabolic effects of HFD consumption (Chaix et al., 2014; Hatori et al., 2012; Zarrinpar et al., 2014). After 8 weeks under these dietary conditions, mice were euthanized every 4 hours for 24hrs and intestinal contents collected (n=3 mice/condition/time point from separate cages; 6 time points). At ZT-13, fasted mice were euthanized prior to feeding. 16S amplicon sequencing was performed on the V1-V3 region using the 454 platform for cecal data. 16S amplicon sequencing was performed on the V4 region using Illumina primers for ileal data. For both regions, rarefaction was set to 1,000 reads to control for sequencing effort. Please see source paper for additional details (Zarrinpar et al., 2014; manuscript under review).

Figure 4.2. Diet and Feeding Pattern Influence Sample Collection Time Results in the Cecum. A) Experimental design. B) BCD for cecal samples comparing HFD ad libitum vs HFD TRF. The dotted line is the average of all shown weighted UniFrac distances. Significance was determined using a paired Mann-Whitney-Wilcoxon test two-sided with Bonferroni correction.C) Heatmap of mean cecal BCD between HFD ad libitum and HFD TRF mice by time point. Highest highlighted in indigo, lowest highlighted in yellow. D) BCD for cecal samples comparing NCD ad libitum vs HFD ad libitum. The dotted line is the average of all shown weighted UniFrac distances. Significance was determined using the Mann-Whitney-Wilcoxon test two-sided with Bonferroni correction.E) Heatmap of mean cecal BCD between NCD ad libitum controls and HFD TRF mice by time point. Highest highlighted in indigo, lowest highlighted in yellow. F) BCD for cecal samples comparing NCD ad libitum vs HFD TRF. The dotted line is the average of all shown weighted UniFrac distances. Significance was determined using the Mann-Whitney-Wilcoxon test two-sided with Bonferroni correction.G) Heatmap of mean cecal BCD between NCD ad libitum controls and HFD TRF mice by time point. Highest highlighted in indigo, lowest highlighted in yellow. Notation: * = $p<0.05$; ** = $p<0.01$; *** = $p<0.001$.

Figure 2

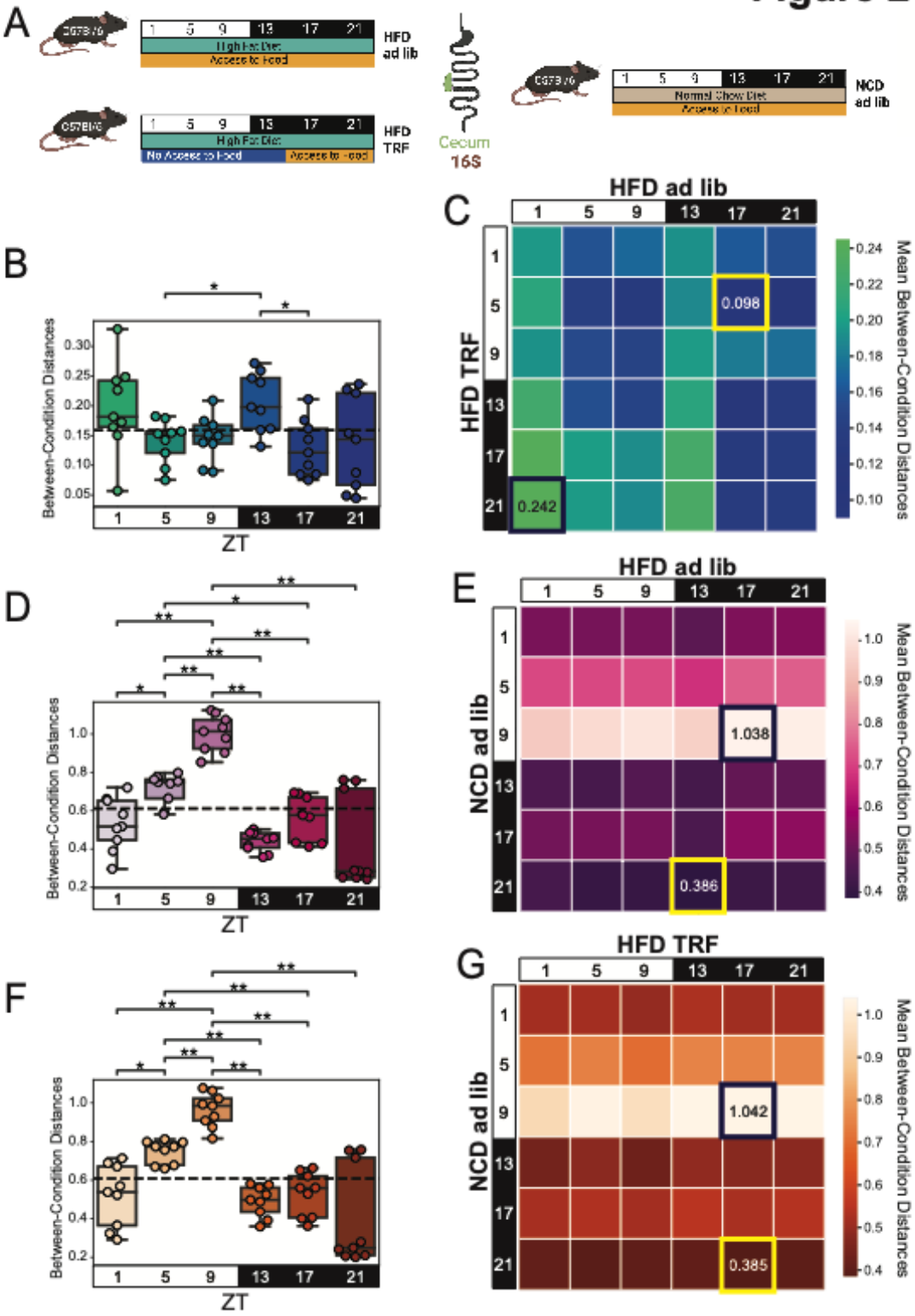

A
HFD ad lib
HFD TRF
Cecum
16S
NCD ad lib
B
Between-Condition Distances
ZT
C
HFD ad lib
HFD TRF
0.098
0.242
Mean Between-Condition Distances
D
Between-Condition Distances
ZT
E
HFD ad lib
NCD ad lib
1.038
0.386
Mean Between-Condition Distances
F
Between-Condition Distances
ZT
G
HFD TRF
NCD ad lib
1.042
0.385
Mean Between-Condition Distances

Figure 4.3 - The study was performed on 8 to 10 week old male C57Bl/6J SPF mice that were maintained in a 12:12 L:D cycle vivarium. The mice were fed *ad libitum* with either a normal chow diet (NCD, Harlan Teklad 2018S, 18% calories from fat, 3.5% crude fiber) or a 37.5% saturated milk fat diet (MFD, Harlan Teklad TD.97222 customized diet). After 5 weeks of being on the NCD or MFD diet, the mice were sacrificed and the cecal and ileal contents harvested every 4 hours for 24 hours (n = 3 mice/treatment). The cecal and ileal contents were flash frozen and stored at -80ºC. 16S amplicon sequencing was performed on the V4-V5 region using standard protocols (https://earthmicrobiome.org/protocols-and-standards/) in a High-Throughput Genome Analysis Core (Institute for Genomics & Systems Biology) at Argonne National Laboratory. Rarefaction was set at 10,000 reads to control for sequencing effort. Please see the source paper for additional details (Leone et al., 2015).

Figure 4.4 - This study was performed on 8 to 12-week old WT C57BL/6 mice that were maintained in a 12:12 L:D cycle vivarium. The mice were fed a normal chow diet (NCD, Harlan Teklad 2018S, 18% calories from fat, 3.5% crude fiber) *ad libitum* for 4 weeks prior to sample collection. The mice were sacrificed and the luminal and mucosal small intestinal samples were collected every 4 hours for 24 hours (except for ZT-8, n = 4-5 mice/time point). The samples were frozen and stored at -80ºC. 16S amplicon sequencing was performed on the V4 region of the genome. Rarefaction was set to 4,200 reads to control sequencing effort. Please see the source paper for additional details (Tuganbaev et al., 2020).

Figure 4.3. Gastrointestinal Region Influence Sample Collection Time Results
A) Experimental design. Mice were fed ad libitum with either NCD or MFD. After 5 weeks, cecal and ileal samples were collected every 4 hours for 24 hours (N=3 mice/condition). B) BCD for both ileal and cecal samples comparing NCD vs MFD. The dotted line is the average of all shown weighted UniFrac distances. Ileal vs Cecal pairwise significance was determined using Mann-Whitney-Wilcoxon test two-sided with Bonferroni correction.C) Heatmap of mean BCD from cecal samples collected from NCD controls and HFD ad libitum mice by time point. Highest highlighted in black, lowest highlighted in yellow. D) Heatmap of mean BCD from ileal samples collected from NCD controls and HFD TRF mice by time point. Highest highlighted in black, lowest highlighted in yellow. Notation: * = $p<0.05$; ** = $p<0.01$; *** = $p<0.001$.

Figure 3

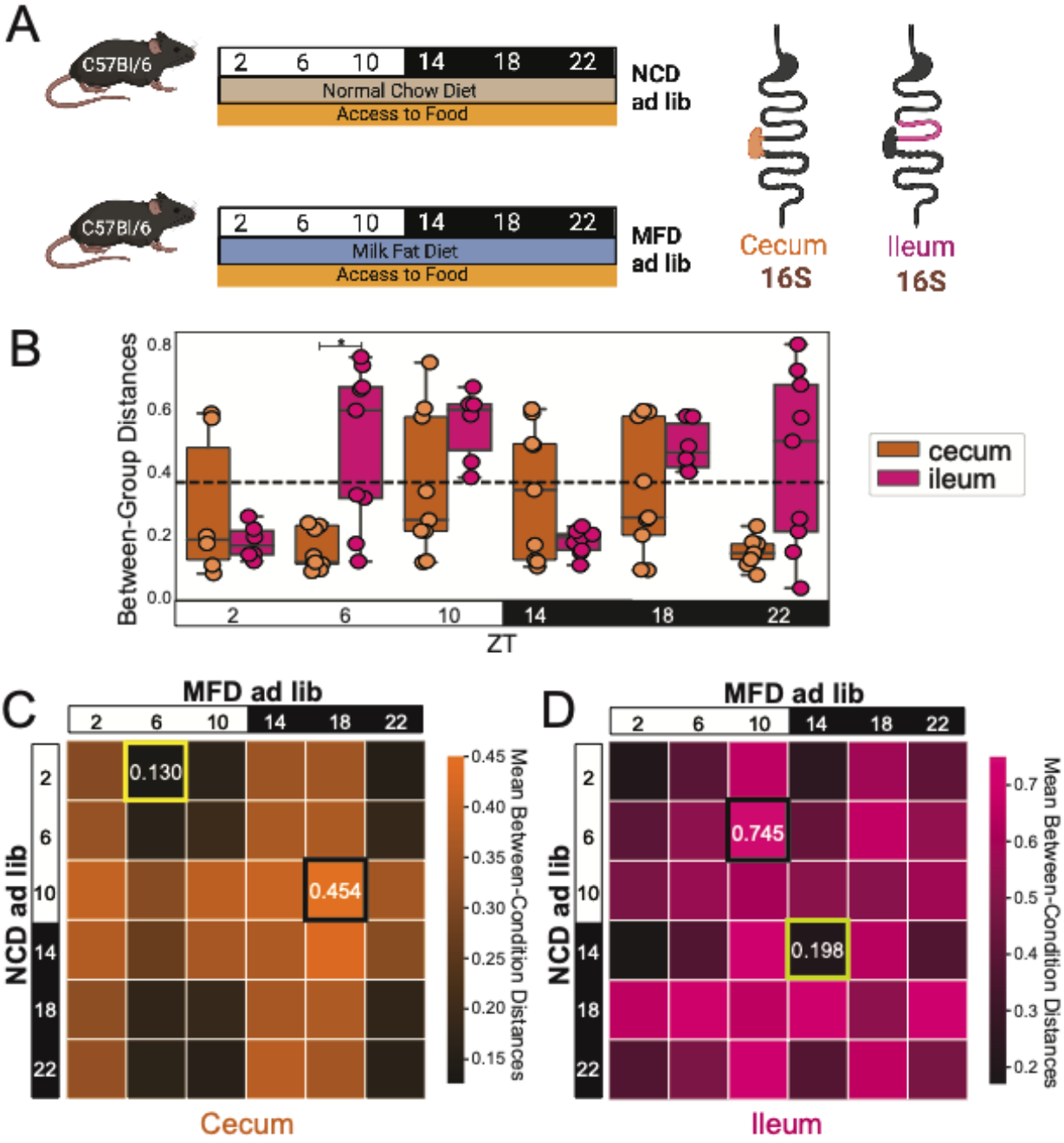

A
C57Bl/6
2 6 10 14 18 22
Normal Chow Diet
Access to Food
NCD
ad lib
C57Bl/6
2 6 10 14 18 22
Milk Fat Diet
Access to Food
MFD
ad lib
Cecum
16S
Ileum
16S
B
Between-Group Distances
0.0 0.2 0.4 0.6 0.8
2 6 10 14 18 22
ZT
cecum
ileum
C
MFD ad lib
2 6 10 14 18 22
NCD ad lib
2 6 10 14 18 22
0.130
0.454
Mean Between-Condition Distances
0.45 0.40 0.35 0.30 0.25 0.20 0.15
Cecum
D
MFD ad lib
2 6 10 14 18 22
NCD ad lib
2 6 10 14 18 22
0.745
0.198
Mean Between-Condition Distances
0.7 0.6 0.5 0.4 0.3 0.2
Ileum

Figure 4.4. Localized changes in BCD between luminal and mucosal contents. A) Experimental design and sample collection for a local site study. Small intestinal samples were collected every 4 hours for 24 hours (N=4-5 mice/condition, skipping ZT8). Mice were fed ad libitum on the same diet (NCD) for 4 weeks before samples were taken. B) BCD for luminal vs mucosal conditions. The dotted line is the average of all shown weighted UniFrac distances. Significance is determined using the Mann-Whitney-Wilcoxon test two-sided with Bonferroni correction. C) Heatmap of mean BCD distances comparing luminal and mucosal by time point. Highest value highlighted in navy, lowest value highlighted in gold. D) Experimentally relevant log ratio, highlighting the changes seen at ZT20.

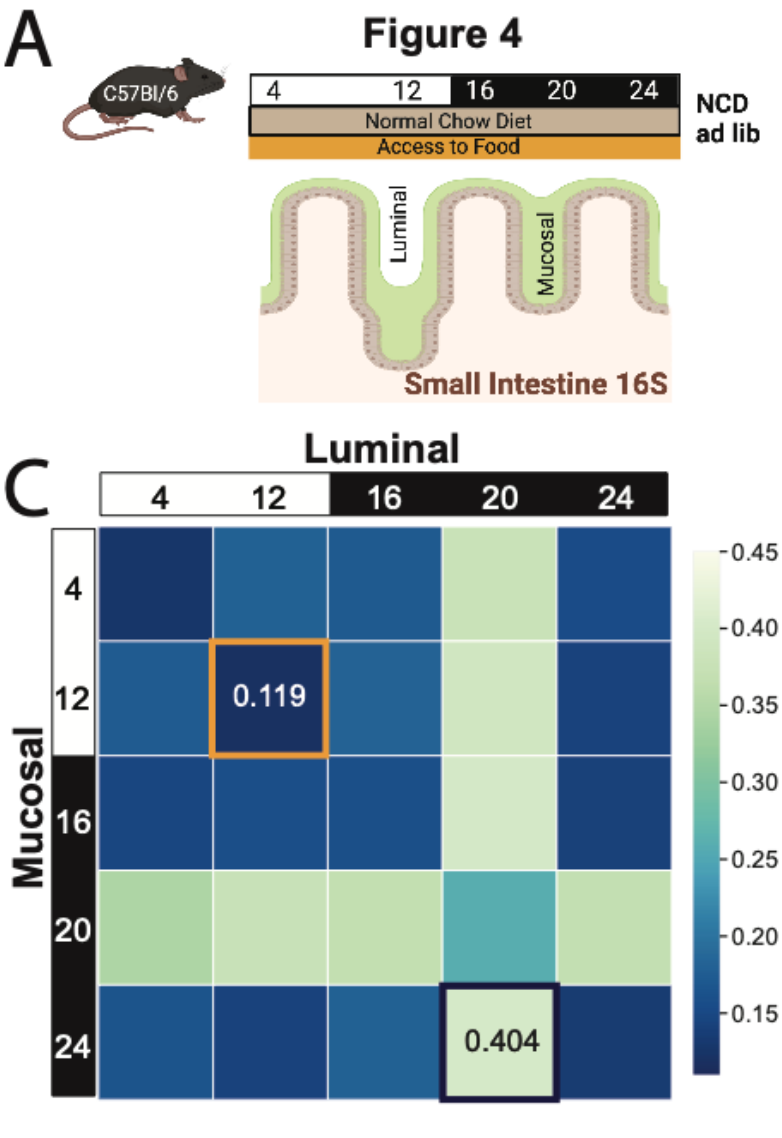

A
C57Bl/6
4
12
16
20
24
Normal Chow Diet
Access to Food
NCD
ad lib
Luminal
Mucosal
Small Intestine 16S
B
Between Group Distances
0.4
0.3
0.2
0.1
****
****
****
**
4
12
16
20
24
ZT
C
Luminal
Mucosal
0.119
0.404
Mean Between-Condition Distances
0.45
0.40
0.35
0.30
0.25
0.20
0.15
D
ln(Lactobacillus/Bifidobacterium)
Luminal
Mucosal
ZT

Figure 4

Figure 4.5A-D, 4.S5 - This study was performed on 10 week old *Ldlr*$^{-/-}$ mice (Jackson Labs) which were fed a high fat, high cholesterol diet (Research Diets D12109i; Clinton/Cybulsky high-fat rodent diet, regular casein, 1.25% added cholesterol, 0.5% sodium cholate). During the experiment, mice were maintained in 12:12 L:D reverse light-cycled cabinets (Phenome Technologies). Mice were either fed *ad libitum* or had feeding restricted to a 9-hour window from ZT-14 to ZT-23. There were two 24hr fecal sample collections (Q4 hrs, 6 time points), after 1 week and 20 weeks of treatment (n = 6 mice/treatment). 16S rRNA was performed on the V4-V5 region using the Earth Microbiome standard protocol (https://earthmicrobiome.org/protocols-and-standards/). Rarefaction was set at 11,498 reads to control for sequencing effort.

Figure 4.5E-F - In brief, two groups of ten-week-old male *Apoe*$^{-/-}$ mice on C57BL/6J background (002052; The Jackson Laboratory, Bar Harbor, ME) were kept in a 12:12 L:D vivarium fed a normal chow diet (Teklad Rodent Diet 8604, 14% calories from fat, 4% crude fiber) before they were switched to an atherosclerotic-promoting diet containing 1.25% cholesterol and 21% milkfat (4.5 Kcal/g; TD.96121; Envigo-Teklad Madison, WI) starting at 10 weeks of age until the end of the study. Mice in the experimental group were exposed to IHC conditions as described in Fig 1 and were administered in a computer-controlled atmosphere chamber (OxyCycler, Reming Bioinstruments, Redfield, NY) for 10 hours per day during the lights on phase (ZT2-ZT12) for 10 weeks. Mice in the control group were exposed to normal room air (21% O_2 and 0.5% CO_2) during that same time period. Fecal samples were collected twice a week for the duration of the study (Tripathi et al., 2019).

Figure 4.5. Longitudinal changes in BCD over the course of a study. A) Experimental design and sample collection for TRF study. Fecal samples were collected every 4 hours for 24 hours (n=6 mice/condition) after 1 week (early; pre-phenotype) and after 20 weeks (late; post-phenotype). The two conditions were both on an atherogenic diet (AD), but one group was fed ad libitum and the other TRF. In TRF, mice were allowed to eat only for 8 hours per day during the dark/active phase of the circadian rhythm between ZT-13 and ZT-21. B) BCD for ad libitum vs TRF conditions at the early (Week 1) and late (Week 20) timepoints. Dotted line is the average of all of the weighted UniFrac distances. Significance is determined using paired Wilcoxon rank-sum test. C) BCD heatmap for early samples, and D) BCD heatmap for late samples. Highest value is highlighted in tan and the lowest value is highlighted in yellow. E) Experimental design and sample collection for longitudinal IHC study. During the 10 weeks of exposure to either normal room air or IHC conditions, samples were collected between ZT-3 and ZT-5 every 3-4 days for the duration of the study (n=12 mice/condition).. F) BCD over the course of the IHC longitudinal study. Dotted line is the mean of all data shown. The only comparison shown is between Age 10.5 weeks and 19.5 weeks; significance was determined using paired Wilcoxon rank-sum tests. Notation: * = $p<0.05$; ** = $p<0.01$; *** = $p<0.001$, ***** = $p<0.00001$.

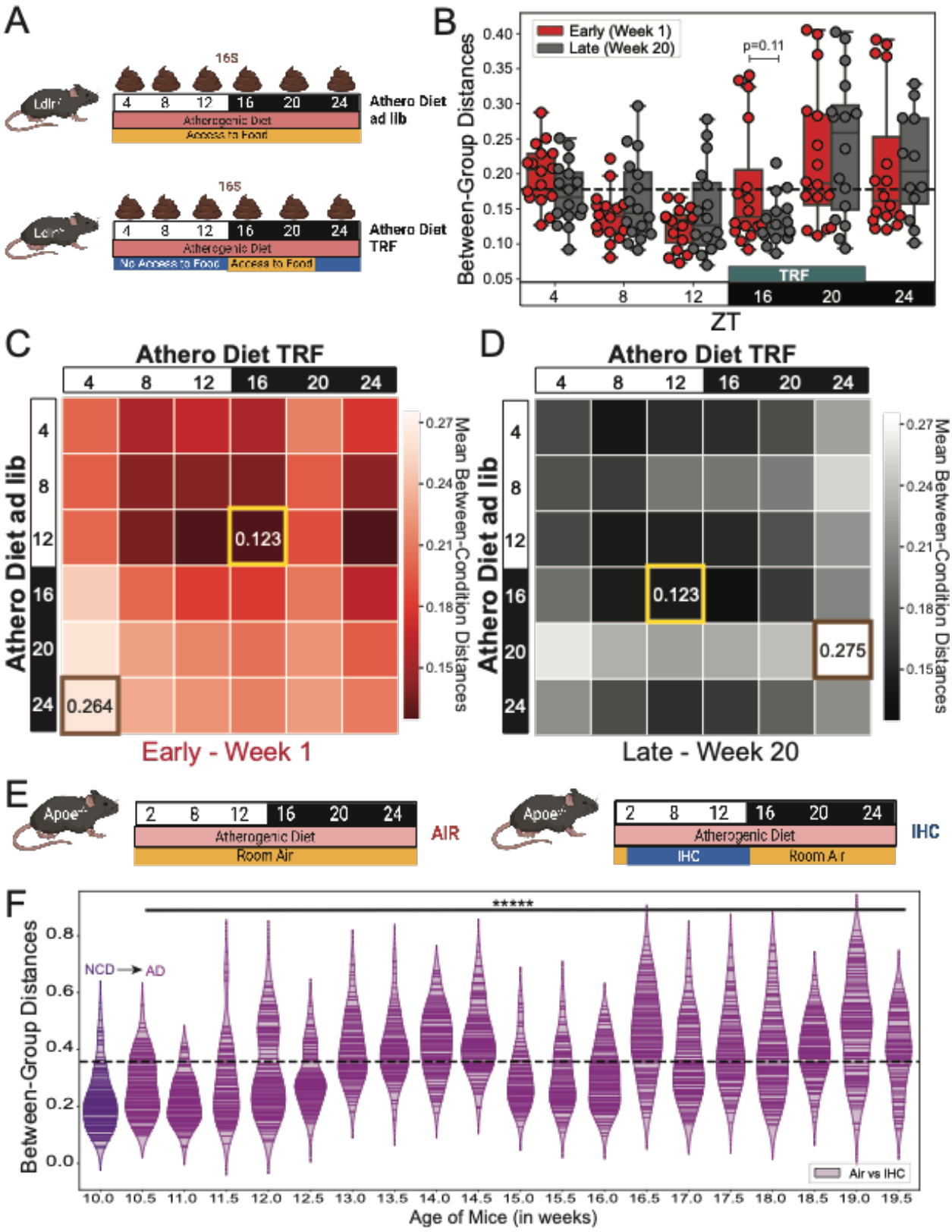

A
16S
4 8 12 16 20 24
Atherogenic Diet
Access to Food
Athero Diet ad lib
16S
4 8 12 16 20 24
Atherogenic Diet
No Access to Food
Access to Food
Athero Diet TRF
B
Early (Week 1)
Late (Week 20)
p=0.11
Between-Group Distances
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
TRF
4 8 12 16 20 24
ZT
C
Athero Diet TRF
4 8 12 16 20 24
Athero Diet ad lib
4 8 12 16 20 24
0.123
0.264
Mean Between-Condition Distances
0.27
0.24
0.21
0.18
0.15
Early - Week 1
D
Athero Diet TRF
4 8 12 16 20 24
Athero Diet ad lib
4 8 12 16 20 24
0.123
0.275
Mean Between-Condition Distances
0.27
0.24
0.21
0.18
0.15
Late - Week 20
E
2 8 12 16 20 24
Atherogenic Diet
Room Air
AIR
2 8 12 16 20 24
Atherogenic Diet
IHC
Room Air
IHC
F
*****
Between-Group Distances
0.8
0.6
0.4
0.2
0.0
NCD → AD
Air vs IHC
10.0 10.5 11.0 11.5 12.0 12.5 13.0 13.5 14.0 14.5 15.0 15.5 16.0 16.5 17.0 17.5 18.0 18.5 19.0 19.5
Age of Mice (in weeks)

Figure 4.S2 - The study was performed on 8-10 week old male C57Bl/6J SPF mice that were maintained in a 12:12 L:D cycle vivarium. The mice were fed *ad libitum* with either a normal chow diet (NCD, Harlan Teklad 2018S, 18% calories from fat, 3.5% crude fiber) or a 37.5% saturated milk fat diet (MFD, Harlan Teklad TD.97222 customized diet). After 5 weeks of being on the NCD or MFD diet, fecal pellets were collected every 4 hours for 24 hours (n=3 mice/treatment). 16S rRNA was performed on the V4-V5 region using standard protocols (https://earthmicrobiome.org/protocols-and-standards/) in a High-Throughput Genome Analysis Core (Institute for Genomics & Systems Biology) at Argonne National Laboratory. Rarefaction was set at 10,000 reads to control for sequencing effort as stated in the source paper. Please see the source paper for additional details (Leone et al., 2015).

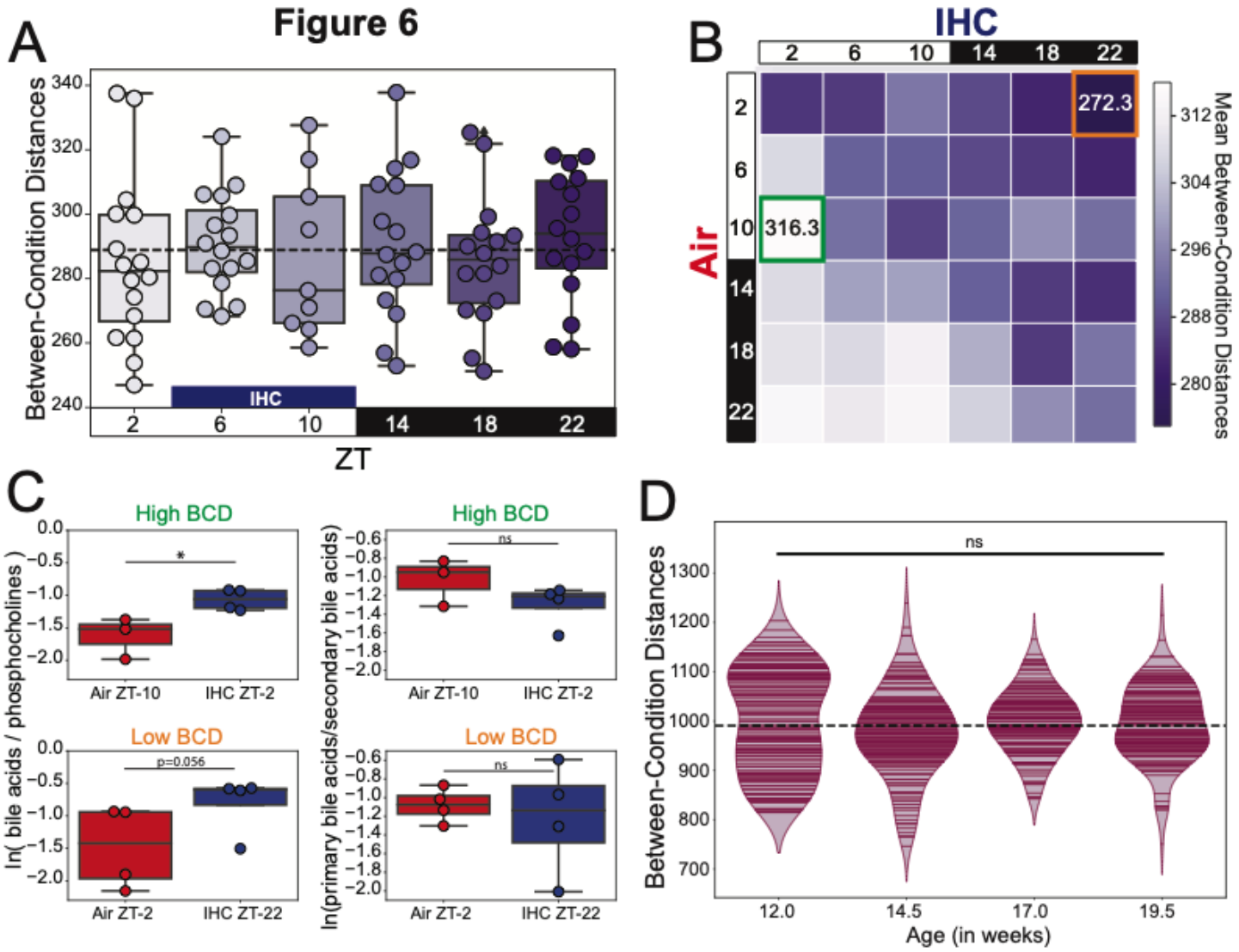
Figure 4.S6 - A total of 12 human subjects underwent 5 days of dietary intervention, either plant or animal based (n = 10 humans/condition). Patients that underwent both dietary interventions did so with a 1 month wash-out period in between interventions (10/12 patients; 9/10 patients per intervention). Two patients only underwent a single intervention (2/12 patients; one plant, one animal; 1/10). Please refer to the source paper for detailed study design and associated protocols (David et al., 2014).

Figure 4.S7 - In brief, two groups of five-week-old male *Balb/c* mice were kept in either a 12:12 L:D or 0:24 L:D vivarium fed a normal chow diet (unspecified in methods) *ad libitum*. After two weeks on condition, mice were anesthetized and sacrificed every 4 hours for 24hrs (n= 4-5 mice per group per time point). Samples from intestinal lumen, mucous layer, epithelial layer, and cecal contents were collected. The phenol-chloroform method was used for DNA extraction. 16S rRNA amplicon sequencing was performed on the V4 region. Rarefaction was set to 1085 reads to control sequencing effort, as performed in the source paper. Please refer to the source paper for detailed study design and associated protocols (Wu et al., 2018).

Metabolome: (**Figure 4.6**) In brief, each fecal sample was examined by untargeted liquid chromatography-tandem mass spectrometry (LC-MS/MS) as described previously (Tripathi et al., 2019). The samples were analyzed on a Vanquish ultrahigh-performance liquid chromatography (UPLC) system coupled to a Q-Exactive orbital ion trap (Thermo Fisher Scientific, Bremen, Germany) after homogenization and extraction. Chromatographic separation was performed using a C18 core shell column (Kinetex column, 50 by 2 mm, 1.7-µm particle size, 100-Å pore size; Phenomenex, Torrance, CA). The raw spectra outputs were converted to *m/z* extensible markup language (mzXML) in centroid mode using MSConvert (part of ProteoWizard). MassIVE links to raw files: Circadian - https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=0434de6d06f5424d8bd567808d069d4e ; longitudinal - https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=7996239533ea48738b550650e8531142**.** MZmine-2.37 was used for isotope peak removal with the alignment of peaks and creation of a feature matrix containing the feature retention times and the exact mass and peak areas of the corresponding extracted ion chromatograms. Identification of molecular features was performed using MS1-based feature detection and MS2-based molecular networking using the GNPS workflow (https://gnps.ucsd.edu) The GNPS jobs can be found at the following URLs: Circadian - https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=b0beb04fc3314f1e93cb69606e35ab64,

Figure 4.6. Untargeted LC-MS/MS shows no changes between conditions with respect to sample collection time. Intermittent hypoxia and hypercapnia = IHC. A) Canberra BCD (Air vs IHC) for diurnal data time points (see Fig 1A for experimental design). Significance was determined using paired Wilcoxon rank sum test, no significantly different pairs found. No circadian rhythmicity of the BCD found (MetaCycle, JTK method, p=1.000). B) Mean between-group weighted Canberra β-diversity distances heatmap. Highest highlighted in green, lowest highlighted in orange. Peak value is 1.3x trough value. C)The log-ratios of differentially abundant metabolites identified by group (phosphocholines vs bile acids) and sub-grouping (primary vs secondary bile acids). D) Experimental design is the same as described in Fig 6E, but untargeted liquid chromatography tandem mass spectrometry was performed on fewer time points than microbiome. N=12 mice per condition. Canberra BCD (Air vs IHC) for longitudinal data. Each horizontal line is a sample value. Significance was tested with a paired Wilcoxon rank-sum test. Notation: ns = not significant, * = $p<0.05$; ** = $p<0.01$; *** = $p<0.001$.

Figure 6


A
Between-Condition Distances
340
320
300
280
260
240
IHC
2
6
10
14
18
22
ZT
B
IHC
2
6
10
14
18
22
Air
272.3
316.3
Mean Between-Condition Distances
312
304
296
288
280
C
High BCD
ln(bile acids / phosphocholines)
*
Air ZT-10
IHC ZT-2
High BCD
ln(primary bile acids/secondary bile acids)
ns
Air ZT-10
IHC ZT-2
Low BCD
p=0.056
Air ZT-2
IHC ZT-22
Low BCD
ns
Air ZT-2
IHC ZT-22
D
Between-Condition Distances
ns
1300
1200
1100
1000
900
800
700
12.0
14.5
17.0
19.5
Age (in weeks)

Longitudinal

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=5b937aefecc14ed7b193ebb9a44193da.

Level 1 identification of bile acid metabolites as defined by the 2007 metabolomics standards initiative (Sumner et al., 2007) was performed using authentic bile acid standards from Cayman Chemical (Ann Arbor, MI). The annotated feature table was analyzed using QIIME2 (version 2020.6) (Bolyen et al., 2019). Canberra calculates overabundance on a feature-by-feature basis. Thus, while canberra is similar to Bray-Curtis, it is more sensitive to rare features. Data was visualized using custom python scripts, which can be found at https://github.com/knightlab-analyses/dynamics. Please refer to the source papers for detailed methods (Allaband et al., 2021; Tripathi et al., 2019).

Data Availability

Microbiome

Figure 4.1 - (Allaband/Zarrinpar 2021) - EBI accession ERP110592.

Figure 4.2 - (Zarrinpar/Panda 2014) - see supplemental excel file attached to source paper [PMID: 25470548]. (Leone/Chang 2015) - The figshare accession number for the 16S amplicon sequence data: http://dx.doi.org/10.6084/m9.figshare.882928

Figure 4.3 - (Tuganbaev/Elinav 2021) - ENA PRJEB38869

Figure 4.4 - (Leone/Chang 2015) - The figshare accession number for the 16S amplicon sequence data: http://dx.doi.org/10.6084/m9.figshare.882928

Figure 4.5 - (longitudinal circadian TRF) EBI**:** ERP123226 ; (longitudinal IHC) - EBI accession ERP110592.

Supplemental 4.4 (David/Turnbaugh 2012) - MG-RAST project ID 6248.

Metabolome

Figure 4.6 MassIVE link to raw data for circadian data - https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=0434de6d06f5424d8bd567808d069d4e. MassIVE link to raw data for longitudinal data - https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=7996239533ea48738b550650e8531142.

Python Notebooks: https://github.com/knightlab-analyses/dynamics.

4.5 Acknowledgments

Author Contributions

CA and AZ conceptualized and designed the analysis. CA, SFR, TK, HJ, MDT, ACDM, and RAR analyzed all the results. CA, AL, and AZ wrote the manuscript. EE, GGH, VAL, PCD, and RK provided critical intellectual input. All authors had access to data for the study. AZ supervised the study and is ultimately responsible for its content.

Disclosures

AZ is a co-founder and holds equity in Endure Biotherapeutics. PCD is an advisor to Cybele and Co-founder and adviser to Ometa and Enveda with prior approval from UC-San Diego.

Correspondence

Amir Zarrinpar, MD, PhD, 9500 Gilman Dr, MC 0983, University of California, San Diego, La Jolla, CA 92093-0983

Grant support

CA is supported by NIH T32 OD017863. SFR is supported by the Soros Foundation. TK is supported by NIH T32 GM719876. ACDM is supported by R01 HL148801-02S1. AZ is supported by the American Heart Association Beginning Grant-in-Aid (16BGIA27760160) and NIH K08 DK102902, R03 DK114536, R21 MH117780, R01 HL148801, R01 EB030134, R01HL157445, and U01 CA265719. All authors receive institutional support from NIH P30 DK120515, P30 DK063491, P30 CA014195, P50 AA011999, and UL1 TR001442.

Chapter 4 has been submitted for publication of the material as it may appear in Cell, "Microbiome Diurnal Dynamics Dominate Phenotype Effects, Enabling Replicability When Controlled." Celeste Allaband, Amulya Lingaraju, Stephany Flores-Ramos, Tanya Kumar, Haniyeh Javaheri, Maria D. Tiu, Ana Carolina Dantas Machado, Roland A. Richter, Emmanuel Elijah, Gabriel G. Haddad, Vanessa A. Leone, Pieter C. Dorrestein, Rob Knight, and Amir Zarrinpar. The dissertation author was the primary investigator and first author of this paper.

4.6 References

1. Allaband C, Lingaraju A, Martino C, Russell B, Tripathi A, Poulsen O, Dantas Machado AC, Zhou D, Xue J, Elijah E, Malhotra A, Dorrestein PC, Knight R, Haddad GG, Zarrinpar A (2021). Intermittent Hypoxia and Hypercapnia Alter Diurnal Rhythms of Luminal Gut Microbiome and Metabolome. MSystems e0011621.

2. Alvarez, Y., Glotfelty, L.G., Blank, N., Dohnalová, L., and Thaiss, C.A. (2020). The Microbiome as a Circadian Coordinator of Metabolism. Endocrinology 161, bqaa059.

3. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R (2017). Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. MSystems 2, e00191-16.

4. Baker, F. (1942). Normal Rumen Microflora and Microfauna of Cattle. Nature 149, 220–220.

5. Beli, E., Prabakaran, S., Krishnan, P., Evans-Molina, C., and Grant, M.B. (2019). Loss of Diurnal Oscillatory Rhythms in Gut Microbiota Correlates with Changes in Circulating Metabolites in Type 2 Diabetic db/db Mice. Nutrients 11, E2310.

6. Bisanz, J.E., Upadhyay, V., Turnbaugh, J.A., Ly, K., and Turnbaugh, P.J. (2019). Meta-Analysis Reveals Reproducible Gut Microbiome Alterations in Response to a High-Fat Diet. Cell Host Microbe 26, 265-272.e4.

7. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu YX, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS 2nd, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. 37, 852–857.

8. Brooks, J.F., and Hooper, L.V. (2020). Interactions among microbes, the immune system, and the circadian clock. Semin. Immunopathol. 42, 697–708.

9. Chaix, A., Zarrinpar, A., Miu, P., and Panda, S. (2014). Time-restricted feeding is a preventative and therapeutic intervention against diverse nutritional challenges. Cell Metab. 20, 991–1005.

10. Collado, M.C., Engen, P.A., Bandín, C., Cabrera-Rubio, R., Voigt, R.M., Green, S.J., Naqib, A., Keshavarzian, A., Scheer, F.A.J.L., and Garaulet, M. (2018). Timing of food intake impacts daily rhythms of human salivary microbiota: a randomized, crossover study. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. 32, 2060–2072.

11. David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V., Devlin, A. S., Varma, Y., Fischbach, M. A., Biddinger, S. B., Dutton, R. J., & Turnbaugh, P. J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. Nature 505, 559–563.

12. Deloris Alexander, A., Orcutt, R.P., Henry, J.C., Baker, J., Bissahoyo, A.C., and Threadgill, D.W. (2006). Quantitative PCR assays for mouse enteric flora reveal strain-dependent differences in composition that are influenced by the microenvironment. Mamm. Genome Off. J. Int. Mamm. Genome Soc. 17, 1093–1104.

13. Frazier, K., and Chang, E.B. (2020). Intersection of the Gut Microbiome and Circadian Rhythms in Metabolism. Trends Endocrinol. Metab. TEM 31, 25–36.

14. Friswell, M.K., Gika, H., Stratford, I.J., Theodoridis, G., Telfer, B., Wilson, I.D., and McBain, A.J. (2010). Site and strain-specific variation in gut microbiota profiles and metabolism in experimental mice. PloS One 5, e8584.

15. Gilbert, J.A., Blaser, M.J., Caporaso, J.G., Jansson, J.K., Lynch, S.V., and Knight, R. (2018). Current understanding of the human microbiome. Nat. Med. 24, 392–400.

16. Goodrich, J.K., Di Rienzi, S.C., Poole, A.C., Koren, O., Walters, W.A., Caporaso, J.G., Knight, R., and Ley, R.E. (2014). Conducting a Microbiome Study. Cell 158, 250–262.

17. Guo, T., Ho, C.-T., Zhang, X., Cao, J., Wang, H., Shao, X., Pan, D., and Wu, Z. (2019). Oolong Tea Polyphenols Ameliorate Circadian Rhythm of Intestinal Microbiome and Liver Clock Genes in Mouse Model. J. Agric. Food Chem. 67, 11969–11976.

18. Hatori, M., Vollmers, C., Zarrinpar, A., DiTacchio, L., Bushong, E. A., Gill, S., Leblanc, M., Chaix, A., Joens, M., Fitzpatrick, J. A., Ellisman, M. H., & Panda, S. (2012). Time-Restricted Feeding without Reducing Caloric Intake Prevents Metabolic Diseases in Mice Fed a High-Fat Diet. Cell Metab. 15, 848–860.

19. Kaczmarek, J.L., Musaad, S.M., and Holscher, H.D. (2017). Time of day and eating behaviors are associated with the composition and function of the human gastrointestinal microbiota. Am. J. Clin. Nutr. 106, 1220–1231.

20. Kenney, T., Gao, J., and Gu, H. (2020). Application of OU processes to modelling temporal dynamics of the human microbiome, and calculating optimal sampling schemes. BMC Bioinformatics 21, 450.

21. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciolek T, McCall LI, McDonald D, Melnik AV, Morton JT, Navas J, Quinn RA, Sanders JG, Swafford AD, Thompson LR, Tripathi A, Xu ZZ, Zaneveld JR, Zhu Q, Caporaso JG, Dorrestein PC. (2018). Best practices for analysing microbiomes. Nat. Rev. Microbiol. 16, 410–422.

22. Kohn, J.N., Kosciolek, T., Marotz, C., Aleti, G., Guay-Ross, R.N., Hong, S.-H., Hansen, S., Swafford, A., Knight, R., and Hong, S. (2020). Differing salivary microbiome diversity, community and diurnal rhythmicity in association with affective state and peripheral inflammation in adults. Brain. Behav. Immun. 87, 591–602.

23. Kohsaka, A., Laposky, A.D., Ramsey, K.M., Estrada, C., Joshu, C., Kobayashi, Y., Turek, F.W., and Bass, J. (2007). High-fat diet disrupts behavioral and molecular circadian rhythms in mice. Cell Metab. 6, 414–421.

24. Leone V, Gibbons SM, Martinez K, Hutchison AL, Huang EY, Cham CM, Pierre JF, Heneghan AF, Nadimpalli A, Hubert N, Zale E, Wang Y, Huang Y, Theriault B, Dinner AR, Musch MW, Kudsk KA, Prendergast BJ, Gilbert JA, Chang EB. (2015). Effects of diurnal variation of gut microbes and high-fat feeding on host circadian clock function and metabolism. Cell Host Microbe 17, 681–689.

25. Ley, R.E., Bäckhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D., and Gordon, J.I. (2005). Obesity alters gut microbial ecology. Proc. Natl. Acad. Sci. U. S. A. 102, 11070–11075.

26. Liang, X., Bushman, F.D., and FitzGerald, G.A. (2015). Rhythmicity of the intestinal microbiota is regulated by gender and the host circadian clock. Proc. Natl. Acad. Sci. U. S. A. 112, 10479–10484.

27. Lozupone, C., and Knight, R. (2005). UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. Appl. Environ. Microbiol. 71, 8228–8235.

28. Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J., and Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. ISME J. 5, 169–172.

29. Mirarab, S., Nguyen, N., and Warnow, T. (2011). SEPP: SATé-Enabled Phylogenetic Placement. In Biocomputing 2012, (Kohala Coast, Hawaii, USA: WORLD SCIENTIFIC), pp. 247–258.

30. Mistry, P., Reitz, C.J., Khatua, T.N., Rasouli, M., Oliphant, K., Young, M.E., Allen-Vercoe, E., and Martino, T.A. (2020). Circadian influence on the microbiome improves heart failure outcomes. J. Mol. Cell. Cardiol. 149, 54–72.

31. Morton, J.T., Marotz, C., Washburne, A., Silverman, J., Zaramela, L.S., Edlund, A., Zengler, K., and Knight, R. (2019). Establishing microbial composition measurement standards with reference frames. Nat. Commun. 10, 2719.

32. Mukherji, A., Kobiita, A., Ye, T., and Chambon, P. (2013). Homeostasis in intestinal epithelium is orchestrated by the circadian clock and microbiota cues transduced by TLRs. Cell 153, 812–827.

33. Musso, G., Gambino, R., and Cassader, M. (2011). Interactions between gut microbiota and host metabolism predisposing to obesity and diabetes. Annu. Rev. Med. 62, 361–380.

34. Piedrahita, J.A., Zhang, S.H., Hagaman, J.R., Oliver, P.M., and Maeda, N. (1992). Generation of mice carrying a mutant apolipoprotein E gene inactivated by gene targeting in embryonic stem cells. Proc. Natl. Acad. Sci. U. S. A. 89, 4471–4475.

35. Reitmeier S, Kiessling S, Clavel T, List M, Almeida EL, Ghosh TS, Neuhaus K, Grallert H, Linseisen J, Skurk T, Brandl B, Breuninger TA, Troll M, Rathmann W, Linkohr B, Hauner H, Laudes M, Franke A, Le Roy CI, Bell JT, Spector T, Baumbach J, O'Toole PW, Peters A, Haller D. (2020). Arrhythmic Gut Microbiome Signatures Predict Risk of Type 2 Diabetes. Cell Host Microbe 28, 258-272.e6.

36. Ren, B., Wang, L., Mulati, A., Liu, Y., Liu, Z., and Liu, X. (2021). Methionine Restriction Improves Gut Barrier Function by Reshaping Diurnal Rhythms of Inflammation-Related Microbes in Aged Mice. Front. Nutr. 8, 746592.

37. Sailani, M. R., Reza Sailani, M., Metwally, A. A., Zhou, W., Rose, S. M. S.-F., Ahadi, S., Contrepois, K., Mishra, T., Zhang, M. J., Kidziński, Ł., Chu, T. J., & Snyder, M. P. (2020). Deep longitudinal multiomics profiling reveals two biological seasonal patterns in California. Nat. Commun. 11, 4933.

38. Schloss, P.D. (2018). Identifying and Overcoming Threats to Reproducibility, Replicability, Robustness, and Generalizability in Microbiome Research. MBio 9, e00525-18.

39. Shao, Y., Shen, Q., Hua, R., Evers, S.S., He, K., and Yao, Q. (2018). Effects of sleeve gastrectomy on the composition and diurnal oscillation of gut microbiota related to the metabolic improvements. Surg. Obes. Relat. Dis. Off. J. Am. Soc. Bariatr. Surg. 14, 731–739.

40. Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, Schwager E, Crabtree J, Ma S; Microbiome Quality Control Project Consortium, Abnet CC, Knight R, White O,

Huttenhower C. (2017). Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat. Biotechnol. 35, 1077–1086.

41. Skarke, C., Lahens, N. F., Rhoades, S. D., Campbell, A., Bittinger, K., Bailey, A., Hoffmann, C., Olson, R. S., Chen, L., Yang, G., Price, T. S., Moore, J. H., Bushman, F. D., Greene, C. S., Grant, G. R., Weljie, A. M., & FitzGerald, G. A. (2017). A Pilot Characterization of the Human Chronobiome. Sci. Rep. 7, 17141.

42. Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., Fan, T. W.-M., Fiehn, O., Goodacre, R., Griffin, J. L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A. N., Lindon, J. C., Marriott, P., Nicholls, A. W., Reily M.D., Thaden J.J., Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics Off. J. Metabolomic Soc. 3, 211–221.

43. Takayasu, L., Suda, W., Takanashi, K., Iioka, E., Kurokawa, R., Shindo, C., Hattori, Y., Yamashita, N., Nishijima, S., Oshima, K., & Hattori, M. (2017). Circadian oscillations of microbial and functional composition in the human salivary microbiome. DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes, 24(3), 261–270.

44. Thaiss CA, Levy M, Korem T, Dohnalová L, Shapiro H, Jaitin DA, David E, Winter DR, Gury-BenAri M, Tatirovsky E, Tuganbaev T, Federici S, Zmora N, Zeevi D, Dori-Bachash M, Pevsner-Fischer M, Kartvelishvily E, Brandis A, Harmelin A, Shibolet O, Halpern Z, Honda K, Amit I, Segal E, Elinav E. (2016). Microbiota Diurnal Rhythmicity Programs Host Transcriptome Oscillations. Cell 167, 1495-1510.e12.

45. The Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. Nature 486, 207–214.

46. Tripathi A, Xu ZZ, Xue J, Poulsen O, Gonzalez A, Humphrey G, Meehan MJ, Melnik AV, Ackermann G, Zhou D, Malhotra A, Haddad GG, Dorrestein PC, Knight R. (2019). Intermittent Hypoxia and Hypercapnia Reproducibly Change the Gut Microbiome and Metabolome across Rodent Model Systems. MSystems 4.

47. Tuganbaev T, Mor U, Bashiardes S, Liwinski T, Nobs SP, Leshem A, Dori-Bachash M, Thaiss CA, Pinker EY, Ratiner K, Adlung L, Federici S, Kleimeyer C, Moresi C, Yamada T, Cohen Y, Zhang X, Massalha H, Massasa E, Kuperman Y, Koni PA, Harmelin A, Gao N, Itzkovitz S, Honda K, Shapiro H, Elinav E. (2020). Diet Diurnally Regulates Small Intestinal Microbiome-Epithelial-Immune Homeostasis and Enteritis. Cell 182, 1441-1459.e21.

48. Uhr, G.T., Dohnalová, L., and Thaiss, C.A. (2019). The Dimension of Time in Host-Microbiome Interactions. MSystems 4, e00216-18.

49. Voigt, R.M., Forsyth, C.B., Green, S.J., Mutlu, E., Engen, P., Vitaterna, M.H., Turek, F.W., and Keshavarzian, A. (2014). Circadian disorganization alters intestinal microbiota. PloS One 9, e97500.

50. Wang L, Ren B, Hui Y, Chu C, Zhao Z, Zhang Y, Zhao B, Shi R, Ren J, Dai X, Liu Z, Liu X. (2020). Methionine Restriction Regulates Cognitive Function in High-Fat Diet-Fed Mice: Roles of Diurnal Rhythms of SCFAs Producing- and Inflammation-Related Microbes. Mol. Nutr. Food Res. 64, e2000190.

51. Weger BD, Gobet C, Yeung J, Martin E, Jimenez S, Betrisey B, Foata F, Berger B, Balvay A, Foussier A, Charpagne A, Boizet-Bonhoure B, Chou CJ, Naef F, Gachon F. (2019). The Mouse Microbiome Is Required for Sex-Specific Diurnal Rhythms of Gene Expression and Metabolism. Cell Metab. 29, 362-382.e8.

52. Wu G, Tang W, He Y, Hu J, Gong S, He Z, Wei G, Lv L, Jiang Y, Zhou H, Chen P. (2018). Light exposure influences the diurnal oscillation of gut microbiota in mice. Biochem. Biophys. Res. Commun. 501, 16–23.

53. Zarrinpar, A., Chaix, A., Yooseph, S., and Panda, S. (2014). Diet and Feeding Pattern Affect the Diurnal Dynamics of the Gut Microbiome. Cell Metab. 20, 1006–1017.

54. Zhang, Y.-K.J., Guo, G.L., and Klaassen, C.D. (2011). Diurnal variations of mouse plasma and hepatic bile acid concentrations as well as expression of biosynthetic enzymes and transporters. PloS One 6, e16683.

Chapter 5. Engineered Native Bacteria are More Efficient at Engraftment and Colonization of a Host and Require Fewer Mutations for Maintenance Long Term

5.1 Introduction

In the fifth part of this thesis, we explore the impact of the host and established intact microbial community on the evolution of a single tagged microbe in a mouse host over its life. The data indicate that the non-native strain struggles with all aspects of adaptation to its new murine host. It has decreased colonization levels (CFU/g feces), increased loss of colonization, increased genome variability, increased genetic drift, increased pangenome variation, increased SNP mutation rate, increased phage presence, increased number of plasmids, and demonstrates loss of at least one of the detected phage defense systems compared to the native strain. The native strain is much more consistently present and genomically stable.

The adaptation and evolution of microbes inside their host is a phenomenon of interest in several fields. It is also of particular importance if we are going to try to start using the microbiome for therapeutic purposes, including personalized medicine. In his famous ongoing long-term evolution experiment (LTEE), Richard Lenski has been studying the adaptation and evolution of *Esherichia coli* (*E.coli*) *in vitro* for the past thirty-five years [refs]. Even in that simple system of minimal media in flasks, myriad mutations could be found and studied. Previous attempts to look into this phenomenon included germ-free [refs] or antibiotic-depleted [refs] models, which have provided great preliminary insight into the host-microbe-microbiome interaction. Unfortunately, these models do not accurately replicate the complexity of real world settings. These models have previously been needed because probiotics and other engineered strains have difficulty with engraftment and maintenance in an intact microbiome without daily repeated dosing.

5.2 Methods and Results

While our study does not fully replicate the complexities of a natural environment - such as environmental bacteria, daily dietary changes, and more - it does bring our model of *in vivo* adaptation and evolution a significant step closer to real-world applications. We were able to overcome many of the previous issues with engraftment by using more wild-type host-native strains of *E.coli*. We used a human-derived Nissle1917 strain (non-native to mouse, AZ20) and a locally cultured mouse native strain similar to NGF-1 (native to mouse, AZ51) that were engineered to have GFP and an antibiotic resistance marker incorporated into each strain's chromosome [ref]. While the strains do have most genes in common, there are about 700 unique genes present in each strain (**Fig A.D.1.S1A**).

The strains were gavaged once into 10-week-old SPF wild-type C57Bl/6 mice that were bred and raised in our double-barrier vivarium (12:12 L:D) together. Their parents were originally purchased from Jackson Labs. While the mice in this study have never received an antibiotic during the entirety of the experiment, the antibiotic resistance marker incorporated into the genome was used to re-isolate our strain on selective media for quantification and analysis. Neither of our strains negatively affected the health of the mice by both subjective observation or objective measurement (weight) (**Fig 5.1A,B**). This indicates that neither strain, once engrafted, conferred a clearly positive or negative effect on the host.

Overall, the non-native strain was worse at long term colonization of a mouse host than the native strain (**Fig 5.1C**). When present, the non-native strain colonized at an average of $\log_{10}$ 4.6 (CFU/g feces) (**Fig 5.1D**). However, colonization levels were irregular for many mice throughout the study and loss of colonization occurred. Loss of colonization was more prominent in females (38%) than in males (13%) (**Fig A.D.1.S2**). Since Nissle1917 was originally isolated from a male soldier in World War I, that might have somehow influenced colonization ability. The native strain colonized at an average of $\log_{10}$ 5.7 (CFU/g feces) (**Fig 5.1E**). Colonization levels

were consistent for all mice with no loss of colonization observed (**Fig 5.1F**). Unlike the native strain, many mice with the non-native strain started to show loss of colonization about 18 weeks (Month 4-5) into the study (**Fig A.D.1.S2**). In addition, this occurs at approximately 27 weeks of age (7 months), which is just after growth plate closure in mice and known to be influenced by estrogen levels (https://doi.org/10.1152%2Fajpendo.00646.2011), and could potentially also explain the sex differences in the non-native strains colonization levels. At the end of the study, we wanted to determine where in the gastrointestinal tract the strains were more likely to live. Both strains were most likely to be found in the cecum, with the ileum being the second most likely site (**Fig A.D.1.S3**).

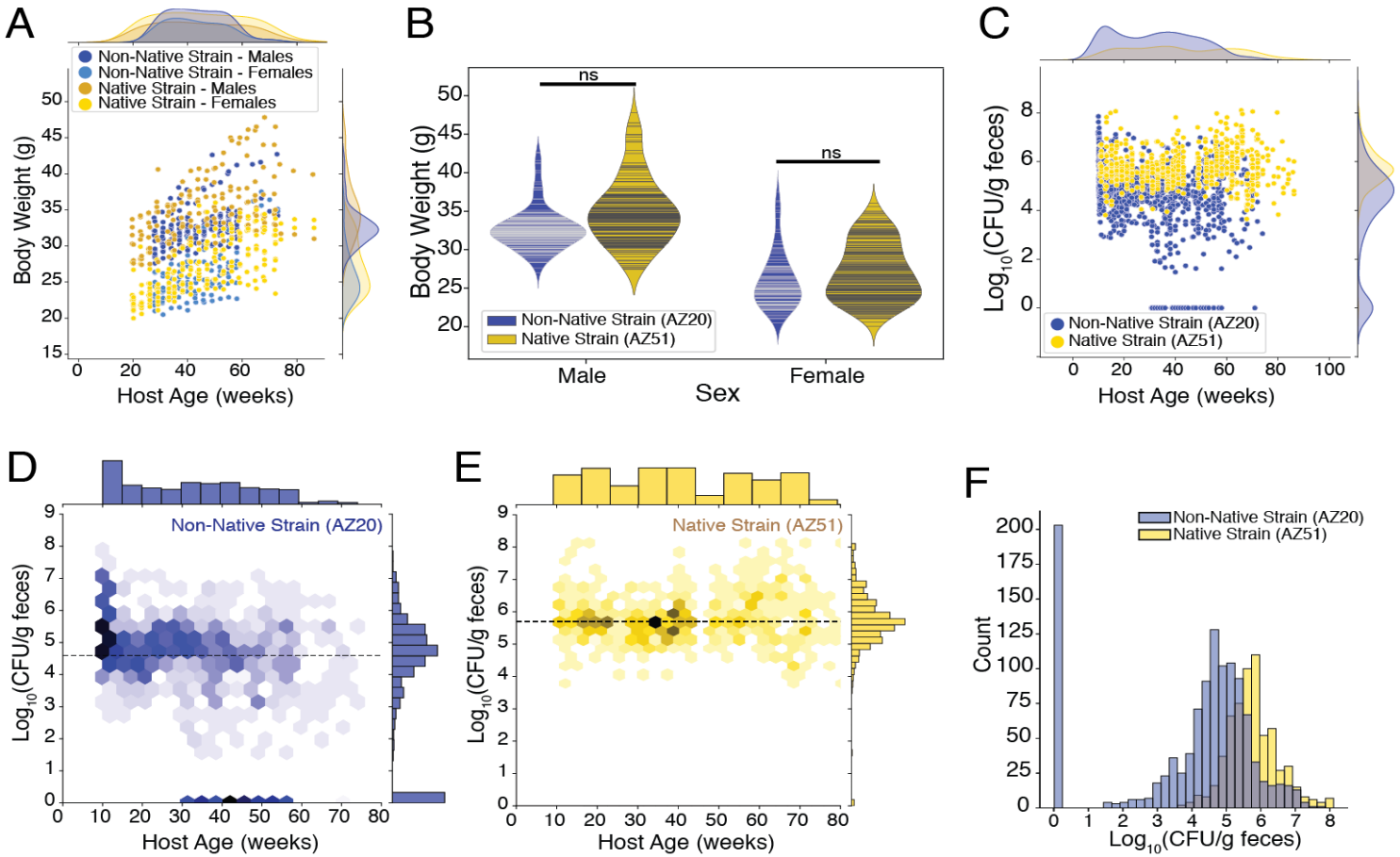


Figure 5.1: Body Weights and Colonization Data A) Jointplot of body weights in grams over time for both native and non-native cohorts. Slight delay in recording body weights for the non-native cohort can be seen. Body weight is used as an indirect objective measurement of host health. B) Violin plot of body weights grouped by strain and sex. Each horizontal bar represents a single measurement. Mann-Whitney-Wilcoxon test used to determine significance. C) Jointplot of Log10(CFU/g feces) over time for both cohorts. D) Jointplot of Log10(CFU/g feces) over time for the non-native strain alone. The dotted line is the median of non-zero colonization levels. E) Jointplot of Log10(CFU/g feces) over time for the native strain alone. The dotted line is the median of non-zero colonization levels. F) Countplot side-by-side comparison of the distribution of colonization levels. Notation: blue for non-native strain, gold for native strain; ns = not significant.

In order to more fully understand the genetic factors that underpin successful engraftment and colonization of these two strains, we took up to 24 individual colony isolates per mouse per time point. There are three major time points (Day 7, Month 3, Month6) where all mice were sampled. There are also three minor time points (Day 3, Month 1, Month 10) where only 4 mice, 2 male and 2 female, were sampled. This resulted in a total of 1319 isolates for the non-native strain and 1969 isolates for the native strain (**Fig A.D.1.1B, C**). Isolate cultures had DNA extracted, library prepped, and sequenced. The sequencing data was assembled, quality-checked, and was further analyzed as follows: pangenome, predicted metabolism, single nucleotide polymorphisms (SNPs), plasmids, phage, and phage defense systems.

The pangenome analysis took all of the detected genes from all the isolates (approximately 60,000) and created a presence/absence table. That table was converted to a jaccard distance matrix and then further analyzed by performing principal coordinates analysis (PCoA). When both groups of samples were analyzed together, PCoA Axis1 showed that there was no overlap between the two groups in gene composition and that explained about 79% of the variance.(**Fig 5.2A**). When the two strains were analyzed separately, both PCoA Axis1 showed clustering by genome size, as expected (**Fig 5.2B, C**). Neither strain appeared to cluster by time point (**Fig 5.2D, E**). There may have been some mild clustering by sex in the non-native strain isolates (PERMANOVA, p=0.027, pseudo-F 2.06) that was more prominent in the native strain cohort (PERMANOVA. p=0.003, pseudo-F 19.41). We also found that the native strain isolates had a much more narrow range (min=4397, max=5639, reference=4649) of genome sizes than the non-native strain isolates (**Fig 5.2F**). The non-native strain isolates had a wider distribution of genome size (min=4371, max=5950, reference=4625). The majority of the native strain isolates had genomes similar to the main reference genome with few outliers at all time points (**Fig 5.2G**), whereas the non-native strain had a smaller proportion of isolates with genomes similar to the reference with a greater proportion of outliers (**Fig 5.2H**).

Moreover, the completeness of major predicted metabolic pathways did not significantly change between early and late time points for either strain (data not shown). However, there was a significant difference in metabolic pathways present when comparing all non-native and native isolates (**Fig A.D.1.S4**). The non-native strain had more complete gamma-aminobutyrate (GABA) shunt pathways than the native strain. The native strain had more complete beta oxidation pathways than the non-native strain.

When collapsing the jaccard presence/absence table to a mouse and time point, RPCA was also calculated. The resulting distances showed that the non-native strain had significantly higher variances in pangenome composition compared to the native strain at all time points (**Fig A.D.1.S5**). The variances were particularly significant during the later time points. In addition, using that sample collapsed data used for RPCA, we found that both strains conformed to a neutral (genetic drift) model of evolution (**Fig A.D.1.S6**). Interestingly, the migration rate for both strains at the first major time point was about 100% and indicated most changes were coming from outside the genome. On the other hand, at the last major time point the non-native strain still had close to 100% outside migration, whereas the native strain dropped to 80%. Both the RPCA variance analysis and the neutral evolution models indicate significant differences in the way the two strains are responding to the environmental stressors.

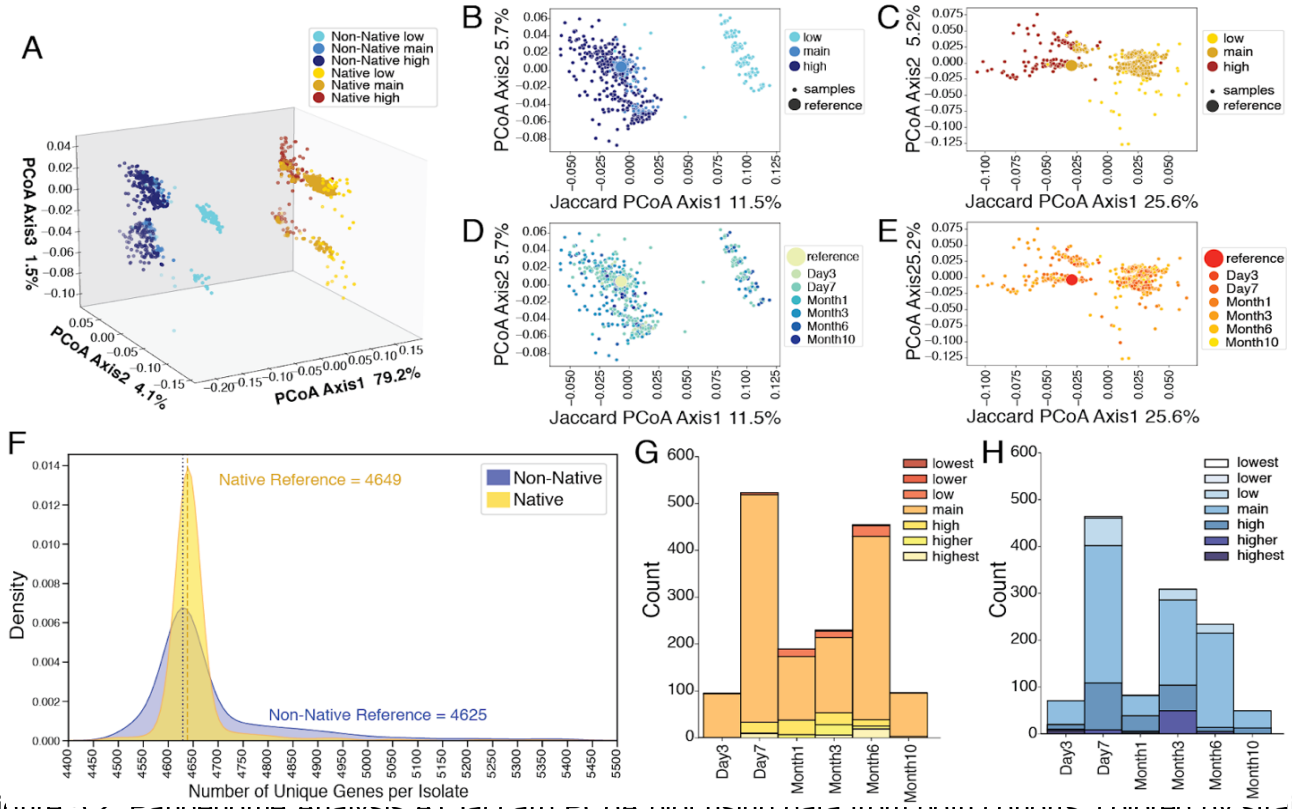


Figure 3.2: Pangenome Analysis A) Jaccard PCoA plot using data from both cohorts, colored by strain and detected number of genes. Each point is a single isolate. Genomes with less than 4000 detected genes are “low”, genomes with 4000 to 5000 detected genes are “main” and genomes with over 5000 detected genes are “high”. B) Jaccard PCoA plot of the non-native strain isolates only, colored by detected number of genes. C) Jaccard PCoA plot of the native strain isolates only, colored by detected number of genes. D) Jaccard PCoA plot of the non-native strain isolates only, colored by time point. The large dot represents the reference genome. E) Jaccard PCoA plot of the non-native strain isolates only, colored by time point. The large dot represents the reference genome. F) Distribution of the detected number of genes in each isolate, colored by strain. Dotted lines indicate the number of detected genes present in the respective reference genomes. G) Countplot of the number of native isolates at each time point in each bin. Each bin indicates the relative number of detected genes. H) Countplot of the number of non-native isolates at each time point in each bin. Each bin indicates the relative number of detected genes. Notation: blue/cool colors for non-native strain, gold/warm colors for native strain.

After examination of the pangenome where we looked at the changes in whole genes, we wanted to examine the prevalence of individual SNP mutations. We found that the overall mutation rate (number of any detected mutations per cell per base per year) was significantly higher in the non-native strain compared to the native strain at all major time points (**Fig 5.3A**). This was most prominent at the Day 7 time point. Interestingly, the pangenome's variation was more pronounced in the later time points but the difference in SNP mutations was more pronounced in the earlier time points. When looking at the raw number of synonymous mutations, which are less likely to confer a phenotype change, the non-native strain had more than the native strain at every major timepoint (**Fig 5.3B**). This was also true when we normalized the number to calculate the synonymous mutation rate (synonymous mutations per cell per base per year) (**Fig 5.3C**). The rate of synonymous mutations was most different at Day 7 between the native and non-native strains. When examining the non-synonymous mutations, which are more likely to confer a phenotype change, the raw number of mutations present was not significantly different between the two strains at early time points but was significantly higher in the non-native strain at the last major time point (Month 6) (**Fig 5.3D**). The rate of non-synonymous mutations was higher in the non-native strain compared to the native strain at all major timepoints (**Fig 5.3E**). This was particularly prominent at the later time points. The synonymous to non-synonymous ratio, thought to indicate either adaptive or purifying selection, was significantly higher in the non-native strain compared to the native strain at all timepoints (**Fig 5.3F**). The native strain's ratio was close to zero, potentially indicating purifying selection. There was no significant differences in mutation rates between isolates cultured from hosts of different sex (**Fig A.D.1.S7**). In general, it appears that the non-native strain responded to environmental stress by increasing the overall rate of mutation.

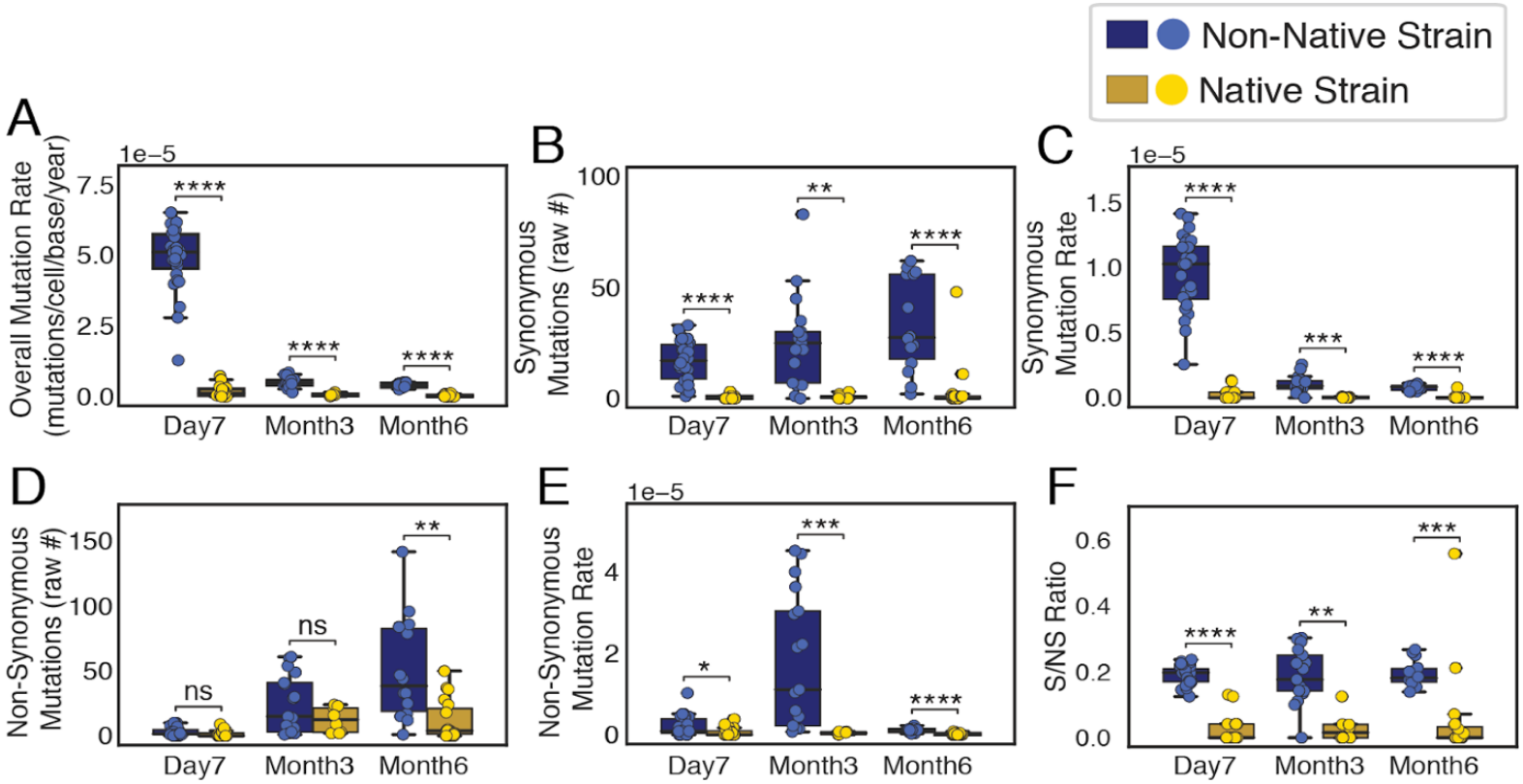


Figure 5.3: SNP Mutations A) Overall mutation rate, calculated as SNP mutations detected by Snippy [ref] per cell per base per year, for both cohorts at the three major timepoints. B) Raw number of synonymous mutations detected in both cohorts at the three major timepoints. C) Synonymous mutation rate, calculated as synonymous SNP mutations detected per cell per base per year, for both cohorts at the three major timepoints. D) Raw number of non-synonymous mutations detected in both cohorts at the three major timepoints. E) Non-synonymous mutation rate, calculated as non-synonymous SNP mutations detected per cell per base per year, for both cohorts at the three major timepoints. F) Synonymous/Non-Synonymous ratio for both cohorts at the three major timepoints. Lower numbers indicate more non-synonymous mutations and higher numbers indicate more synonymous mutations. A two-sided Mann-Whitney-Wilcoxon test with Bonferroni correction was used to test for significance between cohorts. Notation: blue/cool colors for non-native strain, gold/warm colors for native strain; ns: $p>0.05$, *: $0.01 < p <= 0.05$, **: $0.001 < p <= 0.01$, ***: $0.0001 < p <= 0.001$, ****: $p <= 0.0001$.

We wanted to examine more mobile genetic elements, including plasmids. The non-native strain reference genome had 2 original plasmids (5,641 bp/7 genes and 3300bp/5 genes). The native strain reference genome also had 2 original plasmids (42,613 bp/63 genes and 8633bp/7 genes). The non-native strain saw a trend towards an increase in the number of plasmids detected (**Fig 5.4A**). The native strain saw a trend towards a decrease in the number in plasmids detected, although this was a lot more variable (**Fig 5.4B**). The non-native strain had a more narrow range in the number of plasmid genes detected (**Fig 5.4C**). The native strain had a much wider range. This can at least be partially attributed to the small number of genes at the beginning. When looking at some of the top differentially expressed plasmid genes, we see that mbeA, mbeC, and mbeD have opposite trends in the two different strains at the major timepoints (**Fig 5.4D**). MbeA (relaxase) and mbeC (accessory protein) are key interacting proteins on mobilizable plasmid ColE1 and are linked to horizontal gene transfer (https://doi.org/10.1016/j.febslet.2012.01.060). MbeD is a plasmid entry exclusion protein (https://doi.org/10.1128/jb.177.21.6064-6068.1995). These trends suggest that the non-native strain is reducing mobile genetic elements later in the study, whereas the native strain appears to be increasing them in its population.

In addition, the native strain appears to have an increasing plasmid gene population prevalence of colicin over time (**Fig A.D.1.S8A, B, C**). While plasmids are sometimes beneficial (mutualistic relationship), they can also sometimes be a burden on the host bacteria (parasitic relationship). Colicins are known to help regulate the ecological interactions between plasmids and the host bacterial genome and aid in plasmid maintenance (https://doi.org/10.1007/s00438-022-01884-4). In the non-native strain, there is decreasing population prevalence of insN-1 transposon insertion site regulator element of cryptic prophage CP4-6 (**Fig A.D.1.S8D**). Decreases in insN may decrease propionate production (https://doi.org/10.1002/bit.27182). Since short-chain fatty acids, such as propionate, are known to have host effects, this may be important for colonization maintenance. Additionally, insN was seen to have increased expression during

anaerobic growth of E.coli (https://doi.org/10.1371/journal.pgen.1006570). Since the non-native strain is a Nissle1917 derivative that has been kept in freezers and mostly grown aerobically for the last hundred years, this may be critical for host adaptation. Finally, the non-native strain maintained a high population prevalence of the the relE/relB toxin/anti-toxin system over time (**Fig A.D.1.S8E, F**).

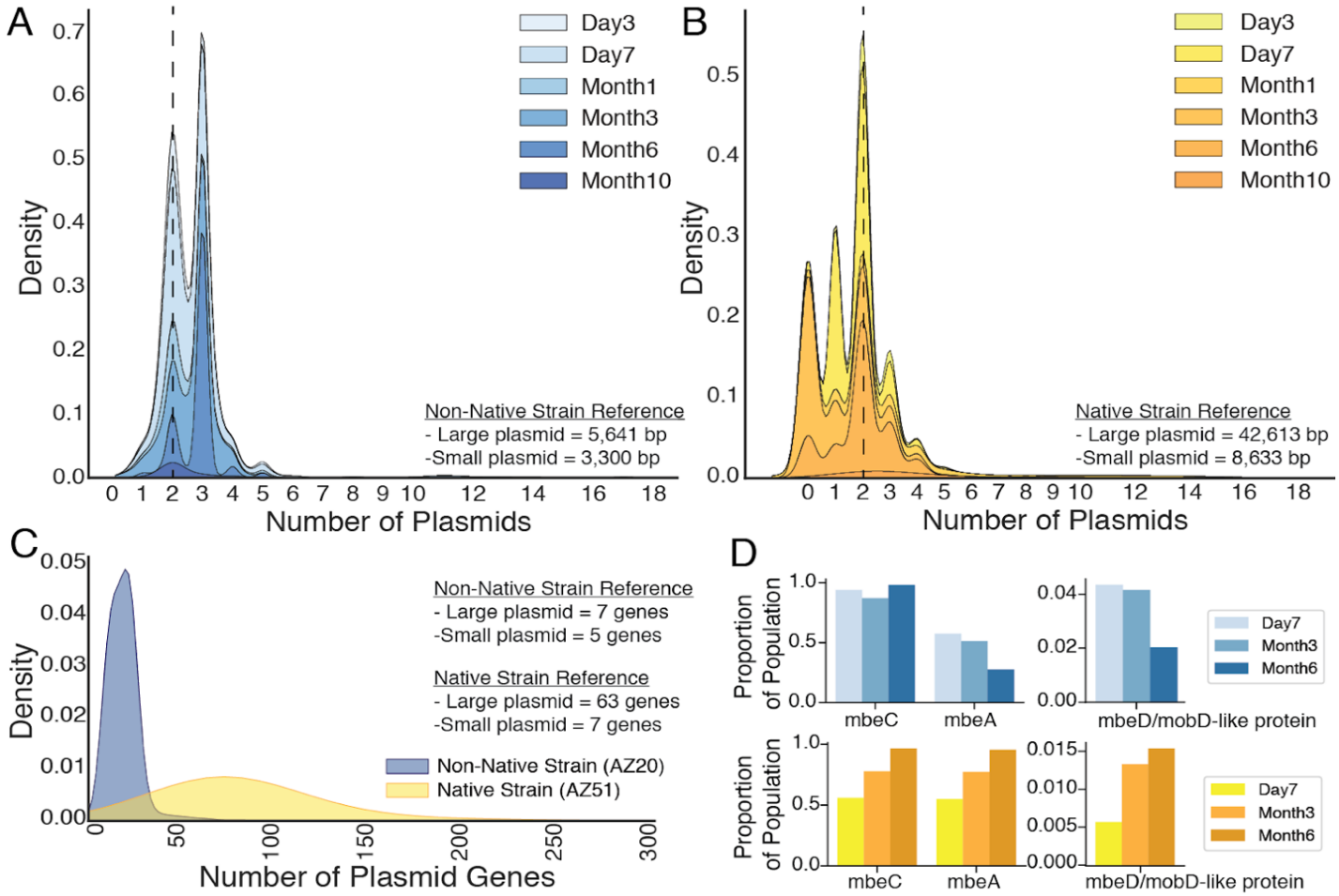


Figure 5.4: PlasmidsA) Stacked density plot of the number of detected plasmids present in the non-native strain cohort across all timepoints. The dotted line indicates the number of plasmids detected in the reference strain. B) Stacked density plot of the number of detected plasmids present in the native strain cohort across all timepoints. The dotted line indicates the number of plasmids detected in the reference strain. C) Layered density plot of the number of detected genes present in any of the plasmids detected. D) Changes in the proportion of isolates containing ColE1 genes (mbeA, mbeC, mbeD) of the cohort plasmid population across the three major time points. Notation: blue/cool colors for non-native strain, gold/warm colors for native strain.

We wanted to examine changes in detected prophage elements. Both strains were found to have 3 viral metagenomic assembled genomes (vMAGs) in their respective reference genomes. The non-native strain generally maintained or increased the vMAGs present in the genome (**Fig 5.5A**). The native strain generally maintained or decreased the vMAGs present in the genome (**Fig 5.5B**). The loss of a vMAG was much more prominent in the native strain than the gain of a vMAG in the non-native strain. The prophage genes present non-native strain reference genome were generally maintained in the non-native genome (**Fig 5.5C**). The native strain saw maintenance or a decrease in the prophage genes present in the native strain reference genome (**Fig 5.5D**). In addition to the core prophage genes present in the reference genomes, the non-native strain had a much larger repertoire of detected phage genes than the native strain (**Fig A.D.1.S9**). Indicating the non-native strain was much more vulnerable to phage attacks than the native strain.

Since the non-native strain appeared more susceptible to phage, we also examined changes in detected phage defense systems. The non-native strain started out with 7 different detected phage defense systems (Lamassu-Fam, Septu, Mokosh, PsyrTA, RM, Thoeris, and Kiwa), which were generally maintained across all time points (**Fig 5.6A**). However, there was a trend towards loss of the Kiwa defense system. It is unknown if that is linked to the trend toward increased prophage incorporation seen in **Fig 5.5A**. The native strain started out with 4 detected phage defense systems (Gabija, RM, Thoeris, and CAS), which were generally maintained across all time points (**Fig 5.6B**). There were some non-native isolates that picked up a few CAS genes, but they were not maintained over time (**Fig 5.6C**). There were also some native strain isolates that picked up two CBASS defense system genes, which were partially maintained in a small subset of the isolates but were trending towards being lost (**Fig 5.6D**).

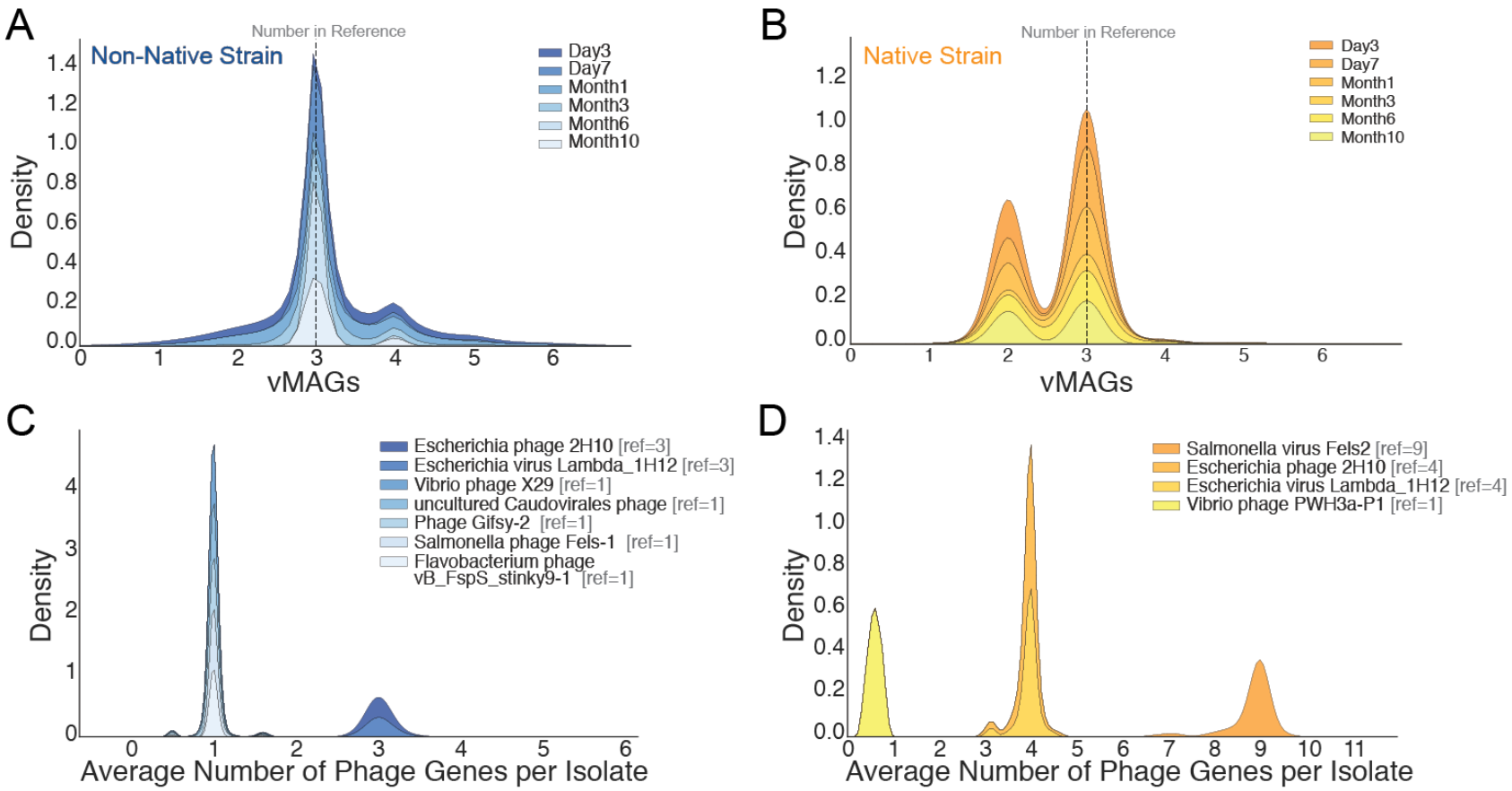


Figure 5.5: Prophage (vMAGs) and gene prevalenceA) Stacked density plot of the number of detected phage (vMAGs) present in the non-native strain cohort across all time points. Dotted line indicates the number of detected phage present in the reference genome. B) Stacked density plot of the number of detected phage (vMAGs) present in the native strain cohort across all time points. Dotted line indicates the number of detected phage present in the reference genome. C) Stacked density plot of the number of detected and annotated phage genes present in the non-native strain cohort across all time points. D) Stacked density plot of the number of detected and annotated phage genes present in the native strain cohort across all time points. Notation: blue/cool colors for non-native strain, gold/warm colors for native strain.

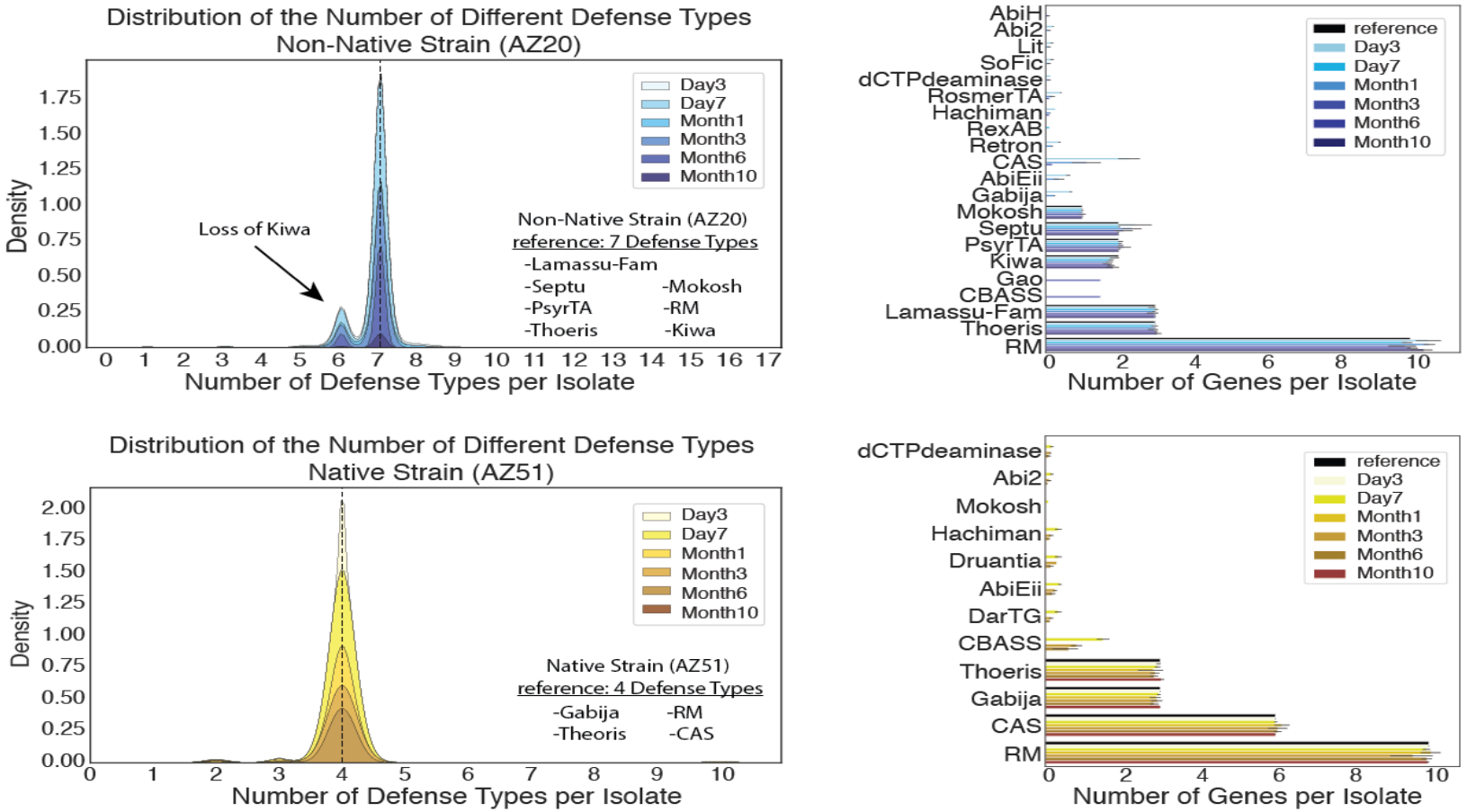


Figure 5.6: Anti-phage systems
A) Stacked density plot of the number of detected anti-phage systems present in the non-native strain cohort across all time points. Dotted line indicates the number of detected anti-phage systems present in the reference genome. B) Countplot/barplot of the number of detected anti-phage systems present per non-native isolate across all timepoints. Only anti-phage systems present in more than one time point are shown. Error bars indicate 95% confidence interval. C) Stacked density plot of the number of detected anti-phage systems present in the native strain cohort across all time points. Dotted line indicates the number of detected anti-phage systems present in the reference genome. D) Countplot/barplot of the number of detected anti-phage systems present per native isolate across all timepoints. Only anti-phage systems present in more than one time point are shown. Error bars indicate 95% confidence interval. Notation: blue/cool colors for non-native strain, gold/warm colors for native strain.

5.3 Conclusion

In conclusion, the non-native strain struggled to maintain colonization by every metric measured compared to the native strain. The non-native strain had about a log fold decrease in colonization levels and increased loss of colonization. The native strain had a narrower distribution of genome sizes, decreased genetic drift over time, and decreased RPCA variation. The non-native strain had an increased SNP mutation rate, especially at earlier time points. The non-native strain also had increased prophage presence in the genome and increased detection of various prophage genes in the genome. The native strain maintained its phage defense systems, whereas the non-native strain showed a loss of one of its defense system types with no clear trend for a replacement system.

These findings will allow us to better understand the ecological pressures present in a more real-world representative system, so that we can better target and manipulate at least one microbe of the microbiome. It is likely that this technique will work with other microbes of the microbiome that may be more prevalent or metabolically important. We envision that this system may work similarly to a CAR T-cell therapeutic system, where host-native strains could be cultured, modified, and returned to their host. It is these authors' hope that engineered native bacteria can eventually be used for long-term engraftment, colonization, and treatment of chronic disease that does not require the daily dosing of medication.

5.4 Acknowledgements

Chapter 5 has been submitted for publication of the material as it may appear in Nature, “Engineered Native Bacteria are More Efficient at Engraftment and Colonization of a Host and Require Less Mutations for Maintenance Long Term.” Celeste Allaband, Maria D. Tiu, Omar Din, Berhard Palsson, Rob Knight, and Amir Zarrinpar. The dissertation author was the primary investigator and first author of this paper.

The authors would like to thank Roland A. Richter for his guidance with coding issues.

Funding

CA is supported by NIH T32 OD017863. AZ is supported by the American Heart Association Beginning Grant-in-Aid (16BGIA27760160) and NIH K08 DK102902, R03 DK114536, R21 MH117780, R01 HL148801, R01 EB030134, R01HL157445, and U01 CA265719. All authors receive institutional support from NIH P30 DK120515, P30 DK063491, P30 CA014195, P50 AA011999, and UL1 TR001442.

Author Contributions

Conceptualization: CA, AZ
Methodology: CA, AZ
Investigation: CA, MDT, OD, CG, AHD
Visualization: CA, MDT, OD, CG, AHD, HJ,
Funding acquisition: AZ
Project administration: AZ
Supervision: RK, AZ
Writing – original draft: CA, AZ
Writing – review & editing: CA, MDT, OD, CG, AHD, BR, IM, HJ, SDB, RK, AZ

Competing Interests

AZ is a co-founder and holds equity in Endure Biotherapeutics.

Data Availability

Isolate Genomes:

Python Notebooks: https://github.com/knightlab-analyses/adaptation

5.5 References

1. Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., Lenski, R. E., & Kim, J. F. Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature* **461**, 1243–7 (2009).

2. Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat. Rev. Genet.* **4**, 457–469 (2003).

3. Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E. & Desai, M. M. The dynamics of molecular evolution over 60,000 generations. *Nature* **551**, 45–50 (2017).

4. Pedro, M., Batista-Barroso, J., Pinto, C., Dias, J., Gordo, I., & Xavier, K. Understanding metabolic processes shaping adaptation of E. coli to the gut. *Access Microbiol.* **1**, (2019).

5. Jones, R. T. & Curtiss, R. Genetic Exchange Between *Escherichia coli* Strains in the Mouse Intestine. *J. Bacteriol.* **103**, 71–80 (1970).

6. Crook, N., Ferreiro, A., Gasparrini, A. J., Pesesky, M. W., Gibson, M. K., Wang, B., Sun, X., Condiotte, Z., Dobrowolski, S., Peterson, D., & Dantas, G. Adaptive Strategies of the Candidate Probiotic E. coli Nissle in the Mammalian Gut. *Cell Host Microbe* **25**, 499-512.e8 (2019).

7. Giraud, A., Matic, I., Tenaillon, O., Clara, A., Radman, M., Fons, M., & Taddei, F. Costs and Benefits of High Mutation Rates: Adaptive Evolution of Bacteria in the Mouse Gut. *Science* **291**, 2606–2608 (2001).

8. Giraud, A., Arous, S., De Paepe, M., Gaboriau-Routhiau, V., Bambou, J.-C., Rakotobe, S., Lindner, A. B., Taddei, F., & Cerf-Bensussan, N. Dissecting the Genetic Components of Adaptation of Escherichia coli to the Mouse Gut. *PLoS Genet.* **4**, e2 (2008).

9. Lescat, M., Launay, A., Ghalayini, M., Magnan, M., Glodt, J., Pintard, C., Dion, S., Denamur, E., & Tenaillon, O. Using long-term experimental evolution to uncover the patterns and determinants of molecular evolution of an *Escherichia coli* natural isolate in the streptomycin-treated mouse gut. *Mol. Ecol.* **26**, 1802–1817 (2017).

10. Ghalayini, M., Magnan, M., Dion, S., Zatout, O., Bourguignon, L., Tenaillon, O., & Lescat, M.. Long-term evolution of the natural isolate of *Escherichia coli* 536 in the mouse gut colonized after maternal transmission reveals convergence in the constitutive expression of the lactose operon. *Mol. Ecol.* **28**, 4470–4485 (2019).

11. Barroso-Batista, J., Sousa, A., Lourenço, M., Bergman, M.-L., Sobral, D., Demengeot, J., Xavier, K. B., & Gordo, I. The First Steps of Adaptation of Escherichia coli to the Gut Are Dominated by Soft Sweeps. *PLoS Genet.* **10**, (2014).

12. Lawrence, J. G. & Ochman, H. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci.* **95**, 9413–9417 (1998).

13. Welch, R. A., Burland, V., Plunkett, G., 3rd, Redford, P., Roesch, P., Rasko, D., Buckles, E. L., Liou, S.-R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G. F., Rose, D. J., Zhou, S., Schwartz, D. C., Perna, N. T., Mobley, H. L. T., Donnenberg, M. S., & Blattner, F. R.. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci.* **99**, 17020–17024 (2002).

14. Russell, B. J., Brown, S. D., Saran, A. R., Mai, I., Lingaraju, A., Siguenza, N., Maissy, E., Dantas Machado, A. C., Pinto, A. F. M., Miyamoto, Y., Alexander Richter, R., Ho, S. B., Eckmann, L., Hasty, J., Saghatelian, A., Knight, R., & Zarrinpar, A. *Intestinal Transgene Delivery with Native* E. coli *Chassis Allows Persistent Physiological Changes.* http://biorxiv.org/lookup/doi/10.1101/2021.11.11.468006 (2021) doi:10.1101/2021.11.11.468006.

15. Ghalayini, M., Launay, A., Bridier-Nahmias, A., Clermont, O., Denamur, E., Lescat, M., & Tenaillon, O. Evolution of a Dominant Natural Isolate of Escherichia coli in the Human Gut over the Course of a Year Suggests a Neutral Evolution with Reduced Effective Population Size. *Appl. Environ. Microbiol.* **84**, (2018).

16. Couce, A. & Tenaillon, O. A. The rule of declining adaptability in microbial evolution experiments. *Front. Genet.* **6**, 99 (2015).

17. Russell, B., Mai, I., Allaband, C. & Zarrinpar, A. *Adaptation Sample Collection and Processing Protocol v1*. https://www.protocols.io/view/adaptation-sample-collection-and-processing-protoc-bwfxpbpn (2021) doi:10.17504/protocols.io.n2bvjx13nlk5/v1.

18. Geissmann, Q. OpenCFU, a New Free and Open-Source Software to Count Cell Colonies and Other Circular Objects. *PLoS ONE* **8**, e54072 (2013).

19. Tiu, M. D., Allaband, C., Mai, I., Russell, B. & Zarrinpar, A. *Colony Counting Protocol (OpenCFU) v1*. https://www.protocols.io/view/colony-counting-protocol-opencfu-bwfzpbp6 (2021) doi:10.17504/protocols.io.bwfzpbp6.

20. Gonzalez, A., Navas-Molina, J. A., Kosciolek, T., McDonald, D., Vázquez-Baeza, Y., Ackermann, G., DeReus, J., Janssen, S., Swafford, A. D., Orchanian, S. B., Sanders, J. G., Shorenstein, J., Holste, H., Petrus, S., Robbins-Pianka, A., Brislawn, C. J., Wang, M., Rideout, J. R., Bolyen, E., … Knight, R. Qiita: rapid, web-enabled microbiome meta-analysis. *Nat. Methods* **15**, 796–798 (2018).

21. Petit, R. A. & Read, T. D. Bactopia: a Flexible Pipeline for Complete Analysis of Bacterial Genomes. *mSystems* **5**, e00190-20 (2020).

22. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput. Biol.* **13**, e1005595 (2017).

23. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

24. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).

25. Titus Brown, C. & Irber, L. sourmash: a library for MinHash sketching of DNA. *J. Open Source Softw.* **1**, 27 (2016).

26. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

27. Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A., & Parkhill, J. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).

28. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu YX, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS 2nd, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology, 37(8), 852–857.. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).

29. Martino C, McDonald D, Cantrell K, Dilmore AH, Vázquez-Baeza Y, Shenhav L, Shaffer JP, Rahman G, Armstrong G, Allaband C, Song SJ, Knight R. (2022). Compositionally Aware Phylogenetic Beta-Diversity Measures Better Resolve Microbiomes Associated with Phenotype. mSystems 7, e0005022..

30. Ruiz-Perez, C. A., Conrad, R. E. & Konstantinidis, K. T. MicrobeAnnotator: a user-friendly, comprehensive functional annotation pipeline for microbial genomes. *BMC Bioinformatics* **22**, 11 (2021).

31. Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol. Biol. Clifton NJ* **1151**, 165–188 (2014).

32. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

33. Bush, S. J. Generalizable characteristics of false-positive bacterial variant calls. *Microb. Genomics* **7**, (2021).

34. Antipov D, Hartwick N, Shen M, Raiko M, Lapidus A, Pevzner PA. plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics* btw493 (2016) doi:10.1093/bioinformatics/btw493.

35. Guo, J., Vik, D., Adjie Pratama, A., Roux, S. & Sullivan, M. *Viral sequence identification SOP with VirSorter2 v3*. https://www.protocols.io/view/viral-sequence-identification-sop-with-virsorter2-bwm5pc86 (2021) doi:10.17504/protocols.io.bwm5pc86.

36. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitúa MC, Vik D, Sullivan MB, Roux S. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 (2021).

37. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2021).

38. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P, Narrowe AB, Rodríguez-Ramos J, Bolduc B, Gazitúa MC, Daly RA, Smith GJ, Vik DR, Pope PB, Sullivan MB, Roux S, Wrighton KC. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res.* **48**, 8883–8900 (2020).

39. Tesson F, Hervé A, Mordret E, Touchon M, d'Humières C, Cury J, Bernheim A. Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat. Commun.* **13**, 2561 (2022).

Appendix A. Supplemental Information for Chapter 2 Intermittent Hypoxia and Hypercapnia Alter Diurnal Rhythms of Luminal Gut Microbiome and Metabolome

A.A.1 Supplementary Figures

Figure A.A.1.S1. Overview of 16S Fecal Microbiome of ApoE-/- Mice on atherogenic diet with IHC treatment.
A) Alpha rarefaction curve. Dotted line indicates 12,000 (rarefaction depth). Shaded regions indicate 95% confidence interval. Mann-Whitney U used to determine statistical significance (statistic=190.5, p-value=0.053). B) Faith's Phylogenetic alpha diversity over time for each sample colored by group. Shaded regions indicate standard error of the mean. Mann-Whitney-Wilcoxon test was used to determine statistical significance. C) Weighted UniFrac PCoA with samples highlighted by timepoint. D) Pie graphs of cycling and non-cycling microbes classified to at least the family level for both conditions. E) Pie graphs of cycling and non-cycling reads attributed to a family level identifier for both conditions.

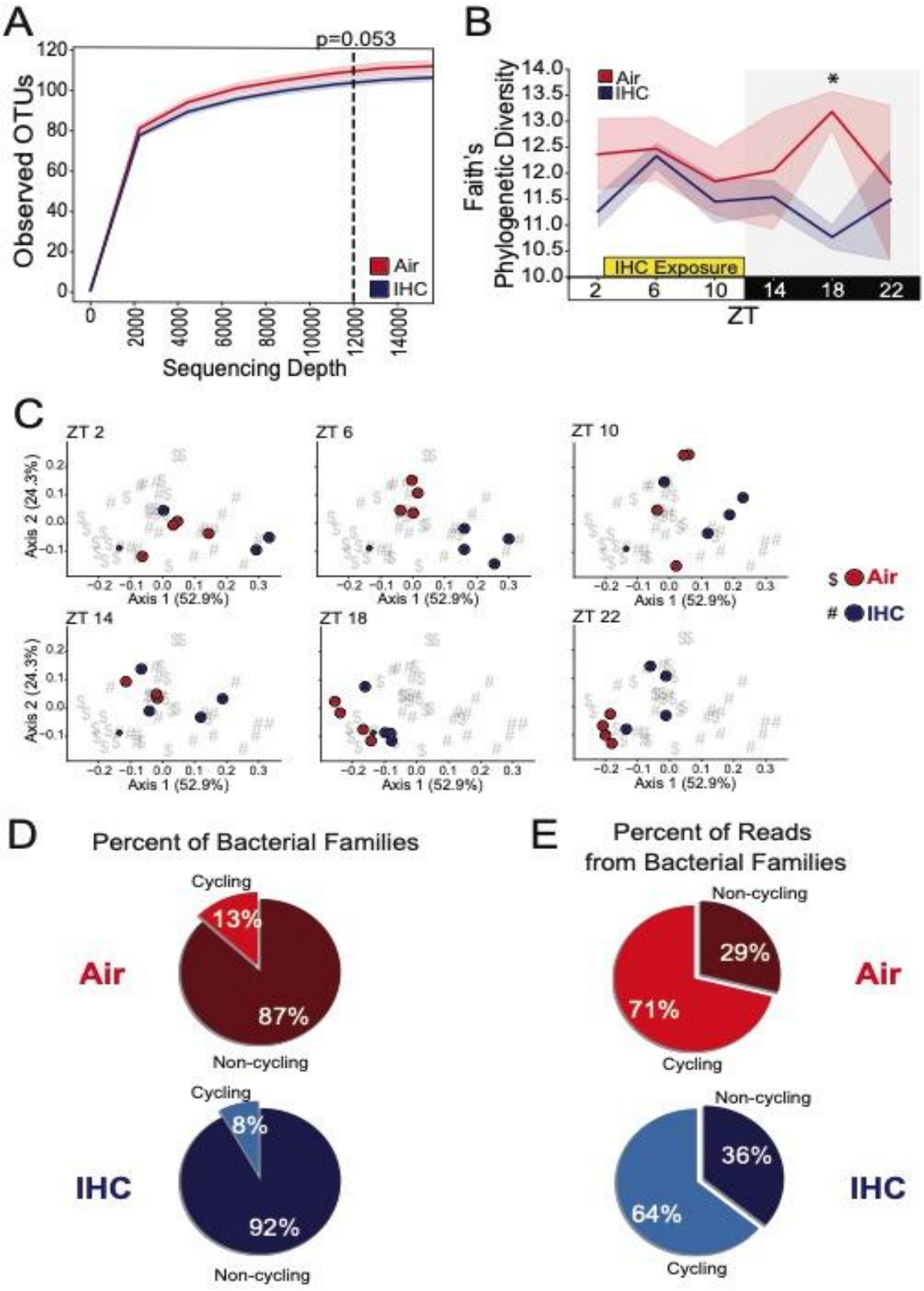

A
p=0.053
Observed OTUs
Sequencing Depth
Air
IHC
B
Faith's Phylogenetic Diversity
Air
IHC
*
IHC Exposure
ZT
C
ZT 2
ZT 6
ZT 10
ZT 14
ZT 18
ZT 22
Axis 1 (52.9%)
Axis 2 (24.3%)
$ Air
# IHC
D
Percent of Bacterial Families
Air
Cycling
13%
87%
Non-cycling
IHC
Cycling
8%
92%
Non-cycling
E
Percent of Reads from Bacterial Families
Air
Non-cycling
29%
71%
Cycling
IHC
Non-cycling
36%
64%
Cycling

Figure A.A.1.S2. IHC Affects the Cyclical Dynamics of Selected Phyla and Families in an OSA Mouse Model. 16S relative abundances. A) Phylum Bacteroidetes B) Phylum Firmicutes C) Phylum Verrucomicrobia D) Family Ruminococcaceae E) Family Lachnospiraceae F) Family S24-7 (“Homeothermaceae”) G) Family Coriobacteriaceae. Solid line represents the mean, error bars indicate standard error of the mean. Individual mice indicated by dashed line tracings. Shading indicates when room lights are off (i.e., active/feeding time for the mice). Yellow square indicates the 10 hours of IHC exposure under the IHC condition. Notation: @ circadian oscillation present ($p < 0.05$ as measured by MetaCycle42, JTK method); * = $p<0.05$ by Mann-Whitney-Wilcoxon test. Control samples with only exposure to normal room air conditions is red (n=4, 5-6 time points per mouse); Experimental samples exposed to IHC conditions for 10 hours per day are blue (n=4, 5-6 time points per mouse).

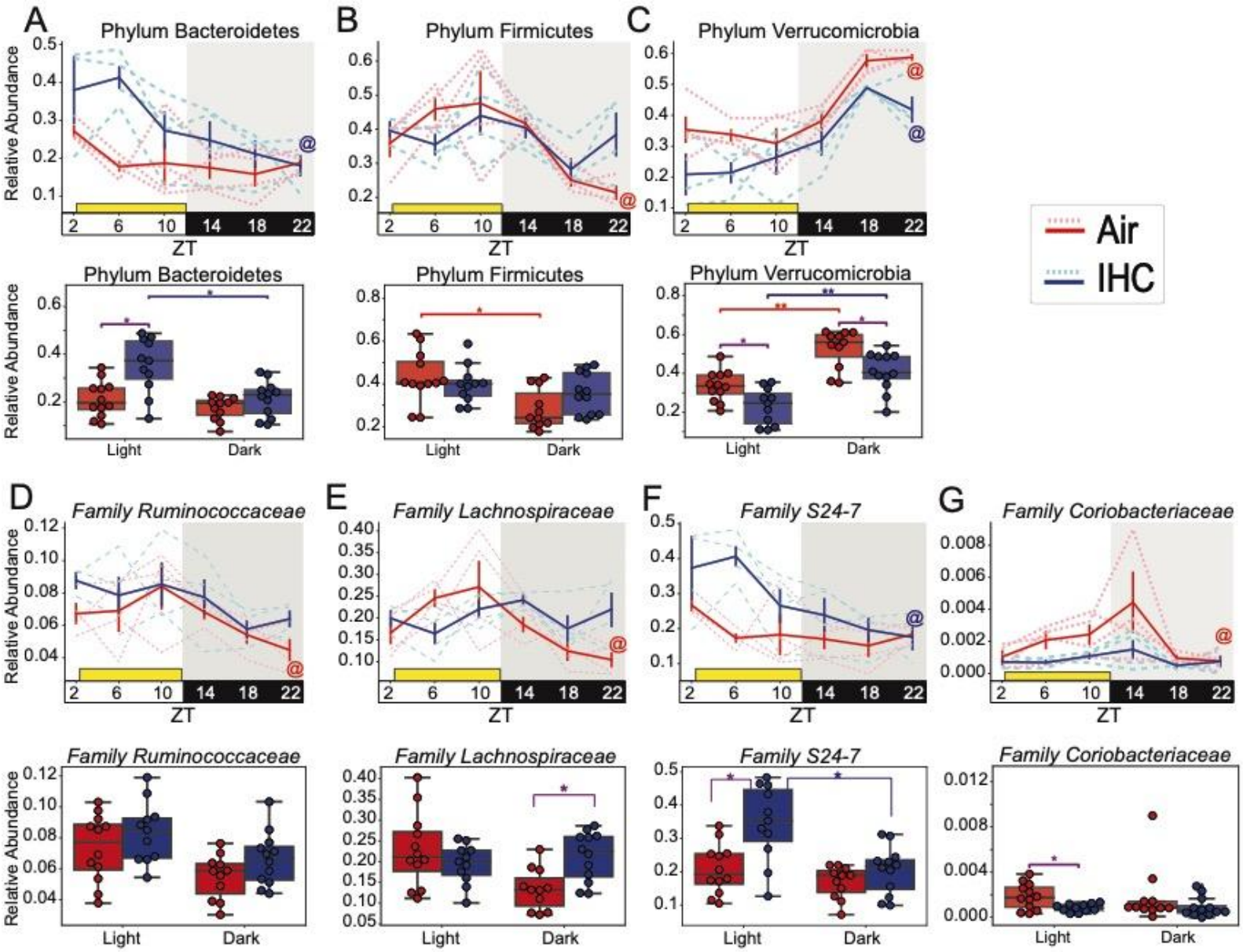
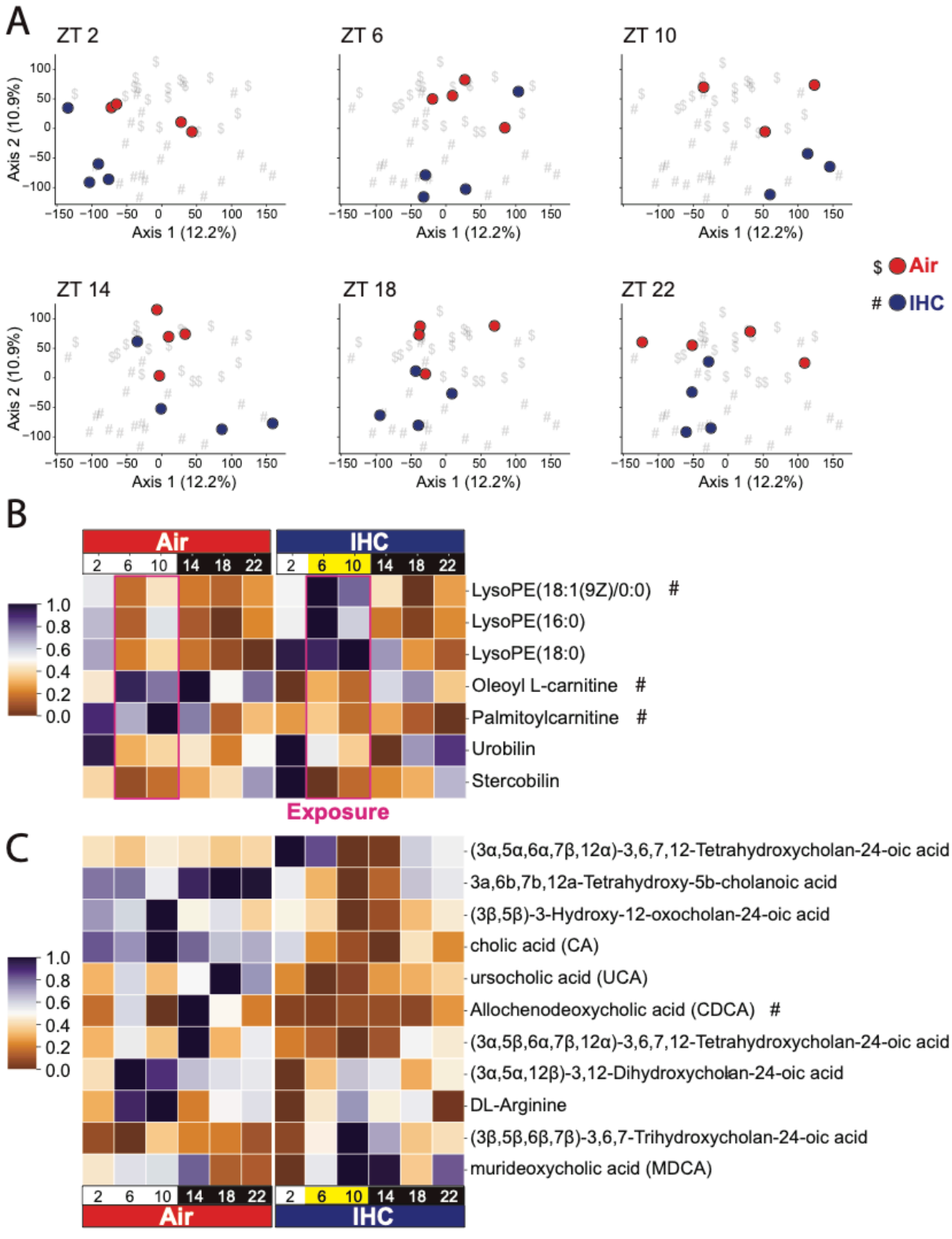

A
Phylum Bacteroidetes
B
Phylum Firmicutes
C
Phylum Verrucomicrobia
Relative Abundance
ZT
Air
IHC
Light
Dark
D
Family Ruminococcaceae
E
Family Lachnospiraceae
F
Family S24-7
G
Family Coriobacteriaceae

Figure A.A.1.S3. Additional Untargeted LC-MS/MS Metabolomics of Fecal Samples from ApoE-/- Mice on atherogenic diet after 1 week of IHC treatment.
A) Heatmap of additional selected metabolites. B) Heatmap of additional selected bile acids, including previously unrecognized bile acids. The value of each square of the heatmap represents the average relative abundance value (total sum normalized) for all mice in that condition for that time point. The heatmaps are also row normalized (row includes both conditions) and placed on a standard scale referenced on the left (0/brown=lowest value, 1/purple=highest value). The MS/MS spectral annotations were determined by using MS/MS-based spectral library matches for GNPS level 2 or 3 identification for all molecules displayed here. C) Canberra PCoA with samples highlighted by timepoint. Notation: # indicates a metabolite that is shown in Figure 2 or Figure S4. Control samples only exposure to normal Air conditions = red (N=4, 5-6 time points per mouse); Experimental samples exposed to IHC conditions = blue (N=4, 5-6 time points per mouse).


A
ZT 2
ZT 6
ZT 10
ZT 14
ZT 18
ZT 22
Axis 2 (10.9%)
Axis 1 (12.2%)
$ Air
# IHC
B
Air
IHC
2 6 10 14 18 22
1.0 0.8 0.6 0.4 0.2 0.0
LysoPE(18:1(9Z)/0:0) #
LysoPE(16:0)
LysoPE(18:0)
Oleoyl L-carnitine #
Palmitoylcarnitine #
Urobilin
Stercobilin
Exposure
C
(3α,5α,6α,7β,12α)-3,6,7,12-Tetrahydroxycholan-24-oic acid
3a,6b,7b,12a-Tetrahydroxy-5b-cholanoic acid
(3β,5β)-3-Hydroxy-12-oxocholan-24-oic acid
cholic acid (CA)
ursocholic acid (UCA)
Allochenodeoxycholic acid (CDCA) #
(3α,5β,6α,7β,12α)-3,6,7,12-Tetrahydroxycholan-24-oic acid
(3α,5α,12β)-3,12-Dihydroxycholan-24-oic acid
DL-Arginine
(3β,5β,6β,7β)-3,6,7-Trihydroxycholan-24-oic acid
murideoxycholic acid (MDCA)

x

Figure A.A.1.S4. IHC Affects the Cyclical Dynamics of Bile Acid Fecal Metabolites in an OSA Mouse Model. Untargeted LC-MS/MS was performed. Selected metabolites and their cyclical dynamics over time (longitudinal plot, upper) and their relative abundances grouped by cycle phase (box plots, lower). A) Deoxycholic acid (DCA). Level 1 identification. B) Tauro-**β**-Muricholic acid (T**β**MCA). Level 1 identification. C) Hyodeoxycholic acid (HDCA). Level 1 identification. D) Ursodeoxycholic acid (UDCA). E) Allochenodeoxycholic acid (CDCA). Level 3 annotation. Solid line represents the mean, error bars indicate standard error of the mean. Individual mice indicated by dashed line. Shading indicates when room lights are off (i.e., active/feeding time for the mice). Yellow square indicates the 10 hours of the day where mice under the IHC condition would be exposed to experimental conditions [ZT2-12]. Notation: @ indicates circadian oscillations as determined by MetaCycle42 (JTK) = $p < 0.05$; * = $p < 0.05$; ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$ by Mann-Whitney-Wilcoxon test. Control samples only exposure to normal Air conditions is red (n=4, 5-6 time points per mouse); Experimental samples exposed to IHC conditions is blue (n=4, 5-6 time points per mouse).

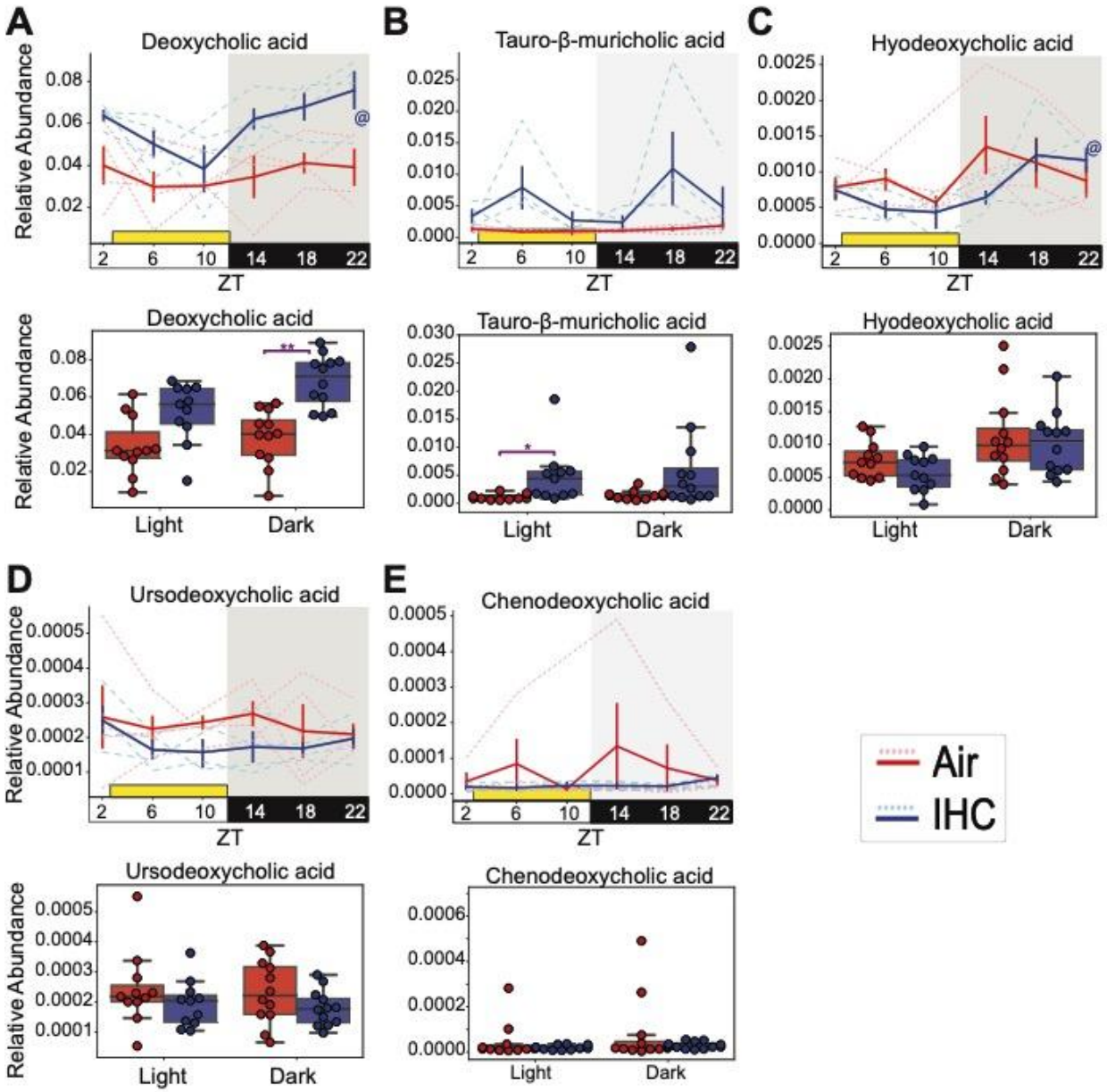
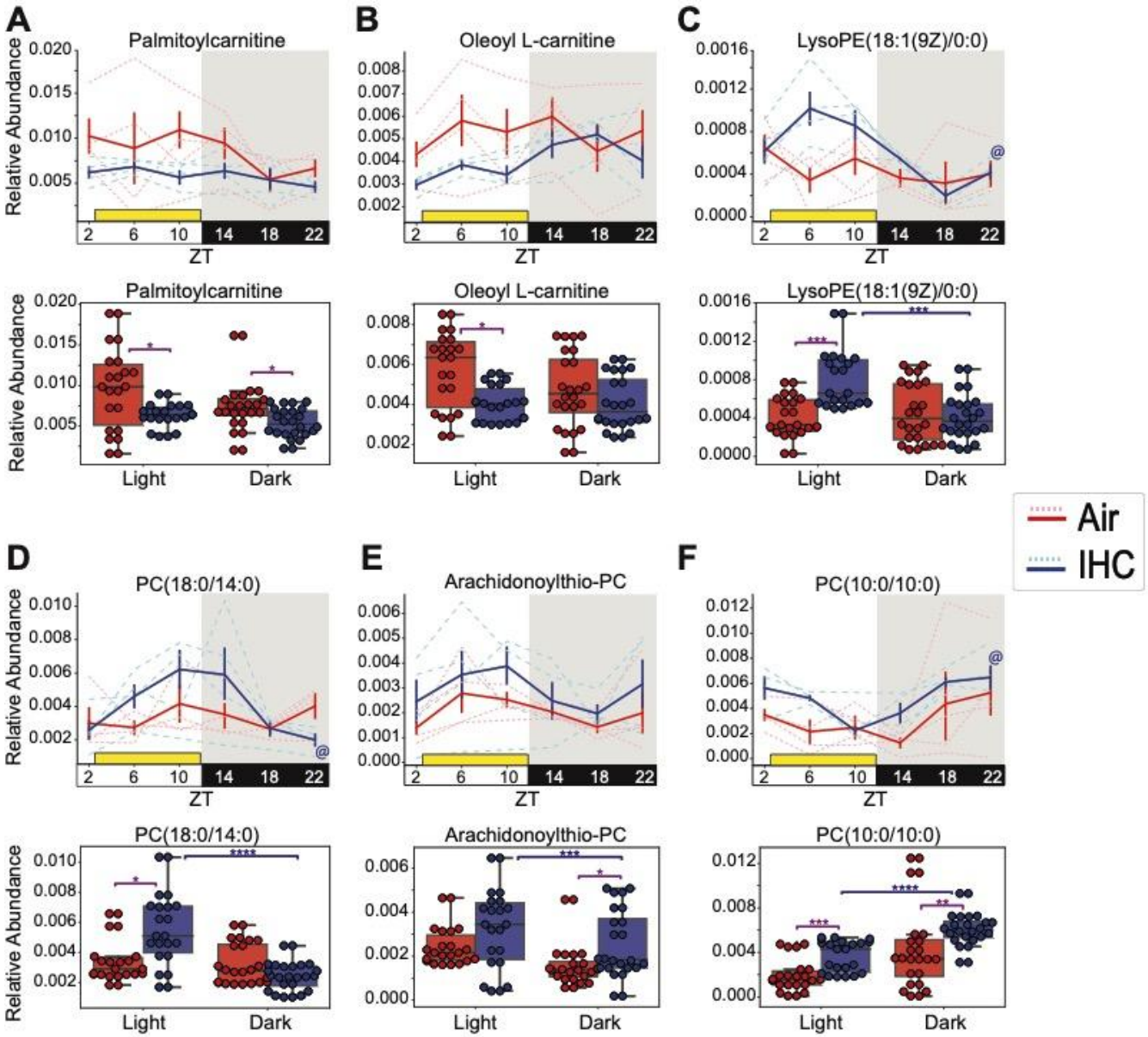

A
Deoxycholic acid
Relative Abundance
0.08
0.06
0.04
0.02
@
2 6 10 14 18 22
ZT
Deoxycholic acid
Relative Abundance
**
Light
Dark
B
Tauro-β-muricholic acid
0.025
0.020
0.015
0.010
0.005
0.000
2 6 10 14 18 22
ZT
Tauro-β-muricholic acid
0.030
0.025
0.020
0.015
0.010
0.005
0.000
*
Light
Dark
C
Hyodeoxycholic acid
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
@
2 6 10 14 18 22
ZT
Hyodeoxycholic acid
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
Light
Dark
D
Ursodeoxycholic acid
Relative Abundance
0.0005
0.0004
0.0003
0.0002
0.0001
2 6 10 14 18 22
ZT
Ursodeoxycholic acid
Relative Abundance
0.0005
0.0004
0.0003
0.0002
0.0001
Light
Dark
E
Chenodeoxycholic acid
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
2 6 10 14 18 22
ZT
Chenodeoxycholic acid
0.0006
0.0004
0.0002
0.0000
Light
Dark
Air
IHC

Figure A.A.1.S5. IHC Affects the Cyclical Dynamics of Selected Fecal Metabolites in an OSA Mouse Model. Untargeted LC-MS/MS was performed. Selected metabolites and their cyclical dynamics over time (longitudinal plot, upper) and their relative abundances grouped by cycle phase (box plots, lower). A) Palmitoylcarnitine. Level 3 annotation. B) 1-(9Z-Octadecenoyl)-sn-glycero-3-phosphoethanolamine (LysoPE (18:1(9Z)/0:0)). Level 3 annotation. C) 1-Stearoyl-2-myristoyl-sn-glycero-3-phosphocholine (PC (18:0/14:0)). Level 3 annotation. D) Oleoyl L-carnitine. Level 3 annotation. E) 1-hexadecyl-2-arachidonoylthio-2-deoxy-sn-glycero-3-phosphocholine (Arachidonoylthio-PC). Level 3 identification. F) Didecanoyl-glycerophosphocholine (PC (10:0/10:0)). Level 3 annotation. Solid line represents the mean, error bars indicate standard error of the mean. Individual mice indicated by dashed line. Shading indicates when room lights are off (i.e., active/feeding time for the mice). Yellow square indicates the 10 hours of the day where mice under the IHC condition would be exposed to experimental conditions [ZT2-12]. Notation: @ indicates circadian oscillations as determined by MetaCycle42 (JTK) = $p < 0.05$; * = $p < 0.05$; ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$ by Mann-Whitney-Wilcoxon test. Control samples only exposure to normal Air conditions is red (n=4, 5-6 time points per mouse); Experimental samples exposed to IHC conditions is blue (n=4, 5-6 time points per mouse).


A
Palmitoylcarnitine
Relative Abundance
ZT
Light
Dark
B
Oleoyl L-carnitine
C
LysoPE(18:1(9Z)/0:0)
D
PC(18:0/14:0)
E
Arachidonoylthio-PC
F
PC(10:0/10:0)
Air
IHC

A.A.2 Supplementary Tables

Table A.A.2.S1 Metabolomic Results Abbreviation Table. The MS/MS spectral annotations were determined by using MS/MS-based spectral library matches for GNPS level 3 identification for all molecules, except for some of the bile acids, as noted - standards were run using the same method as samples for level 1 identification.

Molecule	Abbreviation	Identification Level	Category
(R)-4-((3R,5S,8R,9S,10S,12R,13R,14S,17R)-3,12-dihydroxy-10,13-dimethylhexadecahydro-1H-cyclopenta[a]phenanthren-17-yl)pentanoic acid	(3**α**, 5**α**, 12**β**)-3,12-Dihydroxycholan-24-oic acid	Level 3	Bile acid
α-Muricholic acid	**α**MCA	Level 1	Bile acid
β-Muricholic acid	**β**MCA	Level 1	Bile acid
(4R)-4-((3R,5R,7R,9S,10S,13R,14S,17R)-3,7-dihydroxy-10,13-dimethylhexadecahydro-1H-cyclopenta[a]phenanthren-17-yl)pentanoic acid (also called allochenodeoxycholic acid or simply chenodeoxycholic acid)	CDCA	Level 3	Bile acid
Cholic acid	CA	Level 1	Bile acid
Deoxycholic acid	DCA	Level 1	Bile acid
Hyodeoxycholic acid	HDCA	Level 1	Bile acid
Spectral Match to Palmitoylcarnitine from NIST14	Palmitoylcarnitine	Level 3	Carnitine
Spectral Match to Oleoyl L-carnitine from NIST14	Oleoyl L-carnitine	Level 3	Carnitine
Spectral Match to Arachidonoylthio-PC from NIST14	Arachidonoylthio-PC	Level 3	Phosphocholine
(also called 1-hexadecyl-2-arachidonoylthio-2-deoxy-sn-glycero-3-phosphocholine, arachidonoyl thio-phosphocholine)		Level 3	
PC(20:5/0:0); [M+H]+ C28H49N1O7P1	LPC(20:5)	Level 3	Phosphocholine
PC(O-16:0/18:2); [M+H]+ C42H83N1O7P1	LPC(O-16:0)	Level 3	Phosphocholine
Spectral Match to 1-(9Z-Octadecenoyl)-2-tetradecanoyl-sn-glycero-3-phosphocholine from NIST14	PC(16:0/18:1**Ω**9)	Level 3	Phosphocholine
Spectral Match to 1-(9Z-Octadecenoyl)-sn-glycero-3-phosphocholine from NIST14	LPC(18:1)	Level 3	Phosphocholine

Table A.A.2.S1 Metabolomic Results Abbreviation Table. *Continued*

Molecule	Abbreviation	Identification Level	Category
Spectral Match to 1-Palmitoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine from NIST14	LysoPE(16:0)	Level 3	Phosphoethanolamine
Spectral Match to 1-Stearoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine from NIST14	LysoPE(18:0)	Level 3	Phosphoethanolamine
Spectral Match to 1-(9Z-Octadecenoyl)-sn-glycero-3-phosphoethanolamine from NIST14	LysoPE (18:1(9Z)/0:0)	Level 3	Phosphoethanolamine
Spectral Match to 1-Hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphocholine from NIST14	PC(16:0/18:1)	Level 3	Phosphocholine
Spectral Match to 1-Hexadecyl-sn-glycero-3-phosphocholine from NIST14	Lyso-PAF-C16	Level 3	Phosphocholine
Spectral Match to 1-Myristoyl-sn-glycero-3-phosphocholine from NIST14	LPC(14:0)	Level 3	Phosphocholine
Spectral Match to 1-Octadecanoyl-2-octadecenoyl-sn-glycero-3-phosphocholine from NIST14	GPC(18:0/18:1)	Level 3	Phosphocholine
Spectral Match to 1-Palmitoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine from NIST14	16:0 Lyso PE	Level 3	Phosphoethanolamine
Spectral Match to 1-Pentadecanoyl-sn-glycero-3-phosphocholine from NIST14	LPC(15:0)	Level 3	Phosphocholine
Spectral Match to 1-Stearoyl-2-hydroxy-sn-glycero-3-phosphocholine from NIST14	LPC(18:0/0:0)	Level 3	Phosphocholine
DIDECANOYL-GLYCEROPHOSPHOCHOLINE	PC(10:0/10:0)	Level 3	Phosphocholine
Spectral Match to 1,2-Dilinoleoyl-sn-glycero-3-phosphocholine from NIST14	DL-PC	Level 3	Phosphocholine
Spectral Match to 1,2-Ditetradecanoyl-sn-glycero-3-phosphocholine from NIST14	PC(14:0/14:0)	Level 3	Phosphocholine
Spectral Match to 1-Stearoyl-2-myristoyl-sn-glycero-3-phosphocholine from NIST14	PC(18:0/14:0)	Level 3	Phosphocholine
Stercobilin	Stercobilin	Level 3	Other
Spectral Match to Urobilin from NIST14	Urobilin	Level 3	Other

Tauro-b-muricholic acid [level 1]	TβMCA	Level 3	Bile acid
Taurocholic acid [level 1]	TCA	Level 3	Bile acid
Taurodeoxycholic acid [level 1]	TDCA	Level 3	Bile acid
Ursodeoxycholic acid [level 1]	UDCA	Level 3	Bile acid

Table A.A.2.S2 All unique annotated metabolites.

Metabolites
(4R)-4-((3R,5R,6S,7R,9S,10R,12S,13R,17R)-3,6,7,12-tetrahydroxy-10,13-dimethylhexadecahydro-1H-cyclopenta[a]phenanthren-17-yl)pentanoic acid
(4R)-4-((3R,5R,7R,9S,10S,13R,14S,17R)-3,7-dihydroxy-10,13-dimethylhexadecahydro-1H-cyclopenta[a]phenanthren-17-yl)pentanoic acid
(4R)-4-((3R,5S,6R,7R,9S,10R,12S,13R,14S,17R)-3,6,7,12-tetrahydroxy-10,13-dimethylhexadecahydro-1H-cyclopenta[a]phenanthren-17-yl)pentanoic acid
(4R)-4-((3R,5S,7S,9S,10S,12S,13R,14S,17R)-3,7,12-trihydroxy-10,13-dimethylhexadecahydro-1H-cyclopenta[a]phenanthren-17-yl)pentanoic acid
(4R)-4-((3S,5R,6S,7R,9S,10R,13R,14S,17R)-3,6,7-trihydroxy-10,13-dimethylhexadecahydro-1H-cyclopenta[a]phenanthren-17-yl)pentanoic acid
(4R)-4-((3S,5R,9S,10S,13R,14S,17R)-3-hydroxy-10,13-dimethyl-12-oxohexadecahydro-1H-cyclopenta[a]phenanthren-17-yl)pentanoic acid
(R)-4-((3R,5R,6R,7R,8R,9S,10R,12S,13R,14S,17R)-3,6,7,12-tetrahydroxy-10,13-dimethylhexadecahydro-1H-cyclopenta[a]phenanthren-17-yl)pentanoic acid
(R)-4-((3R,5R,6R,8S,9S,10R,13R,14S,17R)-3,6-dihydroxy-10,13-dimethylhexadecahydro-1H-cyclopenta[a]phenanthren-17-yl)pentanoic acid
(R)-4-((3R,5S,7R,8R,9S,10S,12S,13R,14S,17R)-3,7,12-trihydroxy-10,13-dimethylhexadecahydro-1H-cyclopenta[a]phenanthren-17-yl)pentanoic acid
(R)-4-((3R,5S,8R,9S,10S,12R,13R,14S,17R)-3,12-dihydroxy-10,13-dimethylhexadecahydro-1H-cyclopenta[a]phenanthren-17-yl)pentanoic acid
2-((4R)-4-((3R,5R,6S,7S,9S,10R,13R,17R)-3,6,7-trihydroxy-10,13-dimethylhexadecahydro-1H-cyclopenta[a]phenanthren-17-yl)pentanamido)ethane-1-sulfonic acid
2-((4R)-4-((3R,5R,7R,8R,9S,10S,12S,13R,17R)-3,7,12-trihydroxy-10,13-dimethylhexadecahydro-1H-cyclopenta[a]phenanthren-17-yl)pentanamido)ethane-1-sulfonic acid
4-(3,4-dihydroxyphenyl)-7-methoxy-5-[(2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-[[(2S,3R,4S,5R)-3,4,5-trihydroxyoxan-2-yl]oxymethyl]oxan-2-yl]oxychromen-2-one
Cer(d18:0/16:0(2OH)); [M+H]+ C34H70N1O4
Cer(d18:1/26:0(2OH)); [M+H]+ C44H88N1O4
Ceramide (18:1/16:0)
Cholic acid
DIDECANOYL-GLYCEROPHOSPHOCHOLINE
Deoxycholic acid
Hyodeoxycholic acid
LAUROYLCARNITINE
Massbank:PB000419 Arginine|2-amino-5-(diaminomethylideneamino)pentanoic acid
MassbankEU:SM850301 Sulfamethazine|4-amino-N-(4,6-dimethylpyrimidin-2-yl)benzenesulfonamide
NCGC00385123-01_C22H30O6_7b,9-Dihydroxy-3-(hydroxymethyl)-1,1,6,8-tetramethyl-5-oxo-1,1a,1b,4,4a,5,7a,7b,8,9-decahydro-9aH-cyclopropa[3,4]benzo[1,2-e]azulen-9a-yl acetate
PC(0:0/16:0); [M+H]+ C24H51N1O7P1

Table A.A.2.S2 All unique annotated metabolites. *Continued*

Metabolites
PC(0:0/20:4); [M+H]+ C28H51N1O7P1
PC(16:0/20:5); [M+H]+ C44H79N1O8P1
PC(18:1/20:4); [M+H]+ C46H83N1O8P1
PC(18:2/0:0); [M+H]+ C26H51N1O7P1
PC(18:3/0:0); [M+H]+ C26H49N1O7P1
PC(19:0/0:0); [M+H]+ C27H57N1O7P1
PC(20:3/0:0); [M+H]+ C28H53N1O7P1
PC(O-16:0/18:2); [M+H]+ C42H83N1O7P1
PC(O-18:1/0:0); [M+H]+ C26H55N1O6P1
Phytosphingosine
ReSpect:PS001803 (+-)-alpha-Tocopherol|Vitamin E|DL-all-rac-alpha-Tocopherol|Phytogermine|Profecundin|Syntopherol|Waynecomycin|Almefrol|Denamone|Emipherol
ReSpect:PS001805 (+-)-alpha-Tocopherol|Vitamin E|DL-all-rac-alpha-Tocopherol|Phytogermine|Profecundin|Syntopherol|Waynecomycin|Almefrol|Denamone|Emipherol
SM(d18:1/16:0); [M+H]+ C39H80N2O6P1
SUCROSE
Spectral Match to (+)-.alpha.-Tocopherol from NIST14
Spectral Match to (+)-4-Cholesten-3-one from NIST14
Spectral Match to 1,2-Dilinoleoyl-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-(1Z-Hexadecenyl)-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-(9Z-Octadecenoyl)-2-tetradecanoyl-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-(9Z-Octadecenoyl)-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-(9Z-Octadecenoyl)-sn-glycero-3-phosphoethanolamine from NIST14
Spectral Match to 1-Heptadecanoyl-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-Hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-Hexadecanoyl-2-octadecadienoyl-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-Hexadecyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-Hexadecyl-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-Myristoyl-2-palmitoyl-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-Myristoyl-2-stearoyl-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-Myristoyl-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-Octadecanoyl-2-(5Z,8Z,11Z,14Z-eicosatetraenoyl)-sn-glycero-3-phosphocholine from NIST14

Table A.A.2.S2 All unique annotated metabolites. *Continued*

Metabolites
Spectral Match to 1-Octadecanoyl-2-octadecenoyl-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-Octadecanoyl-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-Oleoyl-2-palmitoyl-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-Palmitoyl-2-docosahexaenoyl-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-Palmitoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine from NIST14
Spectral Match to 1-Palmitoyl-2-myristoyl-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-Stearoyl-2-hydroxy-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-Stearoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine from NIST14
Spectral Match to 1-Stearoyl-2-linoleoyl-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 1-Stearoyl-2-myristoyl-sn-glycero-3-phosphocholine from NIST14
Spectral Match to 12-Ketodeoxycholic acid from NIST14
Spectral Match to 3.alpha.,7.alpha.,12.alpha.-Trihydroxycholestanoic acid from NIST14
Spectral Match to 3.beta.-Hydroxy-5-cholenoic acid from NIST14
Spectral Match to 9-Octadecenamide, (Z)- from NIST14
Spectral Match to Arachidonoylthio-PC from NIST14
Spectral Match to Benzyldodecyldimethylammonium from NIST14
Spectral Match to Benzyltetradecyldimethylammonium from NIST14
Spectral Match to Cholic acid from NIST14
Spectral Match to D-erythro-Sphinganine from NIST14
Spectral Match to Elaidic acid from NIST14
Spectral Match to L-Tyrosine from NIST14
Spectral Match to Lyso-PC(16:0) from NIST14
Spectral Match to N-Tetracosenoyl-4-sphingenine from NIST14
Spectral Match to Oleoyl L-carnitine from NIST14
Spectral Match to Palatinose from NIST14
Spectral Match to Palmitoylcarnitine from NIST14
Spectral Match to Pantothenic acid from NIST14
Spectral Match to Phytosphingosine from NIST14
Spectral Match to Urobilin from NIST14
Spectral Match to Ursodeoxycholic acid from NIST14
Spectral Match to p-Hydroxyphenyllactic acid from NIST14
Stercobilin
THIAMINE
TOP 8 Psoriasis feature - Unknown FeatureID=4262

Table A.A.2.S2 All unique annotated metabolites. *Continued*

Metabolites
TOP19 Psoriasis feature - Unknown FeatureID=3668
Tauro-b-muricholic acid
Taurocholic acid
Taurodeoxycholic acid
Ursodeoxycholic acid
a-Muricholic acid
b-Muricholic acid
sphingosin C16 (Artifact)-emf
taurocholic acid
taurodeoxycholic acid

Table A.A.2.S3 Break down of circadian oscillators (Metacycle, JTK_CYCLE, p<0.05)

Level	Name	Figure Location	number of cycling sOTU	Condition(s)	Lowest level ID(s)
Phylum	Bacteroidetes	Fig 2A, S2A	1	1-both	Family S24-7
Phylum	Firmicutes	Fig 2A, S2B	9	5-Air 2-IHC 2-both	6-Family Lachnospiraceae 1-Family Ruminococcaceae 1-*Lactococcus* sp. 1-*Aldercreutzia* sp.
Phylum	Verrucomicrobia	Fig 2A, S2C*	1	1-both	*Akkermansia muciniphila*

Level	Name	Figure Location	number of cycling sOTU	Condition(s)	Lowest level ID(s)
Family	Lachnospiraceae	Fig 2C	6	4-Air 2-IHC	none ID past family level
Family	Ruminococcaceae	Fig 2B	1	1-both	*Anaerotruncus* sp.
Family	Streptococcaceae	Fig S2E	1	1-both	*Lactococcus* sp.
Family	Coriobacteriaceae	Fig S2D	1	1-Air	*Adlercreutzia* sp.
Family	S24-7	Fig 2D	1	1-both	none ID past family level
Family	Verrucomicrobiaceae	Fig 2A, S2C*	1	1-both	*Akkermansia muciniphila*

*since there is only one member of Phylum Verrucomicrobia/Family Verrucomicrobiaceae (*Akkermansia muciniphila*), all graphs of those levels are essentially of that organism

Appendix B. Supplemental Files for Chapter 3 Intermittent Hypoxia and Hypercapnia Alter Gut Microbiome and Metabolome to Promote Atherosclerosis

A.B.1 Supplemental Figures

Figure A.B.1.S1.. Additional 16S microbiome of ApoE-/- mice on HFD during chronic 10-wk treatment.(A) Axis 1 of the Weighted UniFrac PCoA of the metabolome for all measured time points (B) Relative abundance heatmap (row normalized, set on standard scale) compared by exposure types over time. Longitudinal relative abundance values for: (C) Family Coriobacteriaceae; (D) Family Verrucomicrobiaceae. The only ASV in this family belongs to Akkermansia muciniphila; (E) Family Lachnospiraceae. PERMANOVA used for statistical comparisons at the population level. LME modeling used for statistical comparisons of individual metabolites. The shaded areas in parts C-E represent standard error of the mean. Air/controls are black (n=6), IH is blue (n=12), IC is red (n=12). Statistical significance $p<0.05$, * IC vs Air, # IH vs Air and $ IH vs IC.

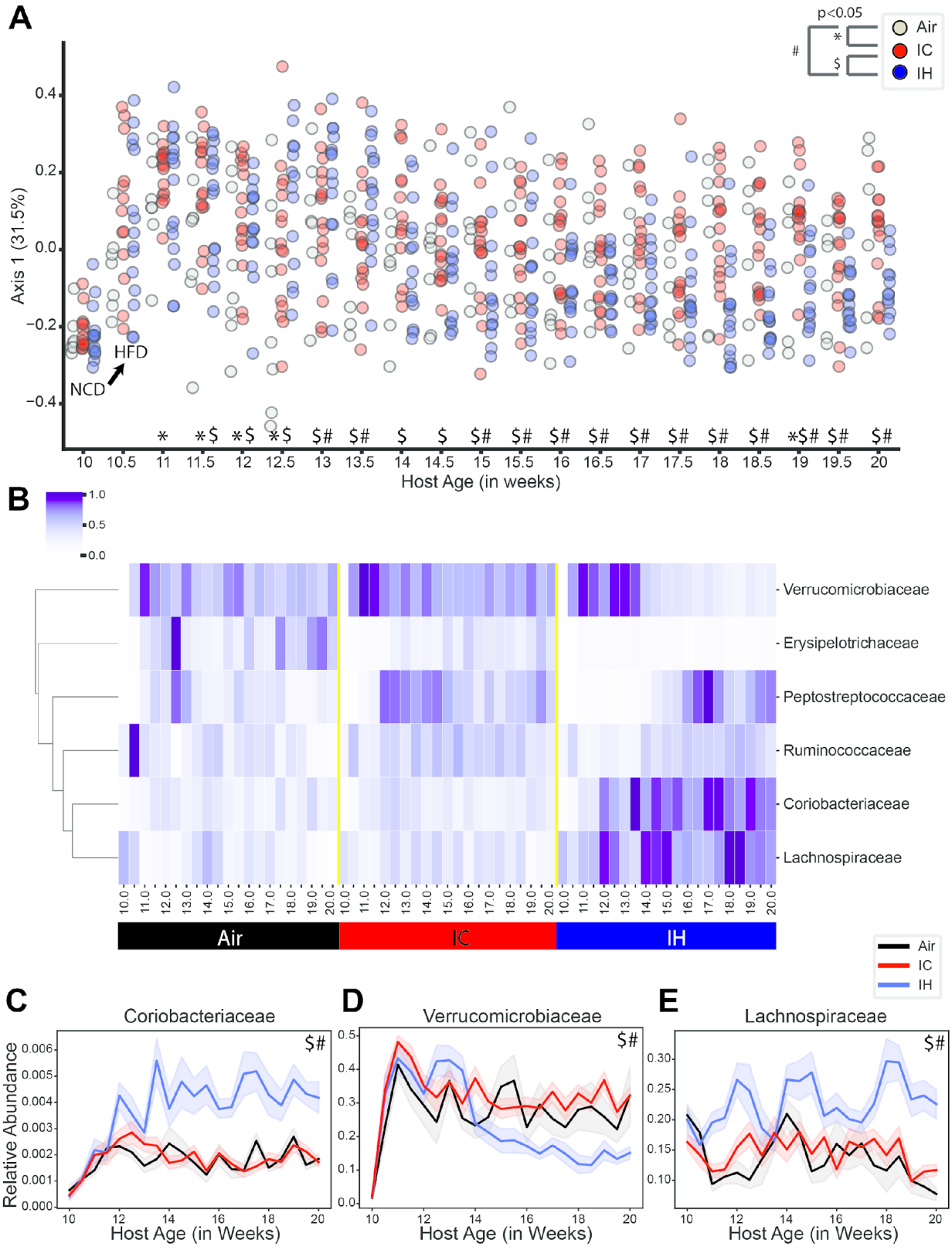

A
p<0.05
Air
IC
IH
Axis 1 (31.5%)
HFD
NCD
Host Age (in weeks)
B
Verrucomicrobiaceae
Erysipelotrichaceae
Peptostreptococcaceae
Ruminococcaceae
Coriobacteriaceae
Lachnospiraceae
Air
IC
IH
C
Coriobacteriaceae
Relative Abundance
Host Age (in Weeks)
D
Verrucomicrobiaceae
Host Age (in Weeks)
E
Lachnospiraceae
Host Age (in Weeks)

Figure A.B.1.S2. Additional Untargeted LC-MS/MS metabolomics of ApoE-/- mice on HFD during chronic 10-wk treatment. (A) Canberra PCoA of the metabolome for all measured time points. (B) Relative abundance heatmap (row normalized, set on standard scale) compared by exposure types over time. Longitudinal relative abundance values for: (C) lysophosphocholine, 1-Hexadecyl-sn-glycero-3-phosphocholine [Lyso-PAF C-16] [Level 3 identification]; (D) acylcarnitine, oleoyl L-carnitine [Level 3 identification]; (E) amino acid, tryptophan [Level 3 identification]. PERMANOVA used for statistical comparisons at the population level. Linear Mixed Effect (LME) modeling used for statistical comparisons of individual metabolites. The shaded areas in parts C-E represent standard error of the mean. Air/controls are black (n=6), IH is blue (n=12), IC is red (n=12). Statistical significance $p<0.05$, * IC vs Air, # IH vs Air and $ IH vs IC.

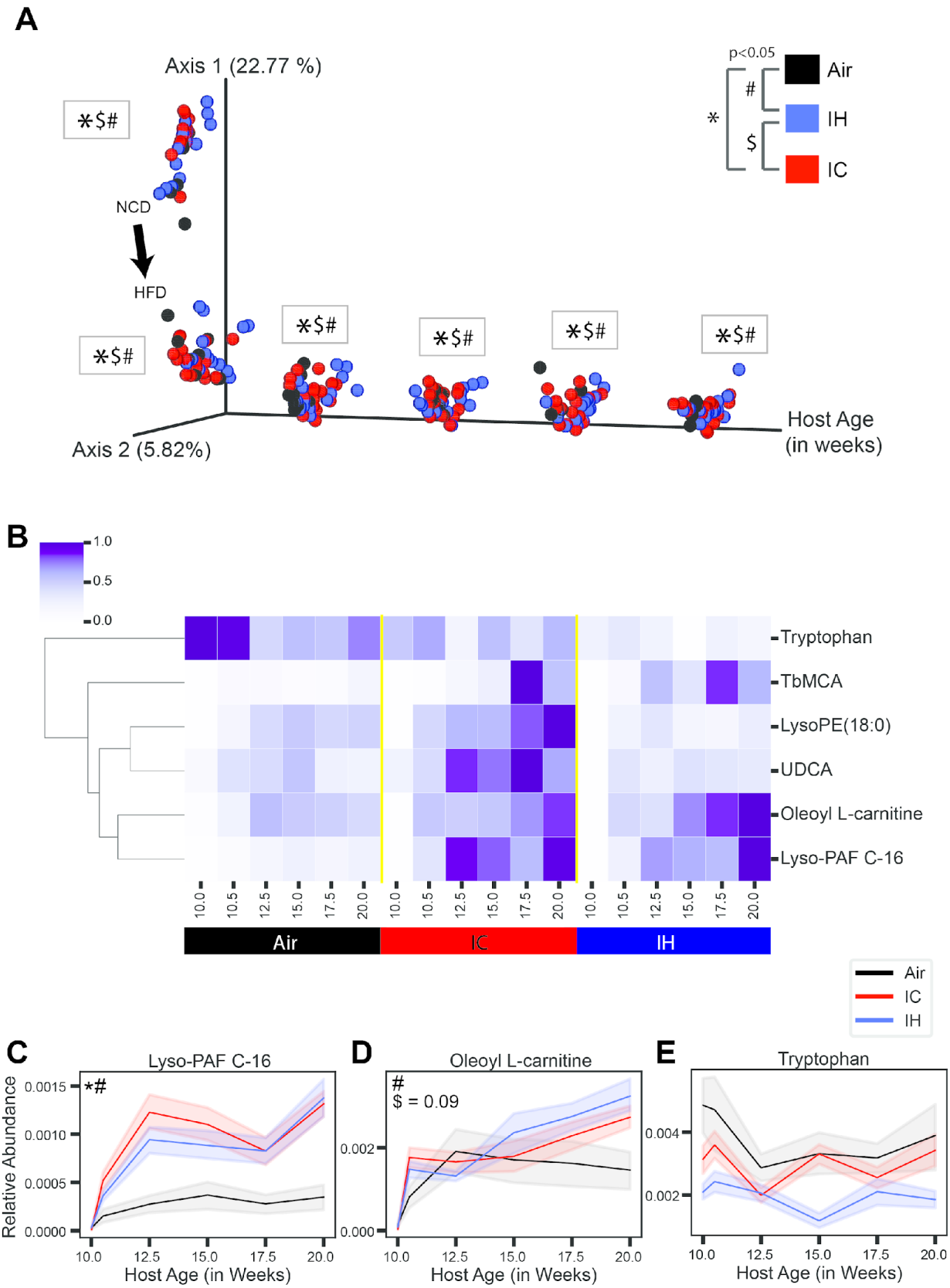
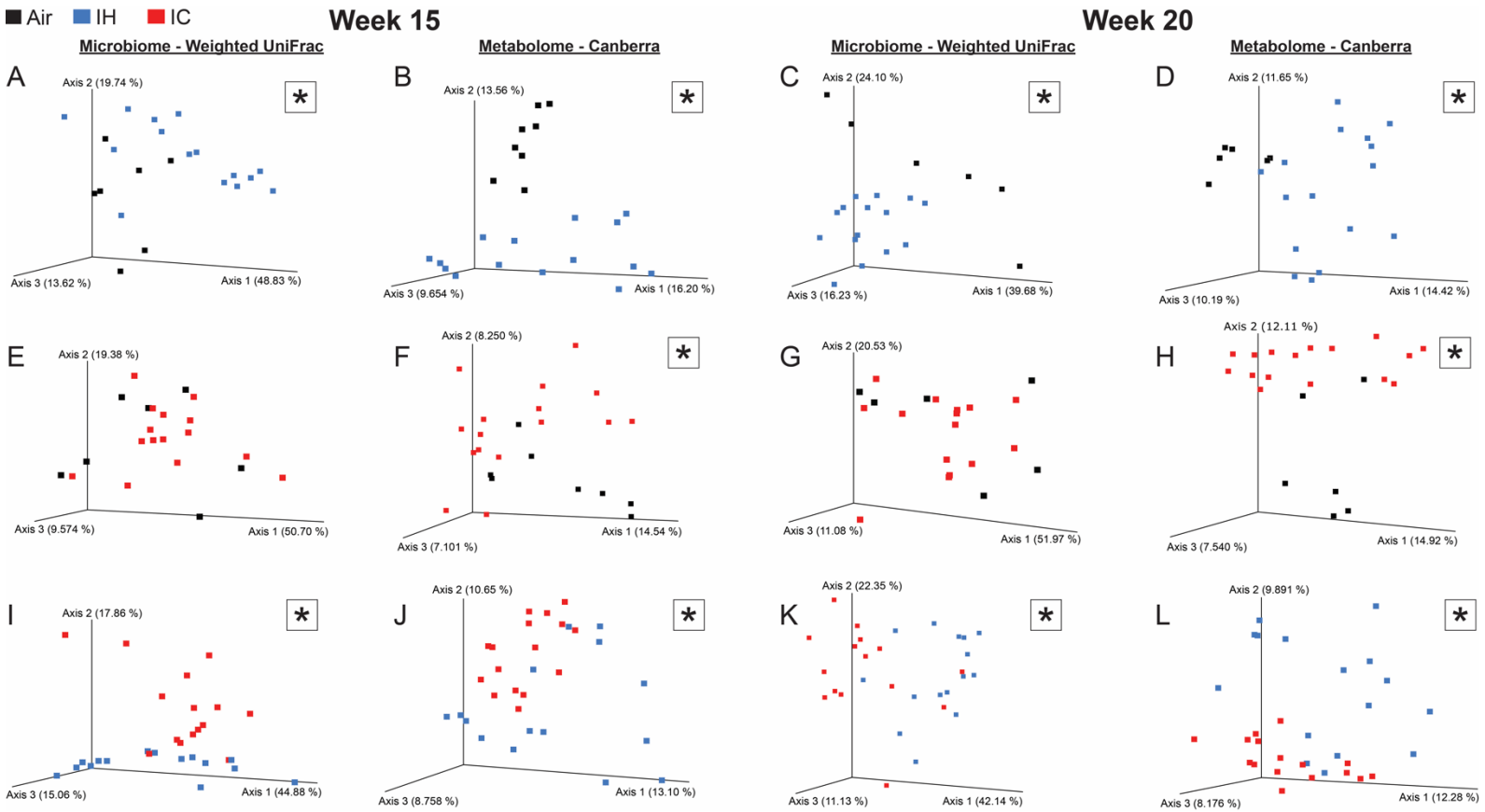

A
Axis 1 (22.77 %)
Axis 2 (5.82%)
Host Age
(in weeks)
NCD
HFD
*$#
p<0.05
Air
IH
IC
B
1.0
0.5
0.0
Tryptophan
TbMCA
LysoPE(18:0)
UDCA
Oleoyl L-carnitine
Lyso-PAF C-16
Air
IC
IH
C
Lyso-PAF C-16
Relative Abundance
Host Age (in Weeks)
D
Oleoyl L-carnitine
$ = 0.09
E
Tryptophan

Figure A.B.1.S3. Additional Microbiome and Metabolome PCoA for Weeks 15 and 20. (A) Microbiome Weighted UniFrac PCoA Week 15 Air vs IH only. (B) Metabolome Canberra PCoA Week 15 Air vs IH only. (C) Microbiome Weighted UniFrac PCoA Week 20 Air vs IH only. (D) Metabolome Canberra PCoA Week 20 Air vs IH only. (E) Microbiome Weighted UniFrac PCoA Week 15 Air vs IC only. (F) Metabolome Canberra PCoA Week 15 Air vs IC only. (G) Microbiome Weighted UniFrac PCoA Week 20 Air vs IC only. (H) Metabolome Canberra PCoA Week 20 Air vs IC only. (I) Microbiome Weighted UniFrac PCoA Week 15 IC vs IH only. (J) Metabolome Canberra PCoA Week 15 IC vs IH only. (K) Microbiome Weighted UniFrac PCoA Week 20 IC vs IH only. (L) Metabolome Canberra PCoA Week 20 IC vs IH only. Air/controls are black (n=6), IH is blue (n=12), IC is red (n=12). Statistical significance * $p<0.05$ by PERMANOVA.

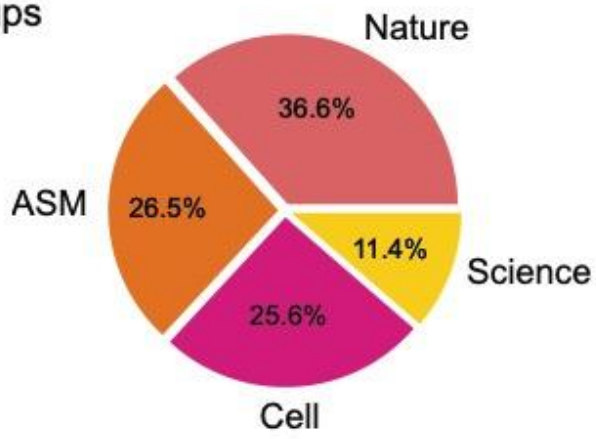

Air
IH
IC
Week 15
Week 20
Microbiome - Weighted UniFrac
Metabolome - Canberra
Microbiome - Weighted UniFrac
Metabolome - Canberra
A
Axis 2 (19.74 %)
Axis 3 (13.62 %)
Axis 1 (48.83 %)
B
Axis 2 (13.56 %)
Axis 3 (9.654 %)
Axis 1 (16.20 %)
C
Axis 2 (24.10 %)
Axis 3 (16.23 %)
Axis 1 (39.68 %)
D
Axis 2 (11.65 %)
Axis 3 (10.19 %)
Axis 1 (14.42 %)
E
Axis 2 (19.38 %)
Axis 3 (9.574 %)
Axis 1 (50.70 %)
F
Axis 2 (8.250 %)
Axis 3 (7.101 %)
Axis 1 (14.54 %)
G
Axis 2 (20.53 %)
Axis 3 (11.08 %)
Axis 1 (51.97 %)
H
Axis 2 (12.11 %)
Axis 3 (7.540 %)
Axis 1 (14.92 %)
I
Axis 2 (17.86 %)
Axis 3 (15.06 %)
Axis 1 (44.88 %)
J
Axis 2 (10.65 %)
Axis 3 (8.758 %)
Axis 1 (13.10 %)
K
Axis 2 (22.35 %)
Axis 3 (11.13 %)
Axis 1 (42.14 %)
L
Axis 2 (9.891 %)
Axis 3 (8.176 %)
Axis 1 (12.28 %)

Appendix C. Supplemental Files for Chapter 4 Microbiome Diurnal Dynamics Dominate Phenotype Effects, Enabling Replicability When Controlled

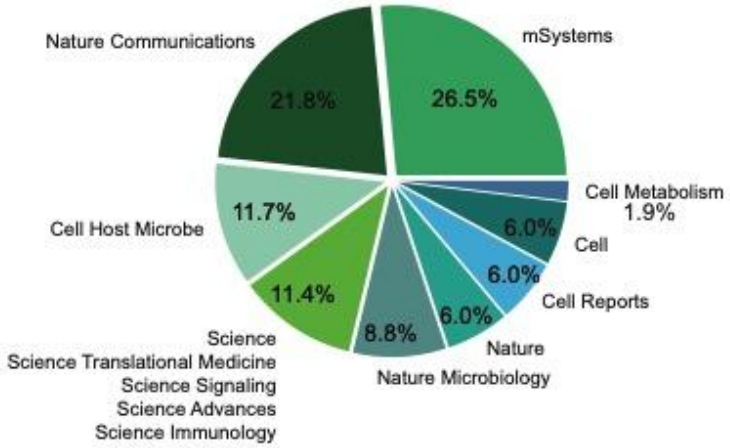
A.C.1 Supplemental Figures

Figure A.C.1.S1. Literature Review A) Of the 586 articles containing microbiome (16S or metagenomic) data, found as described in the methods section, the percentage of microbiome articles from each of the publication groups in 2019. B) The percentage of microbiome articles belonging to each individual journal in 2019. Because the numerous individual journals from Science represented low percentages individually, they were grouped together. The percentage articles where collection time was explicitly stated (yes: 8 AM, ZT4, etc.), implicitly stated (relative: “before surgery”, “in the morning”, etc.), or unstated (not provided: “daily”, “once a week”, etc.).

A Journal Groups



B Journals



C Collection Time Reported

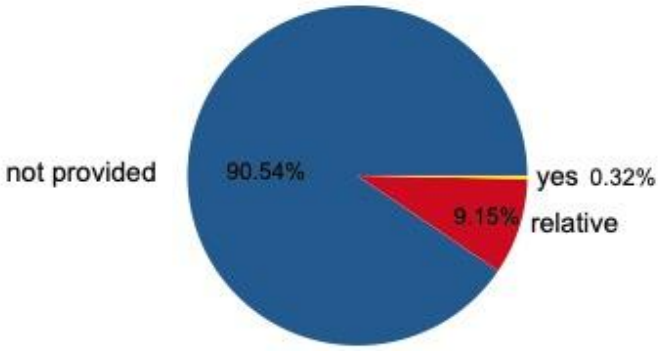


Figure A.C.1.S2. Temporal changes in BCD between NCD and MFDA) Experiment design. C57Bl/6 mice were fed NCD (control) vs MFD ad libitum for 5 weeks before fecal samples were collected for analysis. Samples were collected every 4 hours for 24 hours (N=3 mice/condition). B) BCD for fecal samples comparing NCD vs MFD over 24hrs. The dotted line is the average of all shown weighted UniFrac distances. Significance was determined using the Mann-Whitney-Wilcoxon test two-sided with Bonferroni correction. C) Heatmap of mean BCD from fecal samples collected from NCD vs MFD mice by time point over 24hrs. Highest highlighted in brown, lowest highlighted in orange.

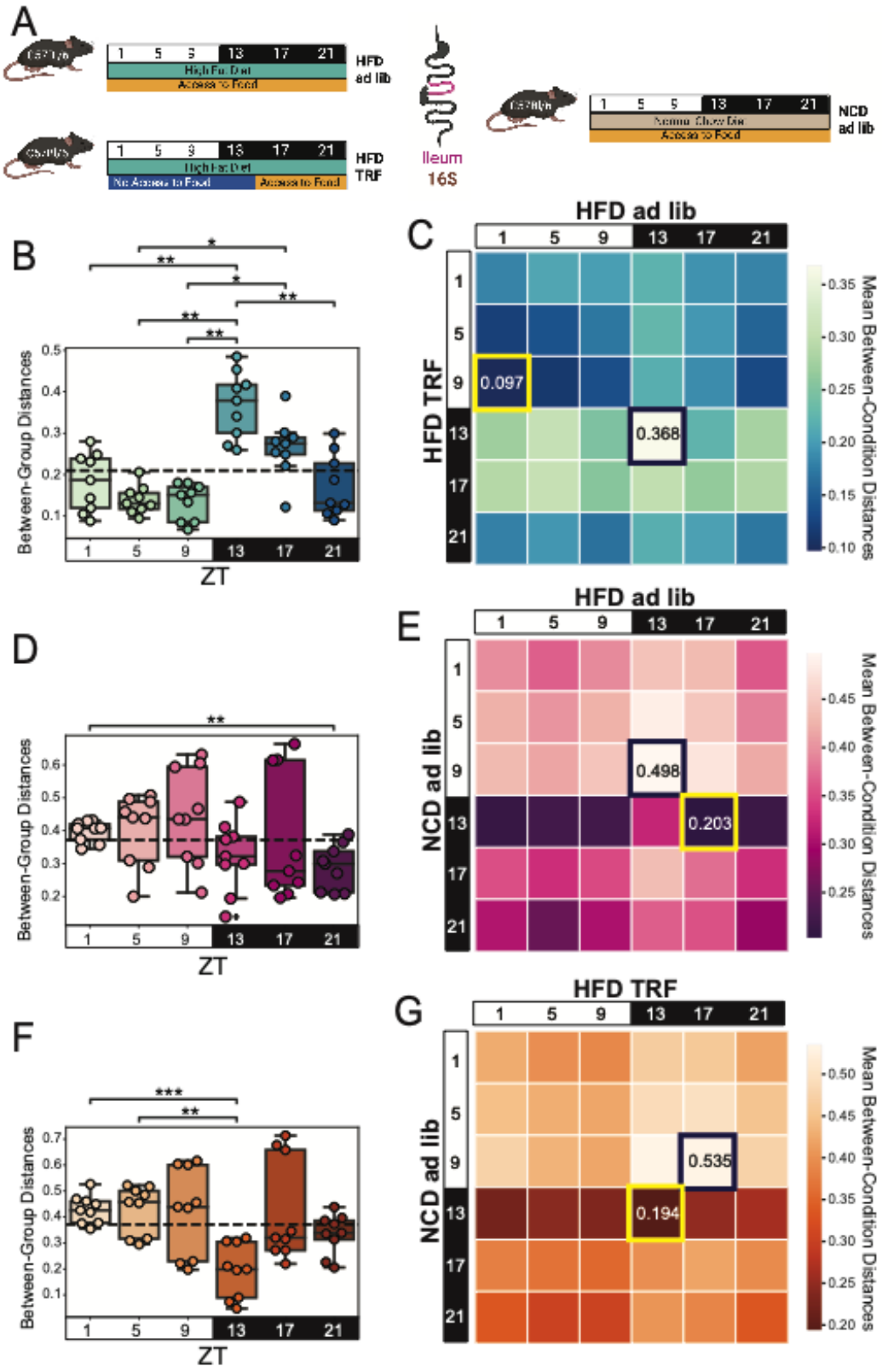

A
1 5 9 13 17 21
HFD ad lib
HFD TRF
Ileum 16S
NCD ad lib
B
Between-Group Distances
ZT
C
HFD ad lib
HFD TRF
0.097
0.368
Mean Between-Condition Distances
D
Between-Group Distances
ZT
E
HFD ad lib
NCD ad lib
0.498
0.203
Mean Between-Condition Distances
F
Between-Group Distances
ZT
G
HFD TRF
NCD ad lib
0.535
0.194
Mean Between-Condition Distances

Figure A.C.1.S3. Diet and Feeding Pattern Influence Sample Collection Time Results in the Ileum A) Experimental design. Mice used are the same as the ones in Fig 2 except this is unpublished ileal study. Mice were fed either ad libitum or TRF (ZT 13-21) access to HFD and compared to NCD ad libitum controls. After 8 weeks, ileal samples were collected every 4 hours for 24 hours (N=3 mice/condition). B) BCD for ileum samples comparing HFD ad libitum vs HFD TRF. Dotted line is the average of all shown weighted UniFrac distances. Significance was determined using a paired Mann-Whitney-Wilcoxon test two-sided with Bonferroni correction.C) Heatmap of mean BCD from ileum samples collected from NCD controls and HFD TRF mice by time point. Highest highlighted in indigo, lowest highlighted in yellow. D) BCD for ileal samples comparing NCD ad libitum vs HFD ad libitum. Dotted line is the average of all shown weighted UniFrac distances. Significance was determined using the Mann-Whitney-Wilcoxon test two-sided with Bonferroni correction. E) Heatmap of mean BCD from ileal samples collected from NCD controls and HFD TRF mice by time point. Highest highlighted in indigo, lowest highlighted in yellow. F) BCD for ileal samples comparing NCD ad libitum vs HFD TRF. Dotted line is the average of all shown weighted UniFrac distances. Significance was determined using the Mann-Whitney-Wilcoxon test two-sided with Bonferroni correction. G) Heatmap of mean BCD from ileal samples collected from NCD controls and HFD TRF mice by time point. Highest highlighted in indigo, lowest highlighted in yellow.

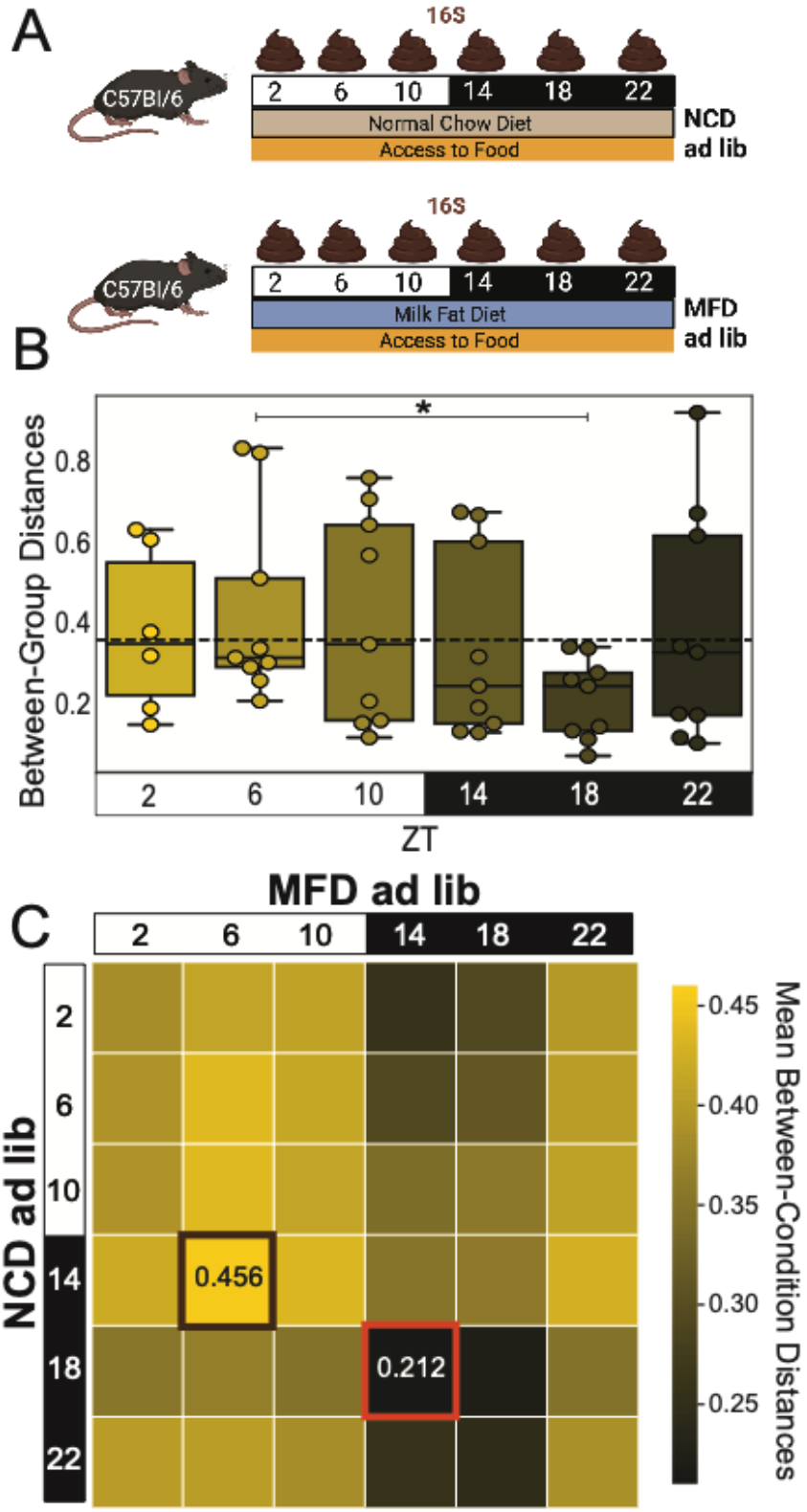
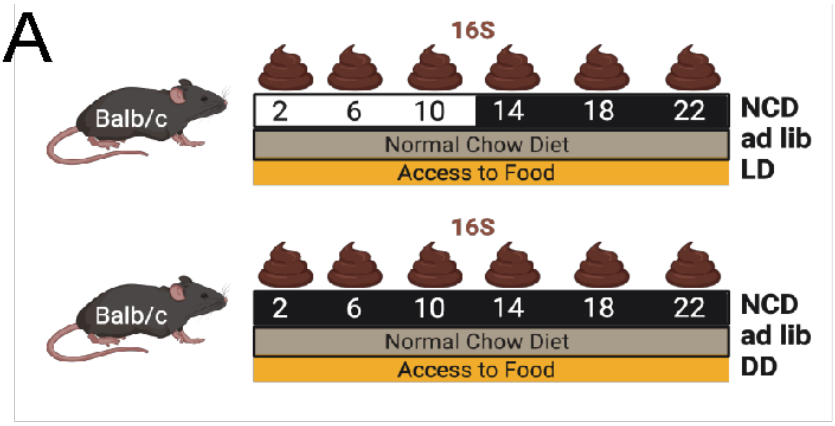
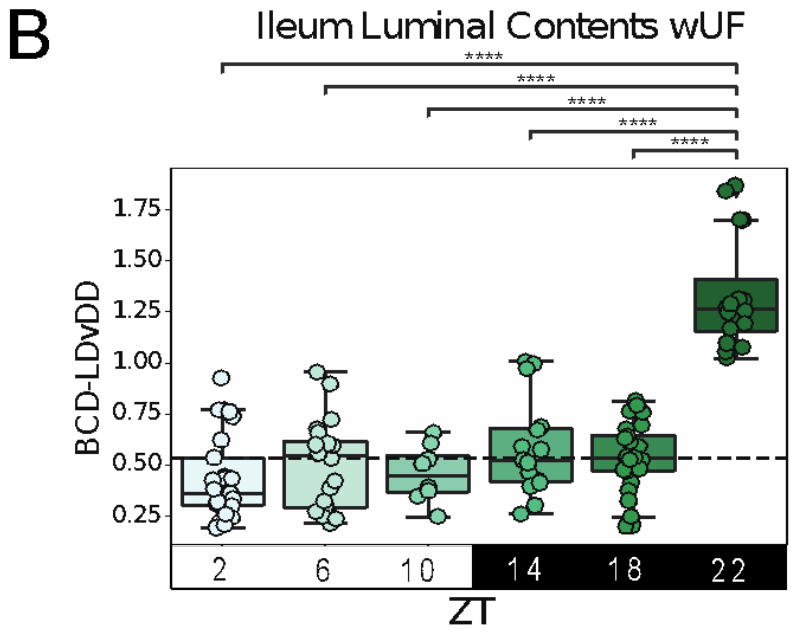

A
16S
C57Bl/6
2 6 10 14 18 22
Normal Chow Diet
Access to Food
NCD
ad lib
16S
C57Bl/6
2 6 10 14 18 22
Milk Fat Diet
Access to Food
MFD
ad lib
B
Between-Group Distances
0.8
0.6
0.4
0.2
*
2 6 10 14 18 22
ZT
C
MFD ad lib
2 6 10 14 18 22
NCD ad lib
2
6
10
14
18
22
0.456
0.212
0.45
0.40
0.35
0.30
0.25
Mean Between-Condition Distances

Figure A.C.1.S4. Irregular differences in diurnal rhythm patterns leads to generally minor shifts in BCD when comparing LD vs DD mice. A) Experimental design. Balb/c mice were fed NCD ad libitum under 0:24 L:D (24hr darkness, DD) experimental conditions and compared to 12:12 L:D (LD) control conditions. After 2 weeks, mice from each group were euthanized every 4 hours for 24 hours (N=4-5 mice/condition) and samples were collected from the proximal small intestine ("jejunum") and distal small intestine ("ileum") contents. B) BCD for luminal contents of proximal small intestine samples comparing LD to DD mice. Dotted line is the average of all shown weighted UniFrac distances. Significance was determined using a paired Mann-Whitney-Wilcoxon test two-sided with Bonferroni correction. C) BCD for luminal contents of distal small intestine samples comparing LD to DD mice. Dotted line is the average of all shown weighted UniFrac distances. Significance was determined using a paired Mann-Whitney-Wilcoxon test two-sided with Bonferroni correction.


A
16S
Balb/c
2
6
10
14
18
22
Normal Chow Diet
Access to Food
NCD
ad lib
LD
16S
Balb/c
2
6
10
14
18
22
Normal Chow Diet
Access to Food
NCD
ad lib
DD


B
Ileum Luminal Contents wUF
****
****
****
****
****
BCD-LDvDD
1.75
1.50
1.25
1.00
0.75
0.50
0.25
2
6
10
14
18
22
ZT

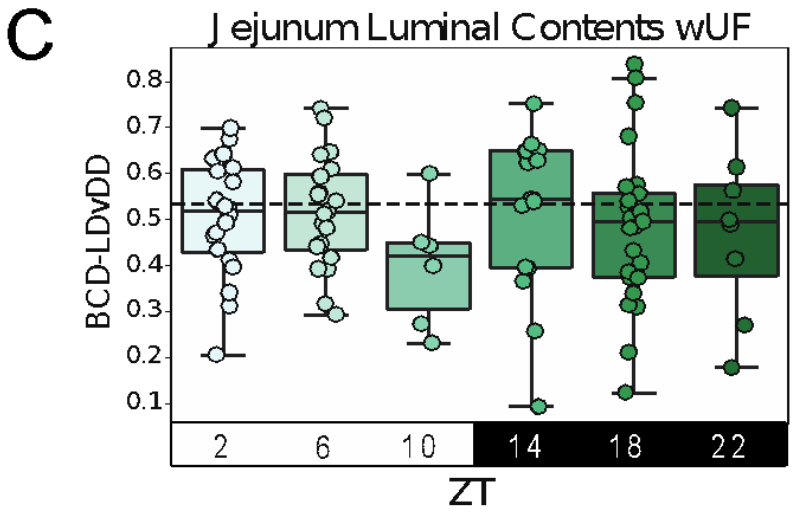

C
Jejunum Luminal Contents wUF
BCD-LDvDD
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
2
6
10
14
18
22
ZT

Figure A.C.1.S5. Line plot of the same values presented in Fig 3B. The shaded region represents the standard error of the mean. The dotted line is the average of all of the weighted UniFrac distances used to calculate this plot. Some of the shifts seen between early and late values may be more easily visualized in this format. Early data: MetaCycle, meta2d/LS method, p=0.0017, amplitude=0.215, period=22.3, adjphase=18.0; Late Data: MetaCycle, meta2d/LS method, p=0.0128, amplitude=0.198, period=25.7, adjphase=16.2.

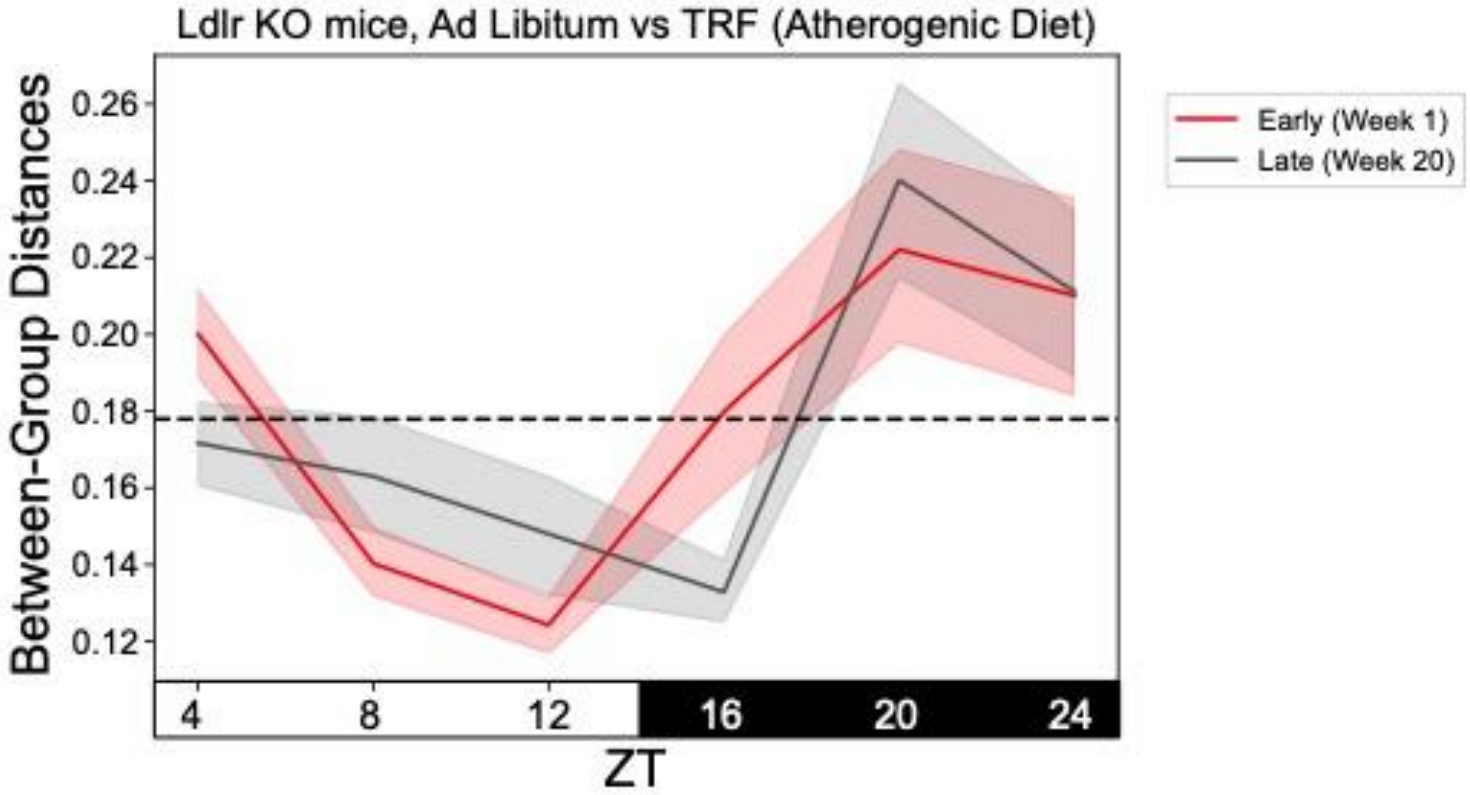

Ldlr KO mice, Ad Libitum vs TRF (Atherogenic Diet)
Between-Group Distances
0.26
0.24
0.22
0.20
0.18
0.16
0.14
0.12
4
8
12
16
20
24
ZT
Early (Week 1)
Late (Week 20)

Figure A.C.1.S6. Human data also shows that a non-continuous intervention affects beta diversity distances over the course of a study. Experimental design: The patients underwent 5 days of dietary intervention, either plant or animal-based (N=10 humans/condition): 9/10 patients underwent both dietary interventions after a 1 month wash-out period, 1/10 patients only underwent a single intervention. See reference (David et al., 2014). A) Weighted UniFrac β-diversity violin plot using between-group distances for plant and animal dietary interventions. Each line on the violin plot is a sample value. The dotted line is the average of all of the weighted UniFrac distances from the time points farthest from the intervention (-4.0 and 10.0). The shaded area represents time points that are not significantly different from each other, except as noted. Significance was determined using the Mann-Whitney-Wilcoxon test two-sided with Bonferroni correction. Notation: light gray line = $p<0.05$; medium gray line = $p<0.01$; black line = $p<0.0001$. B) Mean weighted UniFrac β-diversity distance heatmap using values calculated between plant and animal dietary interventions by time point. Highest value highlighted in purple, lowest highlighted in pink.

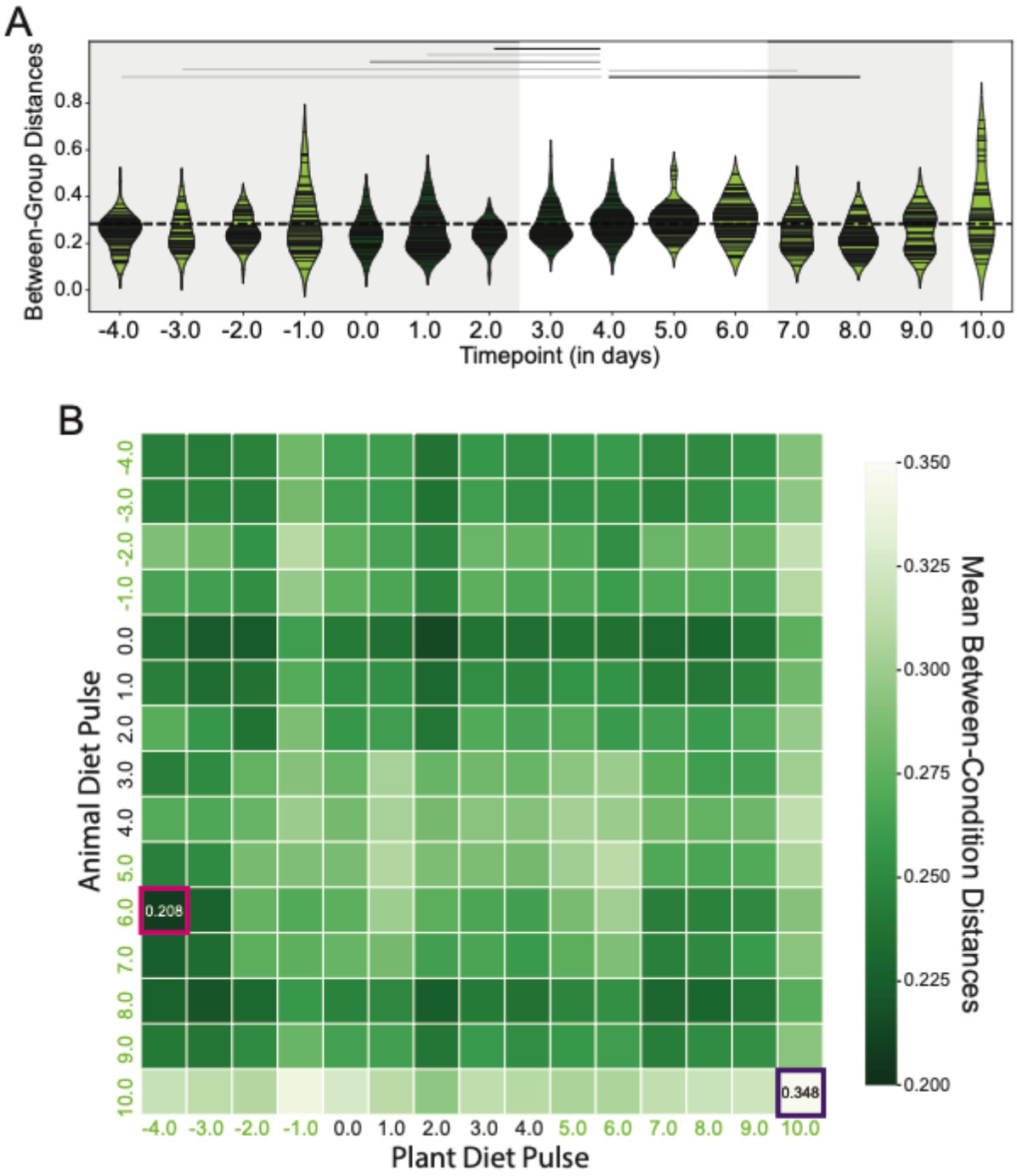
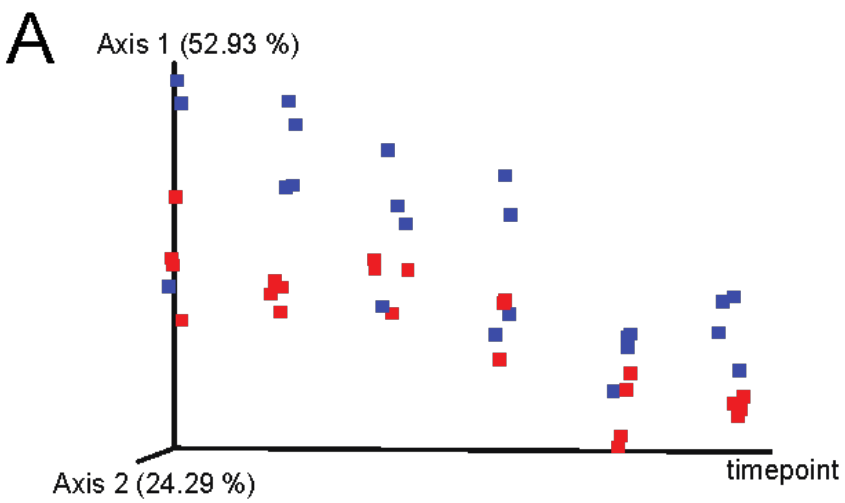
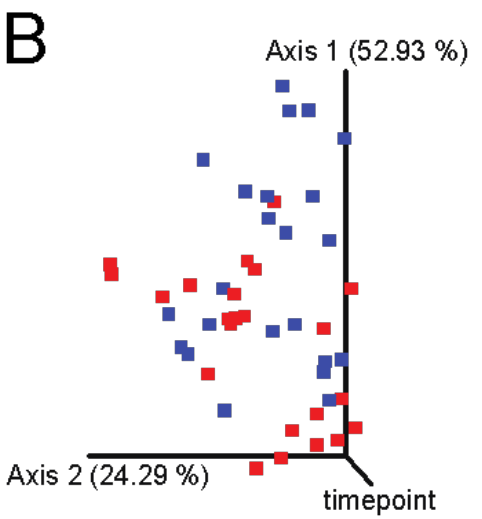
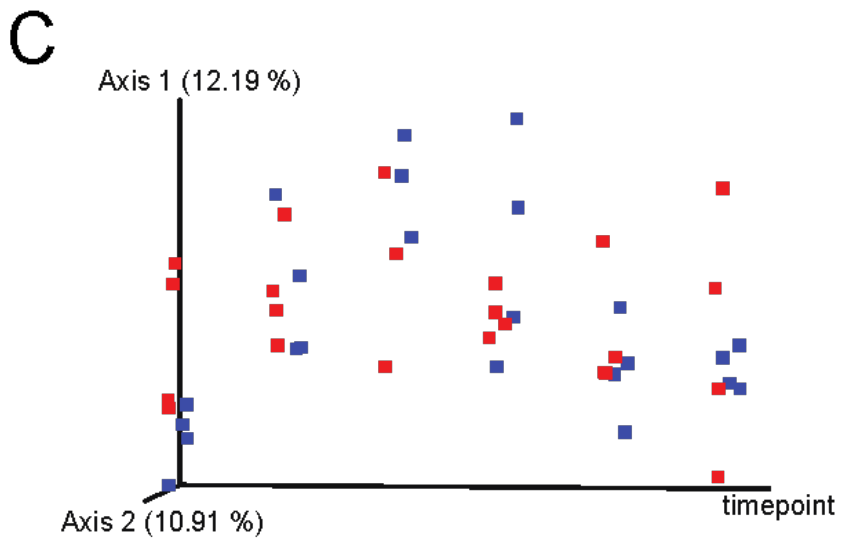
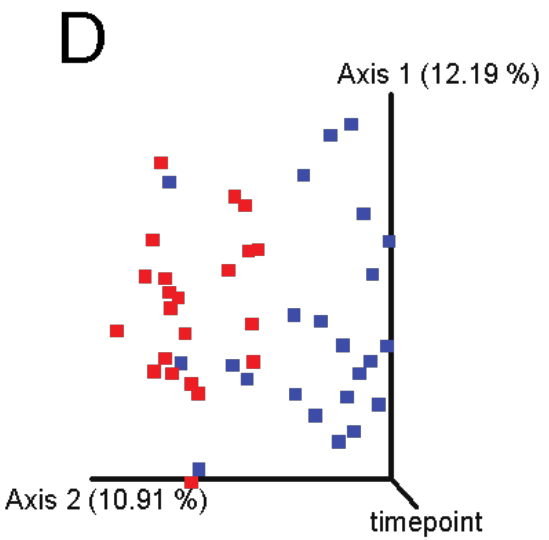

A
Between-Group Distances
0.0
0.2
0.4
0.6
0.8
-4.0 -3.0 -2.0 -1.0 0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0
Timepoint (in days)
B
Animal Diet Pulse
-4.0 -3.0 -2.0 -1.0 0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0
0.208
0.348
-4.0 -3.0 -2.0 -1.0 0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0
Plant Diet Pulse
0.350
0.325
0.300
0.275
0.250
0.225
0.200
Mean Between-Condition Distances

Figure A.C.1.S7. Diurnal IHC Weighted UniFrac PCoA Plots. A) Weighted UniFrac PCoA lateral view. Timepoints as one axis. B)Weighted UniFrac PCoA stacked view. C) Canberra PCoA lateral view. Timepoints as one axis. D) Canberra PCoA stacked view.


A
Axis 1 (52.93 %)
Axis 2 (24.29 %)
timepoint


B
Axis 1 (52.93 %)
Axis 2 (24.29 %)
timepoint


C
Axis 1 (12.19 %)
Axis 2 (10.91 %)
timepoint


D
Axis 1 (12.19 %)
Axis 2 (10.91 %)
timepoint

Appendix D. Supplemental Files for Chapter 5 Engineered Native Bacteria are More Efficient at Engraftment and Colonization of a Host and Require Fewer Mutations for Maintenance Long Term

A.D.1. Supplementary Figures

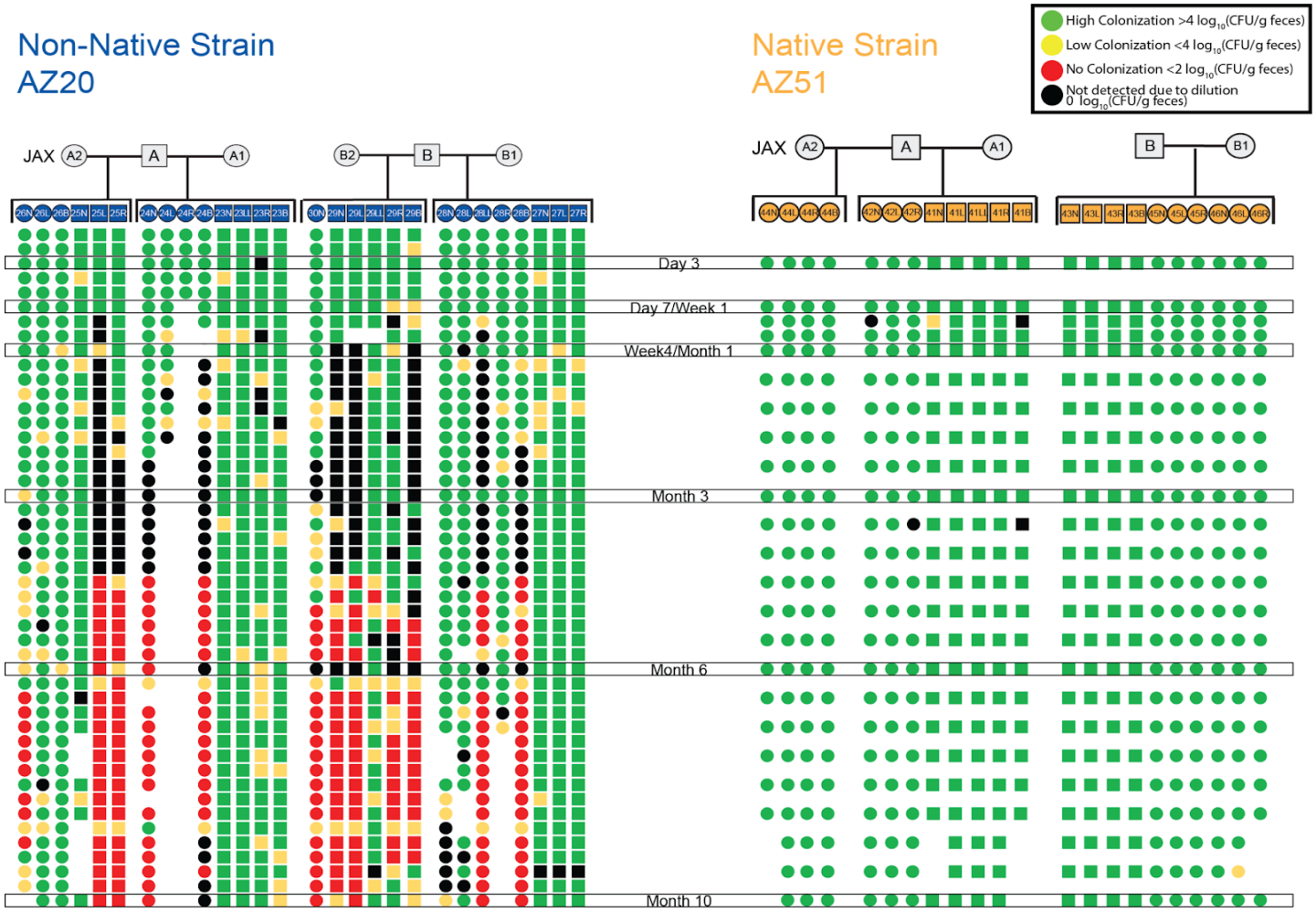


Figure A.D.1.S2 Colonization levels over time by Genealogy.
Both strain cohorts compared side-by-side by time point and genealogy. The number indicates the cage id, the letter indicates which ear had a hole punch. Circles indicate females, squares indicate males. High colonization: greater than 4 log10(CFU/g feces) is indicated in green, Low colonization: less than 4 but higher than 2 log10(CFU/g feces) is indicated in yellow, not colonized: less than 2 log10(CFU/g feces) is indicated in red, no detected due to dilution: 0 log10(CFU/g feces) is indicated in black. Empty spaces indicate sample not collected.

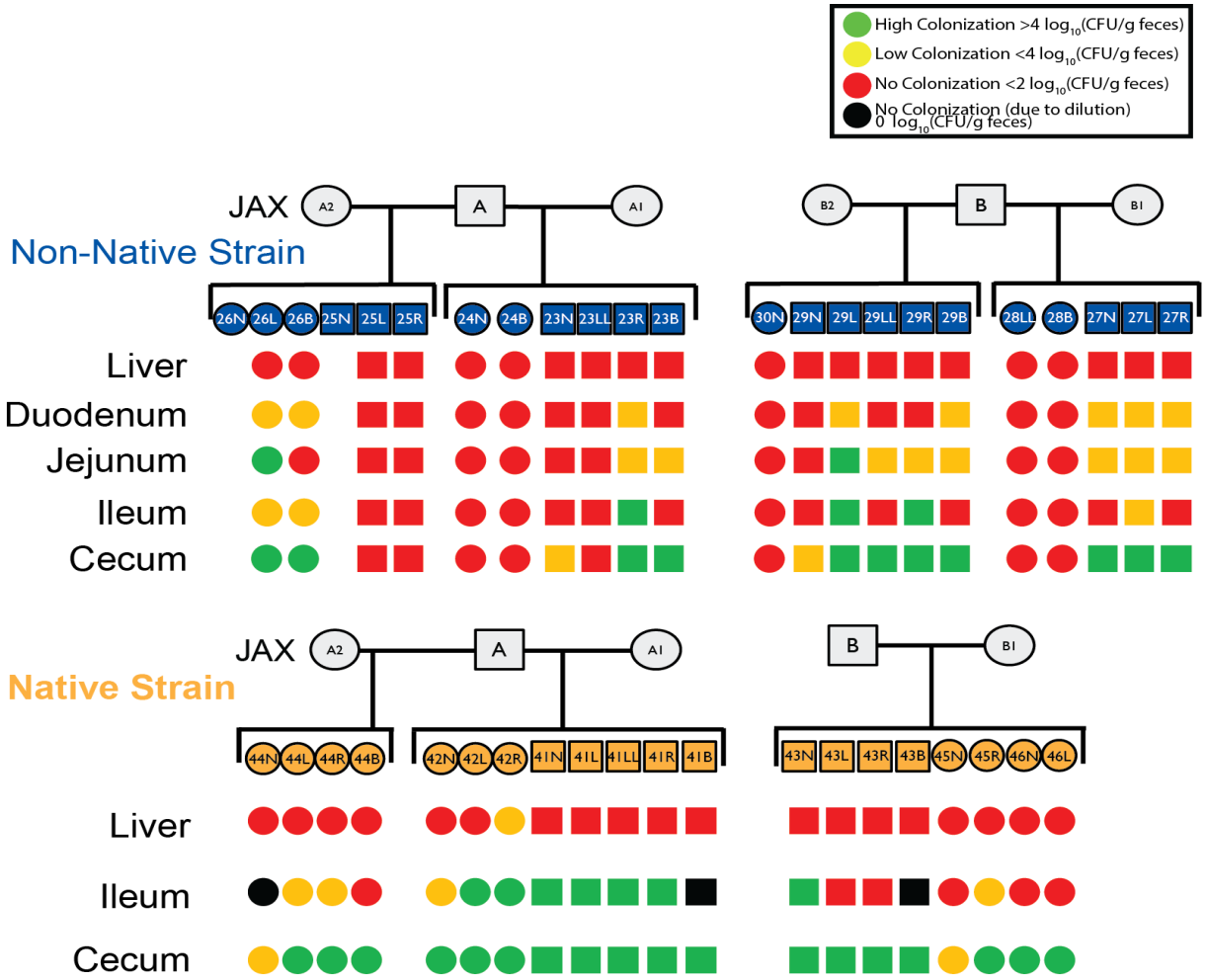


Figure A.D.1.S3 Colonization of Gastrointestinal Tract.
At the end of the study, samples from each of the indicated organs (liver, duodenum, jejunum, cecum) were homogenized and plated on selective media and counted. High colonization: greater than 4 log10(CFU/g feces) is indicated in green, Low colonization: less than 4 but higher than 2 log10(CFU/g feces) is indicated in yellow, not colonized: less than 2 log10(CFU/g feces) is indicated in red, no detected due to dilution: 0 log10(CFU/g feces) is indicated in black. Empty spaces indicate sample not collected.

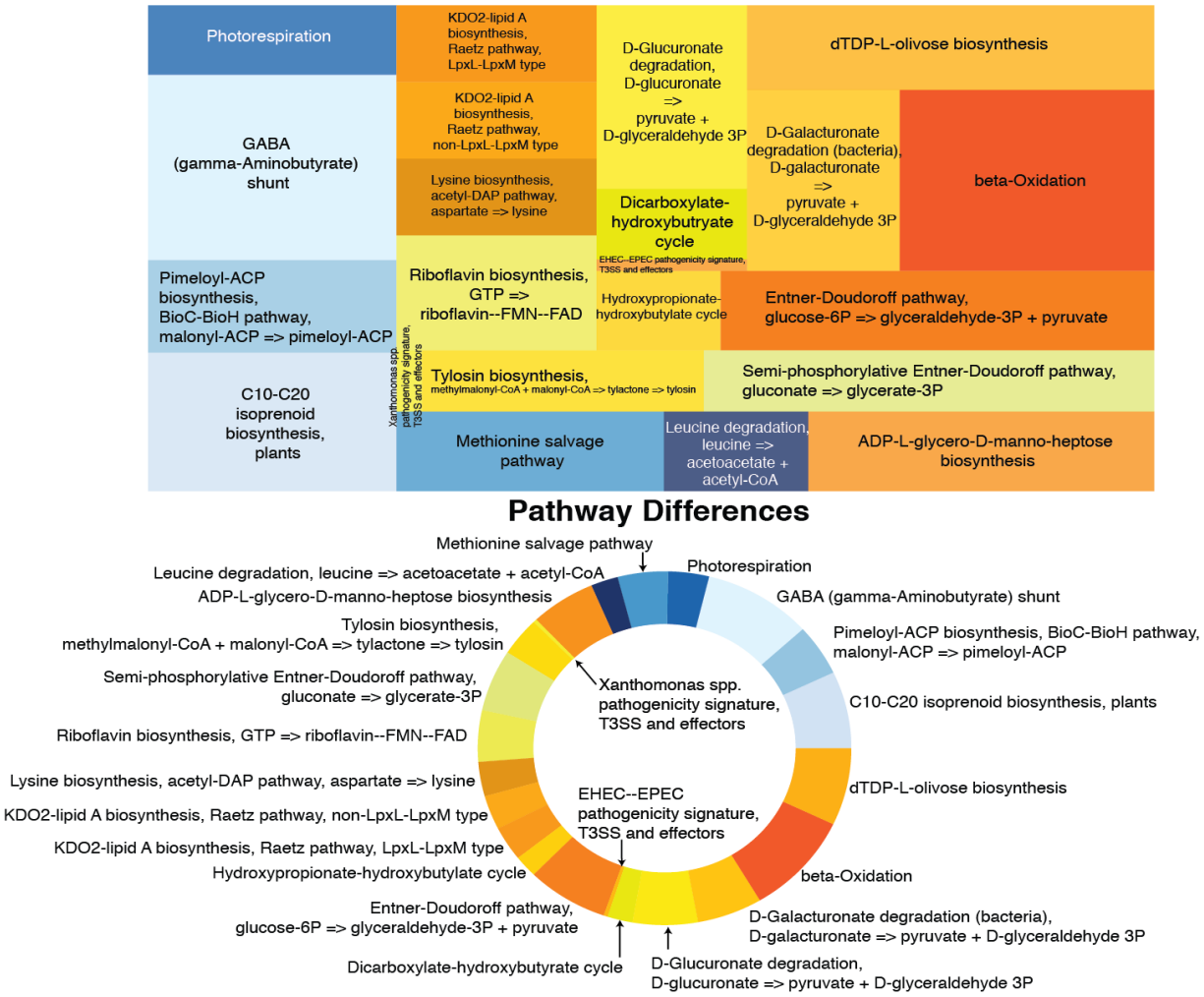


Figure A.D.1.S4: Metabolic Pathway Completeness Comparison. MicrobeAnnotator was used to refine prokka annotations to include KEGG pathways. Blue/cool colors indicate that pathway was more complete for the non-native strain, gold/warm colors indicate that pathway was more complete native strain. The size of the boxes or circular sections indicate the magnitude of the differences between the strains. The mean of all isolates from all timepoints was used, since no significant differences between timepoints were detected within a strain cohort.

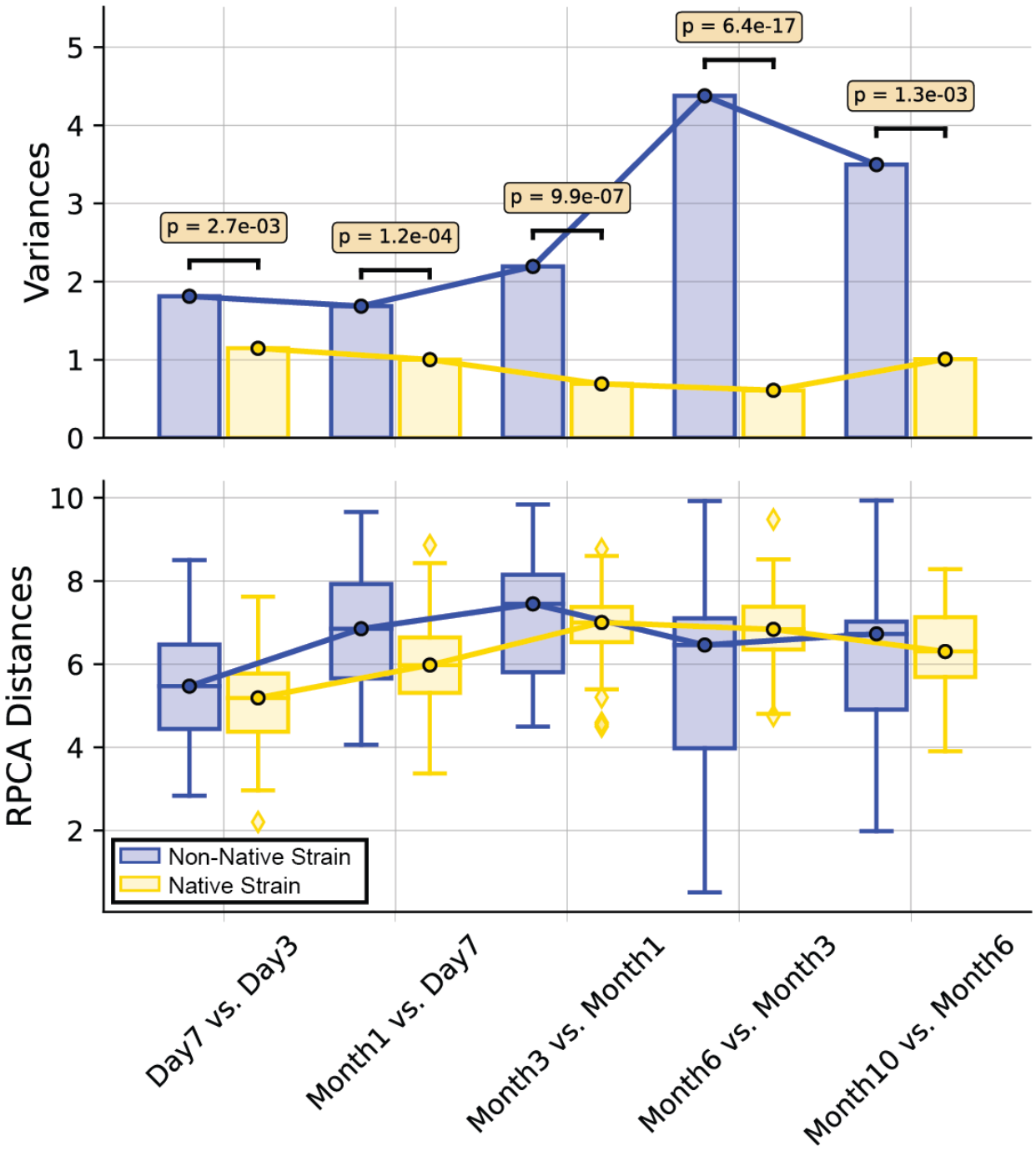


Figure A.D.1.S5: RPCA Analysis. RPCA beta diversity distances were calculated using the collapsed count table described in the methods section. The bottom section has the within group distances between subsequent time points compared between strain cohorts. Top section has within group between subsequent time point variances. Pairwise PERMANOVA used for statistical significance.

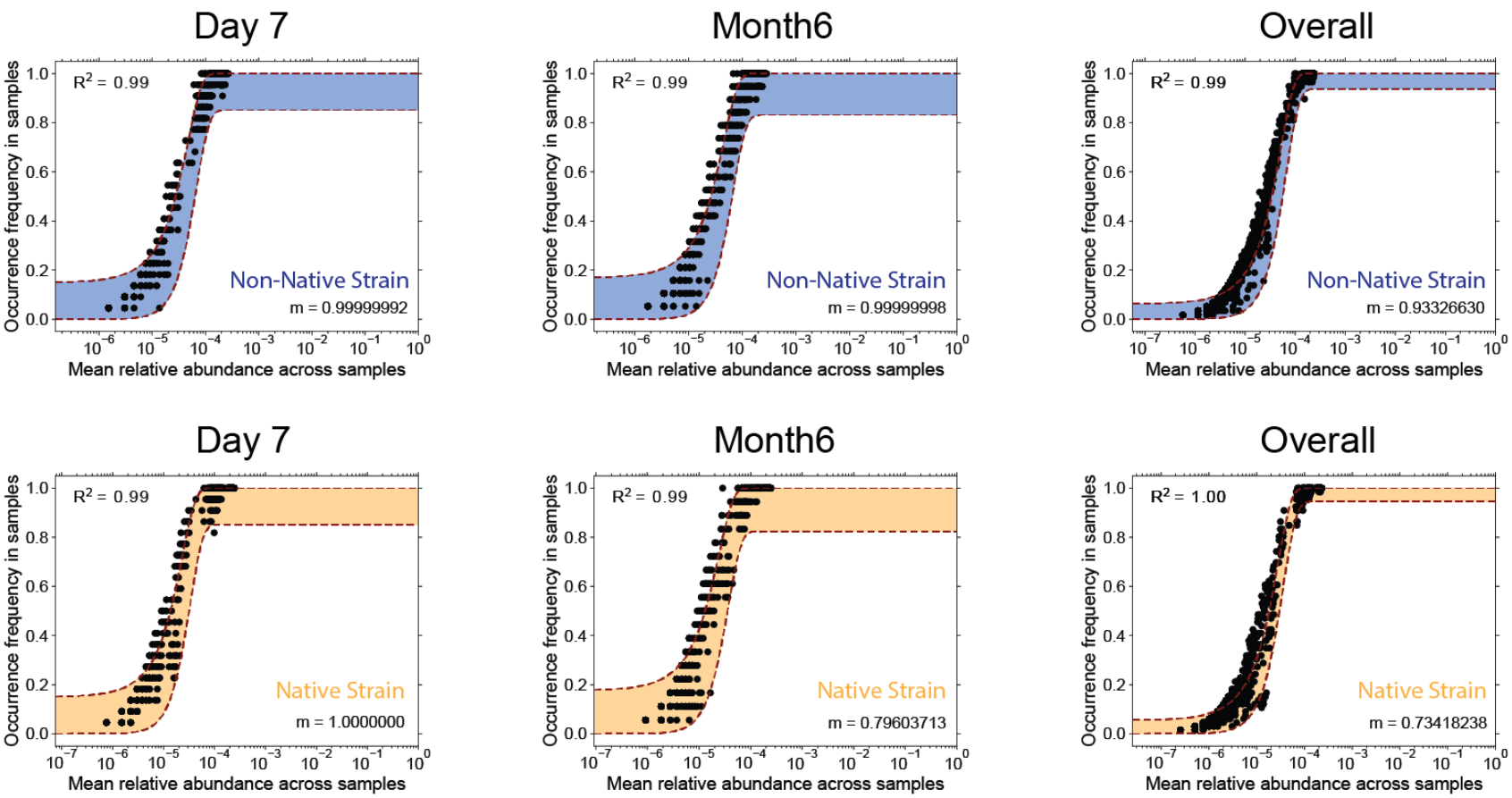


Figure A.D.1.S6: Neutral (Genetic Drift) Modeling. All modeling performed on collapsed count table. A) Modeling performed on non-native strain cohort samples from Day 7. B) Modeling performed on non-native strain cohort samples from Month 6. C) Modeling performed on non-native strain cohort samples from all timepoints. D) Modeling performed on native strain cohort samples from Day 7. E) Modeling performed on native strain cohort samples from Month 6. F) Modeling performed on native strain cohort samples from all timepoints. Notation: R^2 values close to 1 indicate good model fit, likely Genetic Drift; m-values indicate migration percentage, values close to 1 indicate fluctuating genes come from outside the *E.coli* genome.

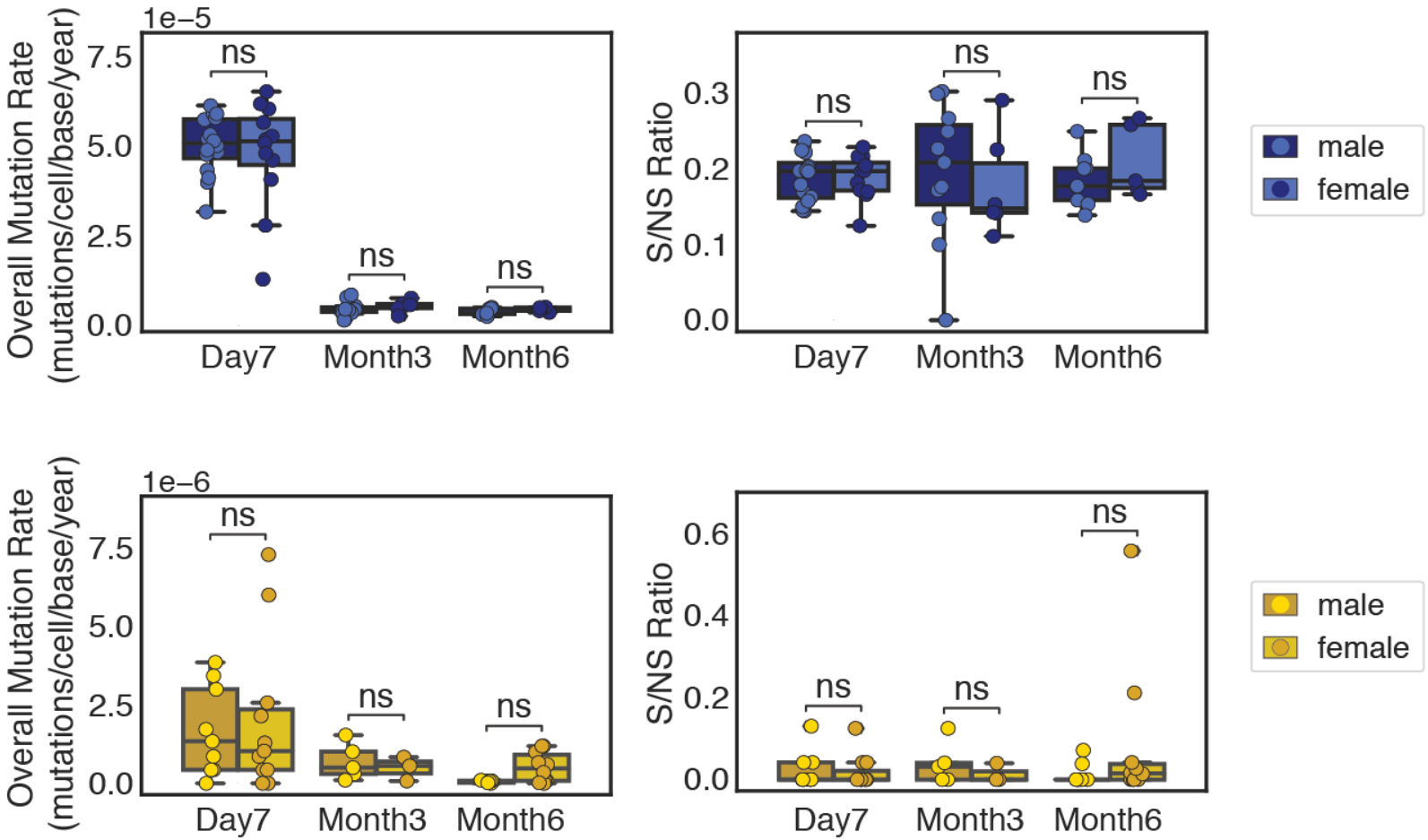


Figure A.D.1.S7: SNP Mutations by Sex. A) Overall mutation rate, calculated as SNP mutations detected by Snippy [ref] per cell per base per year, for the non-native cohort at the three major timepoints and grouped by sex. B) Synonymous/Non-Synonymous ratio for the non-native cohort at the three major timepoints grouped by sex. C) Overall mutation rate, calculated as SNP mutations detected per cell per base per year, for the native cohort at the three major timepoints and grouped by sex. D) Synonymous/Non-Synonymous ratio for the native cohort at the three major timepoints grouped by sex. A two-sided Mann-Whitney-Wilcoxon test with Bonferroni correction was used to test for significance between cohorts. Notation: blue for non-native strain, gold for native strain; ns = p>0.05.

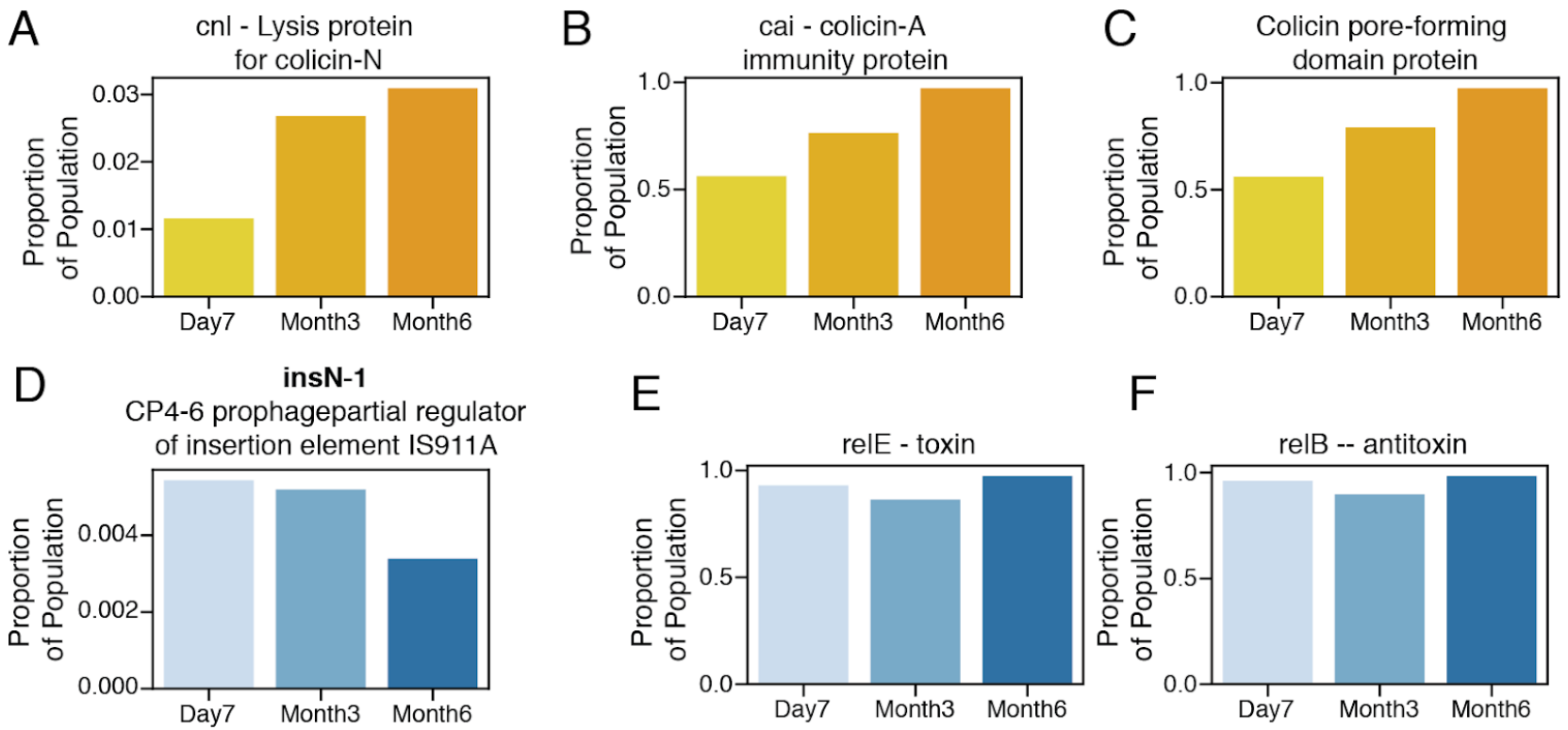


Figure A.D.1.S8: Additional plasmid genes of interest. All graphs indicated changes in the proportion of genes present in the respective strain cohort isolate populations. A) cnl prevalence in the native strain cohort across the three major time points. B) cai prevalence in the native strain cohort across the three major time points. C) Gene for a colicin pore-forming domain protein prevalence in the native strain cohort across the three major time points. D) insN-1 prevalence in the non-native strain cohort across the three major time points. E) relE (toxin of the toxin/anti-toxin system) prevalence in the non-native strain cohort across the three major time points. F) relB (anti-toxin of the toxin/anti-toxin system) prevalence in the non-native strain cohort across the three major time points.

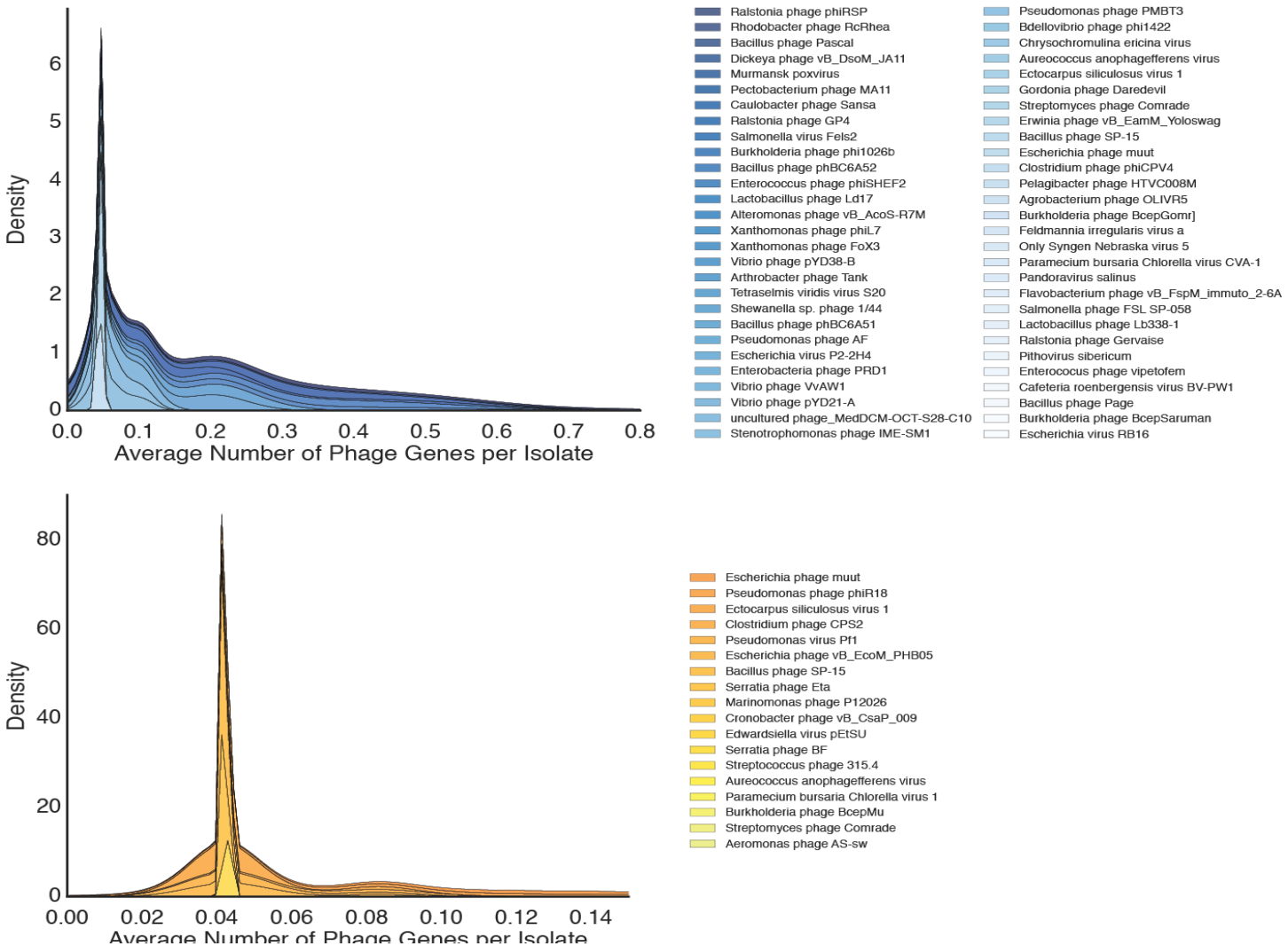


Figure A.D.1.S9: Minor phage. A) Density plot of identified phage genes in the non-native cohort that were not originally present in the reference genome. B) Density plot of identified phage genes in the native cohort that were not originally present in the reference genome.

A.D.2. Supplemental Tables

Table A.D.2.S1. Native isolates that did not pass quality control. Contains checkM, quast, and sourmash results.

sample-id	Marker lineage	# genomes	# markers	# marker sets	Completeness	Contamination	Strain heterogeneity
23LL_Day3_12_S89	Bacteria	5449	104	58	100	107.84	7.83
23LL_Day3_15_S18	Bacteria	5449	104	58	100	100	0
23LL_Day3_8_S57	Bacteria	5449	104	58	100	99.81	2.78
23LL_Day7_13_S95	Bacteria	5449	104	58	69.67	2.82	100
23LL_Month1_24_S186	Bacteria	5449	104	58	100	100	0
24B_Day7_8_S8	Bacteria	5449	104	58	32.84	0	0
24N_Day7_23_S1316	Bacteria	5449	104	58	0	0	0
25L_Day7_14_S108	Bacteria	5449	104	58	0	0	0
25L_Day7_22_S1088	Bacteria	5449	104	58	0	0	0
25N_Day3_10_S75	Bacteria	5449	104	58	100	58.06	7.14
25N_Day3_23_S84	Bacteria	5449	104	58	100	50	31.46
25N_Day3_4_S27	Bacteria	5449	104	58	100	101.38	0.95
25N_Day3_7_S51	Bacteria	5449	104	58	100	99.66	0
25N_Day3_9_S67	Bacteria	5449	104	58	100	18.97	0
25N_Month1_21_S164	Bacteria	5449	104	58	100	97.93	0
25N_Month1_2_S107	Bacteria	5449	104	58	100	113.79	15.33
25N_Month3_16_S257	Bacteria	5449	104	58	98.28	0	0
26B_Month3_23_S295	Bacteria	5449	104	58	3.51	0	0
26L_Day7_3_S86	Bacteria	5449	104	58	36.13	0	0
26N_Day7_14_S1104	Bacteria	5449	104	58	0	0	0

A.D.2.S1. Non-native isolates that did not pass quality control. *Continued*

sample-id	Marker lineage	# genomes	# markers	# marker sets	Completeness	Contamination	Strain heterogeneity
26N_Day7_19_S1109	Bacteria	5449	104	58	82	3.76	100
27L_Day7_10_S1024	Bacteria	5449	104	58	93.65	19.46	95
27N_Day7_10_S99	Bacteria	5449	104	58	6.35	0	0
27N_Day7_11_S99	Bacteria	5449	104	58	82.29	7.05	100
27N_Month6_4_S36	Bacteria	5449	104	58	0	0	0
27R_Month3_14_S328	Bacteria	5449	104	58	88.64	0.16	0
28LL_Day7_8_S1054	Bacteria	5449	104	58	95.69	6.9	100
28LL_Month3_14_S4	Bacteria	5449	104	58	100	0	0
28LL_Month3_17_S406	Bacteria	5449	104	58	100	0	0
28LL_Month3_19_S408	Bacteria	5449	104	58	100	0	0
28LL_Month3_21_S4	Bacteria	5449	104	58	100	0	0
28LL_Month3_24_S41	Bacteria	5449	104	58	100	0	0
28N_Month1_4_S125	Bacteria	5449	104	58	0	0	0

A.D.2.S1. Non-native isolates that did not pass quality control. *Continued*

sample-id	# contigs (>= 0 bp)	# contigs (>= 1000 bp)	# contigs (>= 5000 bp)	# contigs (>= 10000 bp)	# contigs (>= 25000 bp)	# contigs (>= 50000 bp)
23LL_Day3_12_S89	1592	1576	392	123	16	4
23LL_Day3_15_S18	325	317	251	176	92	36
23LL_Day3_8_S57	1910	1905	385	55	1	1
23LL_Day7_13_S95	482	449	244	151	35	1
23LL_Month1_24_S186	105	92	65	58	48	34
24B_Day7_8_S8	356	322	63	13	0	0
24N_Day7_23_S1316	1	1	1	1	1	0
25L_Day7_14_S108	2	2	1	0	0	0
25L_Day7_22_S1088	3	3	0	0	0	0
25N_Day3_10_S75	2181	2176	45	18	11	6
25N_Day3_23_S84	1436	1436	52	15	3	2
25N_Day3_4_S27	124	110	80	68	57	41
25N_Day3_7_S51	165	98	65	53	43	31
25N_Day3_9_S67	78	75	54	35	15	3
25N_Month1_21_S164	466	461	353	248	83	11
25N_Month1_2_S107	1072	1070	480	209	35	7
25N_Month3_16_S257	59	41	30	27	20	13
26B_Month3_23_S295	156	137	10	0	0	0

A.D.2.S1. Non-native isolates that did not pass quality control. *Continued*

sample-id	# contigs (>= 0 bp)	# contigs (>= 1000 bp)	# contigs (>= 5000 bp)	# contigs (>= 10000 bp)	# contigs (>= 25000 bp)	# contigs (>= 50000 bp)
26L_Day7_3_S86	492	459	151	49	1	0
26N_Day7_14_S1104	2	2	0	0	0	0
26N_Day7_19_S1109	403	383	213	136	39	7
27L_Day7_10_S1024	669	573	264	145	42	1
27N_Day7_10_S99	192	159	14	1	0	0
27N_Day7_11_S99	430	401	215	149	47	6
27N_Month6_4_S36	10	9	1	0	0	0
27R_Month3_14_S328	1160	1135	287	57	0	0
28LL_Day7_8_S1054	304	299	184	136	54	15
28LL_Month3_14_S4	104	77	48	42	30	19
28LL_Month3_17_S406	101	75	47	42	30	19
28LL_Month3_19_S408	101	73	46	40	29	19
28LL_Month3_21_S4	101	74	46	41	30	19
28LL_Month3_24_S41	98	73	44	39	29	19
28N_Month1_4_S125	2	2	0	0	0	0

A.D.2.S1. Non-native isolates that did not pass quality control. *Continued*

sample-id	Total length (>= 0 bp)	Total length (>= 1000 bp)	Total length (>= 5000 bp)	Total length (>= 10000 bp)	Total length (>= 25000 bp)	Total length (>= 50000 bp)
23LL_Day3_12_S89	11786997	11779612	8885240	7026978	5492139	5056429
23LL_Day3_15_S18	11968802	11965364	11773229	11206429	9868472	7914269
23LL_Day3_8_S57	11622114	11618282	7976349	5721684	5055204	5055204
23LL_Day7_13_S95	4190950	4164326	3655389	3001004	1163938	59049
23LL_Month1_24_S186	12229807	12224047	12164090	12114822	11958139	11413328
24B_Day7_8_S8	1196587	1168271	518769	184358	0	0
24N_Day7_23_S1316	48468	48468	48468	48468	48468	0
25L_Day7_14_S108	8687	8687	5514	0	0	0
25L_Day7_22_S1088	6521	6521	0	0	0	0
25N_Day3_10_S75	9248196	9245094	5540717	5375014	5267542	5105135
25N_Day3_23_S84	7783416	7783416	5497531	5233154	5068628	5043424
25N_Day3_4_S27	12122783	12115518	12046927	11955838	11786274	11252266
25N_Day3_7_S51	11941923	11918721	11853931	11769529	11604252	11134001
25N_Day3_9_S67	6149619	6148041	6091630	5968504	5619205	5186126

A.D.2.S1. Non-native isolates that did not pass quality control. *Continued*

sample-id	Total length (>= 0 bp)	Total length (>= 1000 bp)	Total length (>= 5000 bp)	Total length (>= 10000 bp)	Total length (>= 25000 bp)	Total length (>= 50000 bp)
25N_Month1_21_S164	11911665	11908989	11559652	10780073	8097256	5700082
25N_Month1_2_S107	12503200	12501566	10916003	8991259	6330147	5436084
25N_Month3_16_S257	2754507	2748024	2721829	2699176	2585248	2316812
26B_Month3_23_S295	336224	320411	64756	0	0	0
26L_Day7_3_S86	2179436	2151394	1425413	719102	25801	0
26N_Day7_14_S1104	2325	2325	0	0	0	0
26N_Day7_19_S1109	4179571	4161868	3742661	3177465	1574247	450795
27L_Day7_10_S1024	4754752	4675912	3932622	3066925	1416921	55846
27N_Day7_10_S99	424837	397715	98703	11881	0	0
27N_Day7_11_S99	4332687	4308205	3865547	3401663	1751986	376643
27N_Month6_4_S36	21789	21308	5497	0	0	0
27R_Month3_14_S328	4448032	4431649	2346196	774969	0	0
28LL_Day7_8_S1054	4352364	4347738	4087793	3743560	2335620	981462
28LL_Month3_14_S4	5819481	5807311	5736323	5687772	5506945	5123786
28LL_Month3_17_S406	5820464	5808516	5739046	5699962	5519101	5135862
28LL_Month3_19_S408	5822311	5810656	5744807	5698677	5532120	5195274
28LL_Month3_21_S4	5822158	5810338	5740864	5701780	5539418	5135909
28LL_Month3_24_S41	5821459	5810420	5739706	5700622	5552566	5196958
28N_Month1_4_S125	4481	4481	0	0	0	0

A.D.2.S1. Non-native isolates that did not pass quality control. *Continued*

sample-id	# contigs	Largest contig	Total length	Reference length	GC (%)	Reference GC (%)
23LL_Day3_12_S89	1583	4628035	11784737	5064668	58.91	50.57
23LL_Day3_15_S18	320	3514664	11967434	5064668	59.53	50.57
23LL_Day3_8_S57	1910	5055204	11622114	5064668	58.81	50.57
23LL_Day7_13_S95	482	59049	4190950	5064668	50.73	50.57
23LL_Month1_24_S186	97	5070988	12227670	5064668	59.38	50.57
24B_Day7_8_S8	353	22648	1195251	5064668	50.87	50.57
24N_Day7_23_S1316	1	48468	48468	5064668	49.84	50.57
25L_Day7_14_S108	2	5514	8687	5064668	46.33	50.57
25L_Day7_22_S1088	3	3173	6521	5064668	48.84	50.57
25N_Day3_10_S75	2180	1517776	9247997	5064668	57.07	50.57
25N_Day3_23_S84	1436	4246038	7783416	5064668	55.16	50.57
25N_Day3_4_S27	118	2434193	12121467	5064668	59.44	50.57
25N_Day3_7_S51	111	5056456	11928468	5064668	59.63	50.57
25N_Day3_9_S67	77	5056480	6149280	5064668	53.01	50.57
25N_Month1_21_S164	464	5087923	11911428	5064668	59.5	50.57
25N_Month1_2_S107	1072	5054894	12503200	5064668	58.88	50.57
25N_Month3_16_S257	47	312900	2752087	5064668	37.49	50.57
26B_Month3_23_S295	155	7883	335813	5064668	51.98	50.57
26L_Day7_3_S86	491	25801	2178995	5064668	52.06	50.57
26N_Day7_14_S1104	2	1311	2325	5064668	48.13	50.57
26N_Day7_19_S1109	403	86488	4179571	5064668	51.13	50.57

A.D.2.S1. Non-native isolates that did not pass quality control. *Continued*

sample-id	# contigs	Largest contig	Total length	Reference length	GC (%)	Reference GC (%)
27L_Day7_10_S1024	667	55846	4753952	5064668	50.64	50.57
27N_Day7_10_S99	192	11881	424837	5064668	53.08	50.57
27N_Day7_11_S99	430	74673	4332687	5064668	50.94	50.57
27N_Month6_4_S36	9	5497	21308	5064668	49.65	50.57
27R_Month3_14_S328	1155	23772	4446526	5064668	50.68	50.57
28LL_Day7_8_S1054	304	104729	4352364	5064668	51.03	50.57
28LL_Month3_14_S4	85	809031	5813639	5064668	59.38	50.57
28LL_Month3_17_S406	84	809031	5815752	5064668	59.38	50.57
28LL_Month3_19_S408	81	809031	5817138	5064668	59.37	50.57
28LL_Month3_21_S4	82	809031	5816685	5064668	59.38	50.57
28LL_Month3_24_S41	81	809031	5816900	5064668	59.38	50.57
28N_Month1_4_S125	2	3173	4481	5064668	48.36	50.57

A.D.2.S1. Non-native isolates that did not pass quality control. *Continued*

sample-id	N50	# misassemblies	# misassembled contigs	Misassembled contigs length	# unaligned contigs	Unaligned length
23LL_Day3_12_S89	17331	9	6	4771952	1530 + 21 part	6204433
23LL_Day3_15_S18	104714	2	1	3514664	313 + 3 part	6786174
23LL_Day3_8_S57	9634	50	49	215672	1711 + 87 part	6071818
23LL_Day7_13_S95	16654	8	7	40806	0 + 1 part	650
23LL_Month1_24_S186	782576	6	4	5348358	89 + 2 part	6867935
24B_Day7_8_S8	4544	2	2	15051	6 + 0 part	10418
24N_Day7_23_S1316	48468	0	0	0	0 + 1 part	35265
25L_Day7_14_S108	5514	0	0	0	0 + 0 part	0
25L_Day7_22_S1088	2136	0	0	0	0 + 0 part	0
25N_Day3_10_S75	485634	7	5	1325992	2139 + 19 part	3811238
25N_Day3_23_S84	4246038	27	26	216423	1335 + 32 part	2242284
25N_Day3_4_S27	332060	3	3	2413849	96 + 1 part	6851818
25N_Day3_7_S51	571739	2	1	5056456	107 + 1 part	6863159
25N_Day3_9_S67	5056480	2	1	5056480	73 + 1 part	1083947
25N_Month1_21_S164	43570	8	1	5087923	458 + 2 part	6736693
25N_Month1_2_S107	25949	10	10	271523	993 + 12 part	6569901
25N_Month3_16_S257	211353	0	0	0	45 + 2 part	2751923
26B_Month3_23_S295	2524	7	5	11118	0 + 0 part	0
26L_Day7_3_S86	6939	10	10	30896	0 + 0 part	0
26N_Day7_14_S1104	1311	0	0	0	0 + 0 part	0

A.D.2.S1. Non-native isolates that did not pass quality control. *Continued*

sample-id	N50	# misassemblies	# misassembled contigs	Misassembled contigs length	# unaligned contigs	Unaligned length
26N_Day7_19_S1109	19776	8	8	51023	0 + 0 part	0
27L_Day7_10_S1024	15661	10	10	49528	0 + 0 part	0
27N_Day7_10_S99	3042	2	2	12781	0 + 0 part	0
27N_Day7_11_S99	21335	9	9	66417	0 + 2 part	1048
27N_Month6_4_S36	3090	0	0	0	0 + 0 part	0
27R_Month3_14_S328	5263	138	120	867262	0 + 2 part	1784
28LL_Day7_8_S1054	27745	12	12	57780	0 + 1 part	518
28LL_Month3_14_S4	327261	0	0	0	84 + 1 part	5813558
28LL_Month3_17_S406	327261	0	0	0	83 + 1 part	5815671
28LL_Month3_19_S408	335517	0	0	0	80 + 1 part	5817057
28LL_Month3_21_S4	327261	0	0	0	81 + 1 part	5816604
28LL_Month3_24_S41	335517	0	0	0	80 + 1 part	5816819
28N_Month1_4_S125	3173	0	0	0	0 + 0 part	0

A.D.2.S1. Non-native isolates that did not pass quality control. *Continued*

sample-id	Genome fraction (%)	Duplication ratio	# N's per 100 kbp	# mismatches per 100 kbp	# indels per 100 kbp	# genomic features
23LL_Day3_12_S89	99.658	1.106	0	20.23	2	4692 + 12 part
23LL_Day3_15_S18	99.866	1.024	0	7.06	0.57	4724 + 6 part
23LL_Day3_8_S57	99.942	1.097	0	29.83	5	4725 + 4 part
23LL_Day7_13_S95	78.826	1.05	0	51.47	2.18	3398 + 541 part
23LL_Month1_24_S186	99.92	1.059	0	6.66	0.67	4721 + 4 part
24B_Day7_8_S8	23.36	1.001	0	93.23	2.7	825 + 510 part
24N_Day7_23_S1316	0.261	1	0	3476.48	151.48	14 + 3 part
25L_Day7_14_S108	0.172	1	0	11.51	0	11 + 1 part
25L_Day7_22_S1088	0.129	1	0	76.68	0	7 + 4 part
25N_Day3_10_S75	99.53	1.079	0	17.87	1.98	4687 + 8 part
25N_Day3_23_S84	99.742	1.097	0	29.89	2.16	4715 + 5 part
25N_Day3_4_S27	99.702	1.044	0	4.2	0.2	4711 + 8 part
25N_Day3_7_S51	99.924	1.001	0	5.75	0.32	4731 + 1 part
25N_Day3_9_S67	99.924	1.001	0	6.36	0.22	4731 + 1 part
25N_Month1_21_S164	99.569	1.026	0	17.31	0.54	4707 + 7 part
25N_Month1_2_S107	99.971	1.172	0	17.95	1.86	4728 + 3 part
25N_Month3_16_S257	0.003	1	0	3658.54	0	0 + 0 part
26B_Month3_23_S295	6.614	1.002	0	156.42	4.48	188 + 236 part

A.D.2.S1. Non-native isolates that did not pass quality control. *Continued*

sample-id	Genome fraction (%)	Duplication ratio	# N's per 100 kbp	# mismatches per 100 kbp	# indels per 100 kbp	# genomic features
26L_Day7_3_S86	42.427	1.014	0	61.29	2.51	1524 + 721 part
26N_Day7_14_S1104	0.045	1.009	0	0	0	3 + 1 part
26N_Day7_19_S1109	79.372	1.04	0	56.44	2.06	3446 + 477 part
27L_Day7_10_S1024	87.345	1.075	0	86.13	5.06	3740 + 603 part
27N_Day7_10_S99	8.386	1	0	101.24	3.06	220 + 292 part
27N_Day7_11_S99	82.357	1.038	0	51.35	2.18	3592 + 473 part
27N_Month6_4_S36	0.417	1.01	0	94.81	4.74	14 + 10 part
27R_Month3_14_S328	87.537	1.003	0	27.59	1.06	3060 + 1289 part
28LL_Day7_8_S1054	83.402	1.03	0	53.76	2.37	3729 + 361 part
28LL_Month3_14_S4	0.002	1	0	0	0	1 + 0 part
28LL_Month3_17_S406	0.002	1	0	0	0	1 + 0 part
28LL_Month3_19_S408	0.002	1	0	0	0	1 + 0 part
28LL_Month3_21_S4	0.002	1	0	0	0	1 + 0 part
28LL_Month3_24_S41	0.002	1	0	0	0	1 + 0 part
28N_Month1_4_S125	0.088	1	0	0	0	6 + 2 part

A.D.2.S1. Non-native isolates that did not pass quality control. *Continued*

sample-id	Largest alignment	Total aligned length	ID	status
23LL_Day3_12_S89	2181346	5572326	AZ20_assemblies/23LL_Day3_12_S89.fasta	disagree
23LL_Day3_15_S18	3339857	5180090	AZ20_assemblies/23LL_Day3_15_S18.fasta	disagree
23LL_Day3_8_S57	5052641	5546188	AZ20_assemblies/23LL_Day3_8_S57.fasta	disagree
23LL_Day7_13_S95	59049	4189820	AZ20_assemblies/23LL_Day7_13_S95.fasta	found
23LL_Month1_24_S186	4615116	5352632	AZ20_assemblies/23LL_Month1_24_S186.fasta	disagree
24B_Day7_8_S8	22411	1183542	AZ20_assemblies/24B_Day7_8_S8.fasta	found
24N_Day7_23_S1316	9961	13203	AZ20_assemblies/24N_Day7_23_S1316.fasta	nomatch
25L_Day7_14_S108	5514	8687	AZ20_assemblies/25L_Day7_14_S108.fasta	nomatch
25L_Day7_22_S1088	3173	6521	AZ20_assemblies/25L_Day7_22_S1088.fasta	nomatch
25N_Day3_10_S75	1511890	5435201	AZ20_assemblies/25N_Day3_10_S75.fasta	disagree
25N_Day3_23_S84	4244238	5534184	AZ20_assemblies/25N_Day3_23_S84.fasta	disagree
25N_Day3_4_S27	2434005	5267721	AZ20_assemblies/25N_Day3_4_S27.fasta	disagree
25N_Day3_7_S51	4694525	5064238	AZ20_assemblies/25N_Day3_7_S51.fasta	disagree
25N_Day3_9_S67	4694548	5064262	AZ20_assemblies/25N_Day3_9_S67.fasta	disagree
25N_Month1_21_S164	1708180	5166252	AZ20_assemblies/25N_Month1_21_S164.fasta	disagree
25N_Month1_2_S107	5053892	5930603	AZ20_assemblies/25N_Month1_2_S107.fasta	disagree
25N_Month3_16_S257	91	164	AZ20_assemblies/25N_Month3_16_S257.fasta	found
26B_Month3_23_S295	7883	334968	AZ20_assemblies/26B_Month3_23_S295.fasta	found
26L_Day7_3_S86	25801	2178387	AZ20_assemblies/26L_Day7_3_S86.fasta	found

A.D.2.S1. Non-native isolates that did not pass quality control. *Continued*

sample-id	Largest alignment	Total aligned length	ID	status
26N_Day7_14_S1104	1290	2304	AZ20_assemblies/26N_Day7_14_S1104.fasta	nomatch

26N_Day7_19_S1109	86488	4178479	AZ20_assemblies/26N_Day7_19_S1109.fasta	found
27L_Day7_10_S1024	55566	4751591	AZ20_assemblies/27L_Day7_10_S1024.fasta	found
27N_Day7_10_S99	11881	424716	AZ20_assemblies/27N_Day7_10_S99.fasta	found
27N_Day7_11_S99	74673	4331023	AZ20_assemblies/27N_Day7_11_S99.fasta	found
27N_Month6_4_S36	5378	21096	AZ20_assemblies/27N_Month6_4_S36.fasta	nomatch
27R_Month3_14_S328	23715	4437738	AZ20_assemblies/27R_Month3_14_S328.fasta	found
28LL_Day7_8_S1054	104729	4350246	AZ20_assemblies/28LL_Day7_8_S1054.fasta	found
28LL_Month3_14_S4	81	81	AZ20_assemblies/28LL_Month3_14_S4.fasta	disagree
28LL_Month3_17_S406	81	81	AZ20_assemblies/28LL_Month3_17_S406.fasta	disagree
28LL_Month3_19_S408	81	81	AZ20_assemblies/28LL_Month3_19_S408.fasta	disagree
28LL_Month3_21_S4	81	81	AZ20_assemblies/28LL_Month3_21_S4.fasta	disagree
28LL_Month3_24_S41	81	81	AZ20_assemblies/28LL_Month3_24_S41.fasta	disagree
28N_Month1_4_S125	3173	4481	AZ20_assemblies/28N_Month1_4_S125.fasta	nomatch

A.D.2.S1. Non-native isolates that did not pass quality control. *Continued*

sample-id	superkingdom	phylum	class
23LL_Day3_12_S89	Bacteria	Proteobacteria	
23LL_Day3_15_S18	Bacteria	Proteobacteria	
23LL_Day3_8_S57	Bacteria	Proteobacteria	
23LL_Day7_13_S95	Bacteria	Proteobacteria	Gammaproteobacteria
23LL_Month1_24_S186	Bacteria	Proteobacteria	
24B_Day7_8_S8	Bacteria	Proteobacteria	Gammaproteobacteria
24N_Day7_23_S1316			
25L_Day7_14_S108			
25L_Day7_22_S1088			
25N_Day3_10_S75	Bacteria	Proteobacteria	
25N_Day3_23_S84	Bacteria	Proteobacteria	
25N_Day3_4_S27	Bacteria	Proteobacteria	
25N_Day3_7_S51	Bacteria	Proteobacteria	
25N_Day3_9_S67	Bacteria	Proteobacteria	
25N_Month1_21_S164	Bacteria	Proteobacteria	
25N_Month1_2_S107	Bacteria	Proteobacteria	
25N_Month3_16_S257	Bacteria	Firmicutes	Bacilli
26B_Month3_23_S295	Bacteria	Proteobacteria	Gammaproteobacteria
26L_Day7_3_S86	Bacteria	Proteobacteria	Gammaproteobacteria
26N_Day7_14_S1104			
26N_Day7_19_S1109	Bacteria	Proteobacteria	Gammaproteobacteria
27L_Day7_10_S1024	Bacteria	Proteobacteria	Gammaproteobacteria
27N_Day7_10_S99	Bacteria	Proteobacteria	Gammaproteobacteria
27N_Day7_11_S99	Bacteria	Proteobacteria	Gammaproteobacteria
27N_Month6_4_S36			
27R_Month3_14_S328	Bacteria	Proteobacteria	Gammaproteobacteria
28LL_Day7_8_S1054	Bacteria	Proteobacteria	Gammaproteobacteria
28LL_Month3_14_S4	Bacteria	Proteobacteria	Alphaproteobacteria
28LL_Month3_17_S406	Bacteria	Proteobacteria	Alphaproteobacteria
28LL_Month3_19_S408	Bacteria	Proteobacteria	Alphaproteobacteria
28LL_Month3_21_S4	Bacteria	Proteobacteria	Alphaproteobacteria
28LL_Month3_24_S41	Bacteria	Proteobacteria	Alphaproteobacteria
28N_Month1_4_S125			

sample-id	Order	family	genus	short-id
23LL_Day3_12_S89				23LL_Day3_12
23LL_Day3_15_S18				23LL_Day3_15
23LL_Day3_8_S57				23LL_Day3_8
23LL_Day7_13_S95	Enterobacterales	Enterobacteriaceae	Escherichia	23LL_Day7_13
23LL_Month1_24_S186				23LL_Month1_24
24B_Day7_8_S8	Enterobacterales	Enterobacteriaceae	Escherichia	24B_Day7_8
24N_Day7_23_S1316				24N_Day7_23
25L_Day7_14_S108				25L_Day7_14
25L_Day7_22_S1088				25L_Day7_22
25N_Day3_10_S75				25N_Day3_10
25N_Day3_23_S84				25N_Day3_23
25N_Day3_4_S27				25N_Day3_4
25N_Day3_7_S51				25N_Day3_7
25N_Day3_9_S67				25N_Day3_9
25N_Month1_21_S164				25N_Month1_21
25N_Month1_2_S107				25N_Month1_2
25N_Month3_16_S257				25N_Month3_16
26B_Month3_23_S295	Enterobacterales	Enterobacteriaceae		26B_Month3_23
26L_Day7_3_S86	Enterobacterales	Enterobacteriaceae	Escherichia	26L_Day7_3
26N_Day7_14_S1104				26N_Day7_14
26N_Day7_19_S1109	Enterobacterales	Enterobacteriaceae	Escherichia	26N_Day7_19
27L_Day7_10_S1024	Enterobacterales	Enterobacteriaceae	Escherichia	27L_Day7_10
27N_Day7_10_S99	Enterobacterales	Enterobacteriaceae		27N_Day7_10
27N_Day7_11_S99	Enterobacterales	Enterobacteriaceae	Escherichia	27N_Day7_11
27N_Month6_4_S36				27N_Month6_4

sample-id	Order	family	genus	short-id
27R_Month3_14_S328	Enterobacterales	Enterobacteriaceae	Escherichia	27R_Month3_14
28LL_Day7_8_S1054	Enterobacterales	Enterobacteriaceae	Escherichia	28LL_Day7_8
28LL_Month3_14_S4	Rhizobiales			28LL_Month3_14
28LL_Month3_17_S406	Rhizobiales			28LL_Month3_17
28LL_Month3_19_S408	Rhizobiales			28LL_Month3_19
28LL_Month3_21_S4	Rhizobiales			28LL_Month3_21
28LL_Month3_24_S41	Rhizobiales			28LL_Month3_24
28N_Month1_4_S125				28N_Month1_4

Table A.D.2.S2. Native isolates that did not pass quality control. Contains checkM, quast, and sourmash results.

sample-id	Marker lineage	# genomes	# markers	# marker sets
41LL_Month3_6_S1469	Bacteria	5449	104	58
41L_Month3_19_S1476	Bacteria	5449	104	58
41L_Month3_20_S1484	Bacteria	5449	104	58
42L_Month3_22_S1598	Bacteria	5449	104	58
42R_Month3_22_S16	Bacteria	5449	104	58
42R_Month3_6_S1567	Bacteria	5449	104	58
43B_Month6_15_S984	Bacteria	5449	104	58
43R_Month6_18_S1006	Bacteria	5449	104	58
44B_Month6_22_S1792	Bacteria	5449	104	58
44N_Month1_11_S189	Bacteria	5449	104	58
44R_Month6_7_S1765	Bacteria	5449	104	58
45L_Day7_11_S1411	Bacteria	5449	104	58
45L_Month3_2_S1627	Bacteria	5449	104	58
45L_Month3_4_S164	Bacteria	5449	104	58
45L_Month3_6_S1659	Bacteria	5449	104	58
45N_Month3_13_S1618	Bacteria	5449	104	58
45N_Month3_14_S1626	Bacteria	5449	104	58
45N_Month3_16_S1642	Bacteria	5449	104	58
45N_Month3_17_S165	Bacteria	5449	104	58
45N_Month3_20_S1674	Bacteria	5449	104	58
45N_Month3_7_S1665	Bacteria	5449	104	58
46L_Day7_10_S361_L002	Bacteria	5449	104	58
46L_Day7_15_S306_L002	Bacteria	5449	104	58
46L_Day7_18_S330_L002	Bacteria	5449	104	58
46L_Day7_4_S313_L002	Bacteria	5449	104	58
46L_Day7_7_S337_L002	Bacteria	5449	104	58
46L_Month6_15_S1076	Bacteria	5449	104	58
46L_Month6_16_S1084	Bacteria	5449	104	58
46L_Month6_21_S1124	Bacteria	5449	104	58
46L_Month6_2_S1067	Bacteria	5449	104	58
46L_Month6_6_S1099	Bacteria	5449	104	58
46N_Month6_14_S1066	Bacteria	5449	104	58
46N_Month6_15_S1074	Bacteria	5449	104	58
46N_Month6_19_S1106	Bacteria	5449	104	58
46R_Day7_10_S363_L002	Bacteria	5449	104	58

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	Marker lineage	# genomes	# markers	# marker sets
46R_Day7_15_S308_L002	Bacteria	5449	104	58
46R_Day7_3_S307_L002	Bacteria	5449	104	58
46R_Day7_5_S323_L002	Bacteria	5449	104	58
46R_Day7_6_S331_L002	Bacteria	5449	104	58
46R_Month6_13_S1062	Bacteria	5449	104	58
46R_Month6_14_S107	Bacteria	5449	104	58
46R_Month6_15_S1078	Bacteria	5449	104	58
46R_Month6_16_S1086	Bacteria	5449	104	58
46R_Month6_17_S1094	Bacteria	5449	104	58
46R_Month6_18_S1102	Bacteria	5449	104	58
46R_Month6_19_S111	Bacteria	5449	104	58
46R_Month6_20_S1118	Bacteria	5449	104	58
46R_Month6_21_S1126	Bacteria	5449	104	58
46R_Month6_22_S1134	Bacteria	5449	104	58
46R_Month6_23_S1142	Bacteria	5449	104	58
46R_Month6_24_S115	Bacteria	5449	104	58

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	Contamination	Strain heterogeneity	# contigs (>= 0 bp)	# contigs (>= 1000 bp)	# contigs (>= 5000 bp)
41LL_Month3_6_S1469	0	0	4	4	1
41L_Month3_19_S1476	0	0	57	46	3
41L_Month3_20_S1484	0.47	66.67	550	496	204
42L_Month3_22_S1598	0	0	581	531	203
42R_Month3_22_S16	2.74	0	1281	1281	258
42R_Month3_6_S1567	0	0	1476	1473	16
43B_Month6_15_S984	82.76	4.35	113	71	58
43R_Month6_18_S1006	72.37	1.92	763	630	373
44B_Month6_22_S1792	0.31	100	1511	1511	27
44N_Month1_11_S189	0	0	1064	1059	232
44R_Month6_7_S1765	1.72	0	1295	1295	232
45L_Day7_11_S1411	0.86	0	856	855	335
45L_Month3_2_S1627	0	0	1459	1459	370
45L_Month3_4_S164	0.34	0	216	215	187
45L_Month3_6_S1659	0	0	678	676	367
45N_Month3_13_S1618	0	0	64	54	49
45N_Month3_14_S1626	0	0	369	369	253
45N_Month3_16_S1642	0	0	154	148	137
45N_Month3_17_S165	0.86	100	405	402	272
45N_Month3_20_S1674	2.74	0	1525	1525	106
45N_Month3_7_S1665	0	0	1400	1400	25
46L_Day7_10_S361_L002	0.16	0	1264	1253	462

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	Contamination	Strain heterogeneity	# contigs (>= 0 bp)	# contigs (>= 1000 bp)	# contigs (>= 5000 bp)
46L_Day7_15_S306_L002	1.88	0	172	169	131
46L_Day7_18_S330_L002	0	0	87	46	40
46L_Day7_4_S313_L002	0	0	104	53	42
46L_Day7_7_S337_L002	0	0	1804	1798	234
46L_Month6_15_S1076	93.3	3.33	865	865	387
46L_Month6_16_S1084	52.1	2.33	2700	2391	547
46L_Month6_21_S1124	5.33	0	446	379	66
46L_Month6_2_S1067	49.29	0	1026	996	394
46L_Month6_6_S1099	0	0	127	125	113
46N_Month6_14_S1066	62.23	7.14	861	848	372
46N_Month6_15_S1074	92.1	9.78	1122	1076	381
46N_Month6_19_S1106	0	0	176	176	73
46R_Day7_10_S363_L002	0	0	108	58	42
46R_Day7_15_S308_L002	0	0	100	51	41
46R_Day7_3_S307_L002	0	0	98	52	43
46R_Day7_5_S323_L002	0	0	168	77	53
46R_Day7_6_S331_L002	0	0	97	47	38
46R_Month6_13_S1062	12.07	0	409	402	194
46R_Month6_14_S107	25.86	0	419	409	276
46R_Month6_15_S1078	1.72	0	339	330	140
46R_Month6_16_S1086	3.45	0	112	111	86
46R_Month6_17_S1094	5.17	66.67	120	117	71
46R_Month6_18_S1102	93.46	2.27	379	366	281

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	Contamination	Strain heterogeneity	# contigs (>= 0 bp)	# contigs (>= 1000 bp)	# contigs (>= 5000 bp)
46R_Month6_19_S111	104.83	6.42	304	297	248
46R_Month6_20_S1118	7.37	12.5	267	263	133
46R_Month6_21_S1126	12.07	0	291	289	229
46R_Month6_22_S1134	38.4	17.86	1642	1630	129
46R_Month6_23_S1142	98.31	5.94	317	315	225
46R_Month6_24_S115	93.68	5.1	426	423	301

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	# contigs (>= 10000 bp)	# contigs (>= 25000 bp)	# contigs (>= 50000 bp)	Total length (>= 0 bp)	Total length (>= 1000 bp)	Total length (>= 5000 bp)
41LL_Month3_6_S1469	0	0	0	13161	13161	7952
41L_Month3_19_S1476	0	0	0	118515	108859	23805
41L_Month3_20_S1484	98	16	0	3331831	3286301	2546318
42L_Month3_22_S1598	87	8	0	3069978	3029056	2195903
42R_Month3_22_S16	44	1	0	4505368	4505368	1995102
42R_Month3_6_S1567	2	0	0	2696877	2696516	109206
43B_Month6_15_S984	53	45	32	8531332	8520589	8481951
43R_Month6_18_S1006	250	86	29	11183780	11130505	10439255
44B_Month6_22_S1792	2	0	0	3074883	3074883	180651
44N_Month1_11_S189	77	15	1	4437760	4436534	2468132
44R_Month6_7_S1765	34	2	0	4274947	4274947	1798841
45L_Day7_11_S1411	115	10	0	4677381	4677258	3306984
45L_Month3_2_S1627	120	12	1	6275118	6275118	3696779
45L_Month3_4_S164	154	83	33	6930821	6930528	6826119
45L_Month3_6_S1659	205	56	16	6867037	6866725	6062257
45N_Month3_13_S1618	45	38	31	7015942	7011829	6998929
45N_Month3_14_S1626	173	77	34	6894356	6894356	6564712
45N_Month3_16_S1642	111	69	36	6956383	6954399	6920121
45N_Month3_17_S165	184	88	24	6879503	6877301	6489892
45N_Month3_20_S1674	6	1	0	3795645	3795645	741933
45N_Month3_7_S1665	2	0	0	2643641	2643641	167309
46L_Day7_10_S361_L002	156	10	0	6479432	6473247	4383411

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	# contigs (>= 10000 bp)	# contigs (>= 25000 bp)	# contigs (>= 50000 bp)	Total length (>= 0 bp)	Total length (>= 1000 bp)	Total length (>= 5000 bp)
46L_Day7_15_S306_L002	102	72	46	6895058	6894230	6769040
46L_Day7_18_S330_L002	37	32	28	7032516	7015466	7001266
46L_Day7_4_S313_L002	40	34	30	7029811	7012772	6987823
46L_Day7_7_S337_L002	19	0	0	5122998	5121288	1633023
46L_Month6_15_S1076	202	48	12	11686056	11686056	10461734
46L_Month6_16_S1084	157	3	0	9286190	9156275	4961746
46L_Month6_21_S1124	46	37	28	5725238	5703080	5052533
46L_Month6_2_S1067	186	55	25	10386487	10370992	8809510
46L_Month6_6_S1099	102	76	42	6937977	6937060	6887621
46N_Month6_14_S1066	148	33	7	10790135	10784094	9520642
46N_Month6_15_S1074	169	37	6	11424652	11410095	9751636
46N_Month6_19_S1106	27	5	2	6138821	6138821	5826061
46R_Day7_10_S363_L002	40	36	30	7037076	7020164	6982785
46R_Day7_15_S308_L002	38	31	27	7031582	7014105	6990451
46R_Day7_3_S307_L002	39	33	30	7030948	7015530	6995706
46R_Day7_5_S323_L002	50	43	35	7021450	6992339	6951680
46R_Day7_6_S331_L002	36	31	29	7032397	7014797	6990615
46R_Month6_13_S1062	86	8	1	7702251	7699956	7112268
46R_Month6_14_S107	160	32	8	9539227	9535374	9100538
46R_Month6_15_S1078	55	6	1	7040277	7036917	6516178
46R_Month6_16_S1086	62	25	6	6852273	6851853	6768018
46R_Month6_17_S1094	32	9	4	6020620	6019779	5883248

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	# contigs (>= 10000 bp)	# contigs (>= 25000 bp)	# contigs (>= 50000 bp)	Total length (>= 0 bp)	Total length (>= 1000 bp)	Total length (>= 5000 bp)
46R_Month6_18_S1102	205	100	49	11756832	11751683	11458232
46R_Month6_19_S111	181	104	44	12353890	12351413	12184393
46R_Month6_20_S1118	78	14	2	7185961	7184848	6812668
46R_Month6_21_S1126	136	35	6	8751842	8750437	8551309
46R_Month6_22_S1134	23	4	1	9103752	9097422	6093887
46R_Month6_23_S1142	166	83	45	11968175	11966797	11708401
46R_Month6_24_S115	209	83	28	11902121	11900618	11517547

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	Total length (>= 10000 bp)	Total length (>= 25000 bp)	Total length (>= 50000 bp)	# contigs	Largest contig	Total length
41LL_Month3_6_S1469	0	0	0	4	7952	13161
41L_Month3_19_S1476	0	0	0	57	8556	118515
41L_Month3_20_S1484	1764873	515423	0	550	49146	3331831
42L_Month3_22_S1598	1354347	214692	0	581	32468	3069978
42R_Month3_22_S16	559076	28883	0	1281	28883	4505368
42R_Month3_6_S1567	26721	0	0	1473	15887	2696516
43B_Month6_15_S984	8446557	8325307	7839021	75	1185431	8523189
43R_Month6_18_S1006	9568889	6976463	4941992	673	495169	11162582
44B_Month6_22_S1792	30905	0	0	1511	17865	3074883
44N_Month1_11_S189	1393903	482927	50164	1059	50164	4436534
44R_Month6_7_S1765	469130	72376	0	1295	40405	4274947
45L_Day7_11_S1411	1749339	332531	0	855	42778	4677258
45L_Month3_2_S1627	1947724	410234	72636	1459	72636	6275118
45L_Month3_4_S164	6583339	5429612	3711588	215	314603	6930528
45L_Month3_6_S1659	4915980	2576700	1295858	676	144749	6866725
45N_Month3_13_S1618	6969163	6857474	6571515	58	685399	7014747
45N_Month3_14_S1626	6005015	4416941	2873345	369	183128	6894356
45N_Month3_16_S1642	6733593	6021059	4732662	150	545347	6955802
45N_Month3_17_S165	5839625	4305948	2090048	405	362936	6879503
45N_Month3_20_S1674	103333	41811	0	1525	41811	3795645
45N_Month3_7_S1665	24140	0	0	1400	12168	2643641
46L_Day7_10_S361_L002	2255846	289666	0	1261	35572	6478769
46L_Day7_15_S306_L002	6555240	6059502	5174689	170	283915	6894820
46L_Day7_18_S330_L002	6977760	6878359	6714630	60	708565	7025310
46L_Day7_4_S313_L002	6977262	6868356	6685131	63	1010288	7019575
46L_Day7_7_S337_L002	220066	0	0	1799	15345	5121874
46L_Month6_15_S1076	9184490	6890953	5637527	865	2274783	11686056
46L_Month6_16_S1084	2226282	87499	0	2495	30468	9232675

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	Total length (>= 10000 bp)	Total length (>= 25000 bp)	Total length (>= 50000 bp)	# contigs	Largest contig	Total length
46L_Month6_21_S1124	4915519	4763271	4418603	390	399099	5710814
46L_Month6_2_S1067	7303443	5340925	4330104	1014	526479	10383339
46L_Month6_6_S1099	6806923	6368429	5105595	126	377614	6937718
46N_Month6_14_S1066	7949115	6306524	5443868	854	5010831	10788440
46N_Month6_15_S1074	8251686	6282240	5259998	1087	2668067	11417390
46N_Month6_19_S1106	5491721	5198996	5092183	176	5011825	6138821
46R_Day7_10_S363_L002	6968960	6900337	6663803	71	686276	7028634
46R_Day7_15_S308_L002	6966945	6838819	6675756	63	686276	7022531
46R_Day7_3_S307_L002	6966931	6852185	6716746	62	590926	7022510
46R_Day7_5_S323_L002	6934103	6816805	6540515	93	684396	7005152
46R_Day7_6_S331_L002	6972426	6873025	6785264	60	590926	7023756
46R_Month6_13_S1062	6344792	5183973	4949035	403	4949035	7700661
46R_Month6_14_S107	8253358	6301292	5452120	412	5011571	9537784
46R_Month6_15_S1078	5907401	5177806	5010518	334	5010518	7039137
46R_Month6_16_S1086	6599130	6017290	5344461	111	5011503	6851853
46R_Month6_17_S1094	5599034	5219694	5060668	117	4149385	6019779
46R_Month6_18_S1102	10933500	9266725	7513145	370	747197	11754984
46R_Month6_19_S111	11693432	10461672	8325029	300	3242715	12353230
46R_Month6_20_S1118	6416407	5476814	5067535	264	5011499	7185446
46R_Month6_21_S1126	7879357	6306767	5304428	290	5011504	8751422
46R_Month6_22_S1134	5396586	5119823	5011404	1637	5011404	9102345
46R_Month6_23_S1142	11288607	9948641	8646384	317	4586989	11968175
46R_Month6_24_S115	10849981	8815570	6868974	425	4410465	11901996

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	Reference length	GC (%)	Reference GC (%)	# misassemblies	# misassembled contigs	Misassembled contigs length
41LL_Month3_6_S1469	5062596	48.7	50.6	0	0	0
41L_Month3_19_S1476	5062596	48.41	50.6	1	1	2080
41L_Month3_20_S1484	5062596	50.69	50.6	6	5	54221
42L_Month3_22_S1598	5062596	50.58	50.6	7	7	38153
42R_Month3_22_S16	5062596	50.5	50.6	67	65	346679
42R_Month3_6_S1567	5062596	50.19	50.6	251	225	494038
43B_Month6_15_S984	5062596	56.76	50.6	0	0	0
43R_Month6_18_S1006	5062596	59.57	50.6	1	1	117826
44B_Month6_22_S1792	5062596	51.25	50.6	151	144	376610
44N_Month1_11_S189	5062596	50.82	50.6	84	80	555537
44R_Month6_7_S1765	5062596	51.1	50.6	32	30	174409
45L_Day7_11_S1411	5062596	50.82	50.6	18	18	195947
45L_Month3_2_S1627	5062596	66.43	50.6	0	0	0
45L_Month3_4_S164	5062596	66.11	50.6	0	0	0
45L_Month3_6_S1659	5062596	66.17	50.6	0	0	0
45N_Month3_13_S1618	5062596	66.09	50.6	0	0	0
45N_Month3_14_S1626	5062596	66.21	50.6	0	0	0
45N_Month3_16_S1642	5062596	66.1	50.6	0	0	0
45N_Month3_17_S165	5062596	66.16	50.6	0	0	0
45N_Month3_20_S1674	5062596	50.46	50.6	176	164	579918
45N_Month3_7_S1665	5062596	50.24	50.6	170	159	337330
46L_Day7_10_S361_L002	5062596	66.38	50.6	0	0	0

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	Reference length	GC (%)	Reference GC (%)	# misassemblies	# misassembled contigs	Misassembled contigs length
46L_Day7_15_S306_L002	5062596	66.22	50.6	0	0	0
46L_Day7_18_S330_L002	5062596	66.09	50.6	0	0	0
46L_Day7_4_S313_L002	5062596	66.1	50.6	0	0	0
46L_Day7_7_S337_L002	5062596	67.13	50.6	0	0	0
46L_Month6_15_S1076	5062596	59.49	50.6	2	2	421547
46L_Month6_16_S1084	5062596	59.52	50.6	107	101	807002
46L_Month6_21_S1124	5062596	53.09	50.6	1	1	103328
46L_Month6_2_S1067	5062596	59.44	50.6	4	4	334595
46L_Month6_6_S1099	5062596	66.25	50.6	0	0	0
46N_Month6_14_S1066	5062596	59.27	50.6	0	0	0
46N_Month6_15_S1074	5062596	59.16	50.6	4	2	2751705
46N_Month6_19_S1106	5062596	53.66	50.6	0	0	0
46R_Day7_10_S363_L002	5062596	66.07	50.6	0	0	0
46R_Day7_15_S308_L002	5062596	66.09	50.6	0	0	0
46R_Day7_3_S307_L002	5062596	66.09	50.6	0	0	0
46R_Day7_5_S323_L002	5062596	66.1	50.6	0	0	0
46R_Day7_6_S331_L002	5062596	66.09	50.6	0	0	0
46R_Month6_13_S1062	5062596	56.78	50.6	1	1	5185
46R_Month6_14_S107	5062596	58.59	50.6	0	0	0
46R_Month6_15_S1078	5062596	55.69	50.6	1	1	5868
46R_Month6_16_S1086	5062596	55.15	50.6	0	0	0
46R_Month6_17_S1094	5062596	53.16	50.6	1	1	196533

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	Reference length	GC (%)	Reference GC (%)	# misassemblies	# misassembled contigs	Misassembled contigs length
46R_Month6_18_S1102	5062596	59.55	50.6	2	2	422527
46R_Month6_19_S111	5062596	59.27	50.6	2	2	38288
46R_Month6_20_S1118	5062596	55.74	50.6	0	0	0
46R_Month6_21_S1126	5062596	57.91	50.6	0	0	0
46R_Month6_22_S1134	5062596	57.67	50.6	21	16	5206987
46R_Month6_23_S1142	5062596	59.53	50.6	2	2	4601040
46R_Month6_24_S115	5062596	59.57	50.6	1	1	377603

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	# local misassemblies	# scaffold gap ext. mis.	# scaffold gap loc. mis.	# unaligned mis. contigs	# unaligned contigs
41LL_Month3_6_S1469	0	0	0	0	0 + 0 part
41L_Month3_19_S1476	0	0	0	0	0 + 0 part
41L_Month3_20_S1484	4	0	0	0	0 + 0 part
42L_Month3_22_S1598	4	0	0	0	0 + 0 part
42R_Month3_22_S16	5	0	0	0	0 + 0 part
42R_Month3_6_S1567	3	0	0	0	0 + 0 part
43B_Month6_15_S984	0	0	0	0	33 + 2 part
43R_Month6_18_S1006	2	0	0	0	609 + 1 part
44B_Month6_22_S1792	7	0	0	0	0 + 1 part
44N_Month1_11_S189	13	0	0	0	0 + 0 part
44R_Month6_7_S1765	9	0	0	0	0 + 0 part
45L_Day7_11_S1411	5	0	0	0	0 + 0 part
45L_Month3_2_S1627	0	0	0	1	1457 + 2 part
45L_Month3_4_S164	0	0	0	2	212 + 3 part
45L_Month3_6_S1659	0	0	0	2	673 + 3 part
45N_Month3_13_S1618	0	0	0	2	55 + 3 part
45N_Month3_14_S1626	0	0	0	2	366 + 3 part
45N_Month3_16_S1642	0	0	0	2	147 + 3 part
45N_Month3_17_S165	0	0	0	1	403 + 2 part
45N_Month3_20_S1674	6	0	0	0	0 + 0 part
45N_Month3_7_S1665	5	0	0	0	0 + 0 part
46L_Day7_10_S361_L002	0	0	0	1	1259 + 2 part

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	# local misassemblies	# scaffold gap ext. mis.	# scaffold gap loc. mis.	# unaligned mis. contigs	# unaligned contigs
46L_Day7_15_S306_L002	0	0	0	2	167 + 3 part
46L_Day7_18_S330_L002	0	0	0	2	57 + 3 part
46L_Day7_4_S313_L002	0	0	0	2	61 + 2 part
46L_Day7_7_S337_L002	0	0	0	1	1798 + 1 part
46L_Month6_15_S1076	8	0	0	1	855 + 2 part
46L_Month6_16_S1084	5	0	0	0	1566 + 2 part
46L_Month6_21_S1124	5	0	0	0	313 + 0 part
46L_Month6_2_S1067	17	0	0	0	926 + 1 part
46L_Month6_6_S1099	0	0	0	2	123 + 3 part
46N_Month6_14_S1066	0	0	0	0	845 + 1 part
46N_Month6_15_S1074	0	0	0	0	1065 + 1 part
46N_Month6_19_S1106	1	0	0	0	170 + 1 part
46R_Day7_10_S363_L002	0	0	0	1	65 + 3 part
46R_Day7_15_S308_L002	0	0	0	2	60 + 3 part
46R_Day7_3_S307_L002	0	0	0	2	59 + 3 part
46R_Day7_5_S323_L002	0	0	0	2	91 + 2 part
46R_Day7_6_S331_L002	0	0	0	2	56 + 3 part
46R_Month6_13_S1062	1	0	0	0	397 + 1 part
46R_Month6_14_S107	0	0	0	1	406 + 3 part
46R_Month6_15_S1078	1	0	0	0	328 + 0 part
46R_Month6_16_S1086	0	0	0	0	107 + 0 part
46R_Month6_17_S1094	1	0	0	0	100 + 0 part
46R_Month6_18_S1102	1	0	0	1	334 + 2 part

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	# local misassemblies	# scaffold gap ext. mis.	# scaffold gap loc. mis.	# unaligned mis. contigs	# unaligned contigs
46R_Month6_19_S111	7	0	0	1	269 + 3 part
46R_Month6_20_S1118	0	0	0	0	255 + 2 part
46R_Month6_21_S1126	0	0	0	0	284 + 2 part
46R_Month6_22_S1134	19	0	0	0	1581 + 0 part
46R_Month6_23_S1142	6	0	0	1	302 + 3 part
46R_Month6_24_S115	6	0	0	1	415 + 2 part

A.D.2.S2. Native isolates that did not pass quality control. Continued

sample-id	Unaligned length	Genome fraction (%)	Duplication ratio	# N's per 100 kbp	# mismatches per 100 kbp	# indels per 100 kbp
41LL_Month3_6_S1469	0	0.26	1.001	0	91.24	15.21
41L_Month3_19_S1476	0	2.341	1	0	77.62	5.06
41L_Month3_20_S1484	0	64.154	1.026	0	34.67	2.31
42L_Month3_22_S1598	0	59.252	1.023	0	29.87	3.07
42R_Month3_22_S16	0	88.843	1.002	0	9.14	0.4
42R_Month3_6_S1567	0	53.118	1.003	0	18.22	0.59
43B_Month6_15_S984	3510958	99.005	1	0	0.28	0
43R_Month6_18_S1006	6164750	98.71	1	0	1.22	0.06
44B_Month6_22_S1792	681	60.648	1.001	0	30.45	0.75
44N_Month1_11_S189	0	87.541	1.001	0	17.1	0.47
44R_Month6_7_S1765	0	84.357	1.001	0	18.64	0.68
45L_Day7_11_S1411	0	92.34	1.001	0	10.29	0.62
45L_Month3_2_S1627	6274662	0.009	1	0	2850.88	0
45L_Month3_4_S164	6929913	0.011	1.143	0	2788.1	0
45L_Month3_6_S1659	6866110	0.012	1	0	2439.02	0
45N_Month3_13_S1618	7014132	0.011	1.143	0	2788.1	0
45N_Month3_14_S1626	6893741	0.011	1.143	0	2788.1	0
45N_Month3_16_S1642	6955187	0.011	1.143	0	2788.1	0
45N_Month3_17_S165	6879047	0.009	1	0	2850.88	0
45N_Month3_20_S1674	0	74.771	1.003	0	13.95	0.32
45N_Month3_7_S1665	0	52.098	1.002	0	17.63	0.45
46L_Day7_10_S361_L002	6478313	0.009	1	0	2850.88	0
46L_Day7_15_S306_L002	6894205	0.012	1	0	2439.02	0
46L_Day7_18_S330_L002	7024695	0.012	1	0	2439.02	0
46L_Day7_4_S313_L002	7019037	0.011	1	0	2788.1	0
46L_Day7_7_S337_L002	5121495	0.007	1	0	3430.08	0

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	Unaligned length	Genome fraction (%)	Duplication ratio	# N's per 100 kbp	# mismatches per 100 kbp	# indels per 100 kbp
46L_Month6_15_S1076	6609747	99.86	1.004	0	3.13	0.32
46L_Month6_16_S1084	4721388	89.06	1.001	0	8.98	0.33
46L_Month6_21_S1124	723794	98.495	1	0	1.64	0.08
46L_Month6_2_S1067	5396079	98.507	1	0	6.18	0.2
46L_Month6_6_S1099	6937103	0.012	1	0	2439.02	0
46N_Month6_14_S1066	5603305	99.987	1.024	0	3.65	0.45
46N_Month6_15_S1074	6169865	99.879	1.038	0	5.85	0.95
46N_Month6_19_S1106	992086	100	1.017	0	5.53	0.45
46R_Day7_10_S363_L002	7018553	0.18	1.109	0	142.98	0
46R_Day7_15_S308_L002	7021916	0.011	1.143	0	2788.1	0
46R_Day7_3_S307_L002	7021895	0.011	1.143	0	2788.1	0
46R_Day7_5_S323_L002	7004614	0.011	1	0	2788.1	0
46R_Day7_6_S331_L002	7022722	0.012	1.687	0	2120.72	0
46R_Month6_13_S1062	2649112	99.768	1	0	4.79	0.61
46R_Month6_14_S107	4474525	100	1	0	4.48	0.4
46R_Month6_15_S1078	1968508	99.98	1.002	0	7.01	0.63
46R_Month6_16_S1086	1790098	99.983	1	0	3.06	0.16
46R_Month6_17_S1094	822230	99.813	1.029	0	2.28	0.91
46R_Month6_18_S1102	6740192	99.054	1	0	0.86	0
46R_Month6_19_S111	6910184	99.847	1.077	0	3.58	0.97
46R_Month6_20_S1118	2113883	100	1.002	0	6	0.41
46R_Month6_21_S1126	3684857	99.814	1.003	0	3.88	0.26
46R_Month6_22_S1134	3689509	99.818	1.071	0	10.55	3.34
46R_Month6_23_S1142	6780950	99.863	1.026	0	2.75	0.59

A.D.2.S2. Native isolates that did not pass quality control. Continued

sample-id	Unaligned length	Genome fraction (%)	Duplication ratio	# N's per 100 kbp	# mismatches per 100 kbp	# indels per 100 kbp
46R_Month6_24_S115	6834300	99.796	1.003	0	1.52	0.04

A.D.2.S2. Native isolates that did not pass quality control. Continued

sample-id	# genomic features	Largest alignment	Total aligned length	NA50	NGA50	NA75	NGA75
41LL_Month3_6_S1469	0 + 0 part	7952	13161	7952	-	2885	-
41L_Month3_19_S1476	0 + 0 part	8556	118515	2369	-	1469	-
41L_Month3_20_S1484	0 + 0 part	49146	3329966	10735	5024	5134	-
42L_Month3_22_S1598	0 + 0 part	32468	3067739	8901	3507	4603	-
42R_Month3_22_S16	0 + 0 part	28883	4500299	4300	3826	2599	1986
42R_Month3_6_S1567	0 + 0 part	15887	2690283	1715	1004	1260	-
43B_Month6_15_S984	0 + 0 part	1185431	5012231	198107	747198	-	358910
43R_Month6_18_S1006	0 + 0 part	495169	4997832	-	215324	-	124669
44B_Month6_22_S1792	0 + 0 part	17865	3071073	2085	1284	1428	-
44N_Month1_11_S189	0 + 0 part	47492	4432259	5405	4527	2846	2012
44R_Month6_7_S1765	0 + 0 part	40199	4271268	3961	3399	2357	1663
45L_Day7_11_S1411	0 + 0 part	42778	4675144	7699	7103	4283	3486
45L_Month3_2_S1627	0 + 0 part	314	456	-	-	-	-
45L_Month3_4_S164	0 + 0 part	314	615	-	-	-	-
45L_Month3_6_S1659	0 + 0 part	314	615	-	-	-	-
45N_Month3_13_S1618	0 + 0 part	314	615	-	-	-	-
45N_Month3_14_S1626	0 + 0 part	314	615	-	-	-	-
45N_Month3_16_S1642	0 + 0 part	314	615	-	-	-	-
45N_Month3_17_S165	0 + 0 part	314	456	-	-	-	-
45N_Month3_20_S1674	0 + 0 part	41811	3787648	2661	1982	1686	-
45N_Month3_7_S1665	0 + 0 part	12168	2638743	1852	993	1327	-
46L_Day7_10_S361_L002	0 + 0 part	314	456	-	-	-	-

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	# genomic features	Largest alignment	Total aligned length	NA50	NGA50	NA75	NGA75
46L_Day7_15_S306_L002	0 + 0 part	314	615	-	-	-	-
46L_Day7_18_S330_L002	0 + 0 part	314	615	-	-	-	-
46L_Day7_4_S313_L002	0 + 0 part	314	538	-	-	-	-
46L_Day7_7_S337_L002	0 + 0 part	314	379	-	-	-	-
46L_Month6_15_S1076	0 + 0 part	2274307	5074853	-	2126354	-	2126354
46L_Month6_16_S1084	0 + 0 part	30468	4509441	-	5986	-	2771
46L_Month6_21_S1124	0 + 0 part	399099	4986841	139153	185443	77050	91937
46L_Month6_2_S1067	0 + 0 part	525747	4986060	-	168267	-	96677
46L_Month6_6_S1099	0 + 0 part	314	615	-	-	-	-
46N_Month6_14_S1066	0 + 0 part	5010831	5185091	-	501083 1	-	501083 1
46N_Month6_15_S1074	0 + 0 part	2285980	5245388	-	138665 2	-	127650 4
46N_Month6_19_S1106	0 + 0 part	5011550	5146460	501155 0	501155 0	501155 0	501155 0
46R_Day7_10_S363_L002	0 + 0 part	8556	9169	-	-	-	-
46R_Day7_15_S308_L002	0 + 0 part	314	615	-	-	-	-
46R_Day7_3_S307_L002	0 + 0 part	314	615	-	-	-	-
46R_Day7_5_S323_L002	0 + 0 part	314	538	-	-	-	-
46R_Day7_6_S331_L002	0 + 0 part	169	613	-	-	-	-
46R_Month6_13_S1062	0 + 0 part	4949035	5051549	494903 5	494903 5	-	494903 5
46R_Month6_14_S107	0 + 0 part	5011571	5063259	501157 1	501157 1	-	501157 1
46R_Month6_15_S1078	0 + 0 part	5010486	5070596	501048 6	501048 6	-	501048 6
46R_Month6_16_S1086	0 + 0 part	5011503	5061755	501150 3	501150 3	-	501150 3
46R_Month6_17_S1094	0 + 0 part	4149070	5196987	414907 0	414907 0	662018	414907 0

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	# genomic features	Largest alignment	Total aligned length	NA50	NGA50	NA75	NGA75
46R_Month6_18_S1102	0 + 0 part	747197	5014792	-	386544	-	347638
46R_Month6_19_S111	0 + 0 part	3241595	5440868	-	3241595	-	975634
46R_Month6_20_S1118	0 + 0 part	5011499	5071549	5011499	5011499	-	5011499
46R_Month6_21_S1126	0 + 0 part	5011504	5066512	5011504	5011504	-	5011504
46R_Month6_22_S1134	0 + 0 part	4565292	5410343	4565292	4565292	-	4565292
46R_Month6_23_S1142	0 + 0 part	4540774	5186124	-	4540774	-	4540774
46R_Month6_24_S115	0 + 0 part	4409841	5067019	-	4409841	-	4409841

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	LA50	LGA50	LA75	LGA75	ID	status
41LL_Month3_6_S1469	1	-	2	-	scratch2/AZ51_assemblies/41LL_Month3_6_S1469.fasta	nomatch
41L_Month3_19_S1476	15	-	31	-	scratch2/AZ51_assemblies/41L_Month3_19_S1476.fasta	nomatch
41L_Month3_20_S1484	89	203	197	-	scratch2/AZ51_assemblies/41L_Month3_20_S1484.fasta	found
42L_Month3_22_S1598	107	285	228	-	scratch2/AZ51_assemblies/42L_Month3_22_S1598.fasta	found
42R_Month3_22_S16	327	396	661	845	scratch2/AZ51_assemblies/42R_Month3_22_S16.fasta	found
42R_Month3_6_S1567	534	1451	996	-	scratch2/AZ51_assemblies/42R_Month3_6_S1567.fasta	found
43B_Month6_15_S984	8	3	-	6	scratch2/AZ51_assemblies/43B_Month6_15_S984.fasta	disagree
43R_Month6_18_S1006	-	8	-	15	scratch2/AZ51_assemblies/43R_Month6_18_S1006.fasta	disagree
44B_Month6_22_S1792	502	1115	948	-	scratch2/AZ51_assemblies/44B_Month6_22_S1792.fasta	found
44N_Month1_11_S189	195	259	483	681	scratch2/AZ51_assemblies/44N_Month1_11_S189.fasta	found
44R_Month6_7_S1765	316	423	663	960	scratch2/AZ51_assemblies/44R_Month6_7_S1765.fasta	found
45L_Day7_11_S1411	185	212	388	462	scratch2/AZ51_assemblies/45L_Day7_11_S1411.fasta	found
45L_Month3_2_S1627	-	-	-	-	scratch2/AZ51_assemblies/45L_Month3_2_S1627.fasta	found
45L_Month3_4_S164	-	-	-	-	scratch2/AZ51_assemblies/45L_Month3_4_S164.fasta	found
45L_Month3_6_S1659	-	-	-	-	scratch2/AZ51_assemblies/45L_Month3_6_S1659.fasta	found
45N_Month3_13_S1618	-	-	-	-	scratch2/AZ51_assemblies/45N_Month3_13_S1618.fasta	found
45N_Month3_14_S1626	-	-	-	-	scratch2/AZ51_assemblies/45N_Month3_14_S1626.fasta	found

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	LA50	LGA50	LA75	LGA75	ID	status
45N_Month3_16_S1642	-	-	-	-	scratch2/AZ51_assemblies/45N_Month3_16_S1642.fasta	found
45N_Month3_17_S165	-	-	-	-	scratch2/AZ51_assemblies/45N_Month3_17_S165.fasta	found
45N_Month3_20_S1674	452	728	901	-	scratch2/AZ51_assemblies/45N_Month3_20_S1674.fasta	found
45N_Month3_7_S1665	467	1367	889	-	scratch2/AZ51_assemblies/45N_Month3_7_S1665.fasta	found
46L_Day7_10_S361_L002	-	-	-	-	scratch2/AZ51_assemblies/46L_Day7_10_S361_L002.fasta	found
46L_Day7_15_S306_L002	-	-	-	-	scratch2/AZ51_assemblies/46L_Day7_15_S306_L002.fasta	found
46L_Day7_18_S330_L002	-	-	-	-	scratch2/AZ51_assemblies/46L_Day7_18_S330_L002.fasta	found
46L_Day7_4_S313_L002	-	-	-	-	scratch2/AZ51_assemblies/46L_Day7_4_S313_L002.fasta	found
46L_Day7_7_S337_L002	-	-	-	-	scratch2/AZ51_assemblies/46L_Day7_7_S337_L002.fasta	found
46L_Month6_15_S1076	-	2	-	2	scratch2/AZ51_assemblies/46L_Month6_15_S1076.fasta	disagree
46L_Month6_16_S1084	-	252	-	560	scratch2/AZ51_assemblies/46L_Month6_16_S1084.fasta	disagree
46L_Month6_21_S1124	12	10	26	21	scratch2/AZ51_assemblies/46L_Month6_21_S1124.fasta	disagree
46L_Month6_2_S1067	-	9	-	19	scratch2/AZ51_assemblies/46L_Month6_2_S1067.fasta	disagree
46L_Month6_6_S1099	-	-	-	-	scratch2/AZ51_assemblies/46L_Month6_6_S1099.fasta	found
46N_Month6_14_S1066	-	1	-	1	scratch2/AZ51_assemblies/46N_Month6_14_S1066.fasta	disagree
46N_Month6_15_S1074	-	2	-	3	scratch2/AZ51_assemblies/46N_Month6_15_S1074.fasta	disagree

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	LA50	LGA50	LA75	LGA75	ID	status
46N_Month6_19_S1106	1	1	1	1	scratch2/AZ51_assemblies/46N_Month6_19_S1106.fasta	disagree
46R_Day7_10_S363_L002	-	-	-	-	scratch2/AZ51_assemblies/46R_Day7_10_S363_L002.fasta	found
46R_Day7_15_S308_L002	-	-	-	-	scratch2/AZ51_assemblies/46R_Day7_15_S308_L002.fasta	found
46R_Day7_3_S307_L002	-	-	-	-	scratch2/AZ51_assemblies/46R_Day7_3_S307_L002.fasta	found
46R_Day7_5_S323_L002	-	-	-	-	scratch2/AZ51_assemblies/46R_Day7_5_S323_L002.fasta	found
46R_Day7_6_S331_L002	-	-	-	-	scratch2/AZ51_assemblies/46R_Day7_6_S331_L002.fasta	found
46R_Month6_13_S1062	1	1	-	1	scratch2/AZ51_assemblies/46R_Month6_13_S1062.fasta	disagree
46R_Month6_14_S107	1	1	-	1	scratch2/AZ51_assemblies/46R_Month6_14_S107.fasta	disagree
46R_Month6_15_S1078	1	1	-	1	scratch2/AZ51_assemblies/46R_Month6_15_S1078.fasta	disagree
46R_Month6_16_S1086	1	1	-	1	scratch2/AZ51_assemblies/46R_Month6_16_S1086.fasta	disagree
46R_Month6_17_S1094	1	1	2	1	scratch2/AZ51_assemblies/46R_Month6_17_S1094.fasta	disagree
46R_Month6_18_S1102	-	5	-	8	scratch2/AZ51_assemblies/46R_Month6_18_S1102.fasta	disagree
46R_Month6_19_S111	-	1	-	2	scratch2/AZ51_assemblies/46R_Month6_19_S111.fasta	disagree
46R_Month6_20_S1118	1	1	-	1	scratch2/AZ51_assemblies/46R_Month6_20_S1118.fasta	disagree
46R_Month6_21_S1126	1	1	-	1	scratch2/AZ51_assemblies/46R_Month6_21_S1126.fasta	disagree

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	LA50	LGA50	LA75	LGA75	ID	status
46R_Month6_22_S1134	1	1	-	1	scratch2/AZ51_assemblies/46R_Month6_22_S1134.fasta	disagree
46R_Month6_23_S1142	-	1	-	1	scratch2/AZ51_assemblies/46R_Month6_23_S1142.fasta	disagree
46R_Month6_24_S115	-	1	-	1	scratch2/AZ51_assemblies/46R_Month6_24_S115.fasta	disagree

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	superkingdom	phylum	class	order
41LL_Month3_6_S1469				
41L_Month3_19_S1476				
41L_Month3_20_S1484	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales
42L_Month3_22_S1598	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales
42R_Month3_22_S16	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales
42R_Month3_6_S1567	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales
43B_Month6_15_S984	Bacteria	Proteobacteria	Gammaproteobacteria	
43R_Month6_18_S1006	Bacteria	Proteobacteria	Gammaproteobacteria	
44B_Month6_22_S1792	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales
44N_Month1_11_S189	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales
44R_Month6_7_S1765	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales
45L_Day7_11_S1411	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales
45L_Month3_2_S1627	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
45L_Month3_4_S164	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
45L_Month3_6_S1659	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
45N_Month3_13_S1618	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
45N_Month3_14_S1626	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
45N_Month3_16_S1642	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
45N_Month3_17_S165	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
45N_Month3_20_S1674	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales
45N_Month3_7_S1665	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales
46L_Day7_10_S361_L002	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
46L_Day7_15_S306_L002	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	superkingdom	phylum	class	order
46L_Day7_18_S330_L002	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
46L_Day7_4_S313_L002	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
46L_Day7_7_S337_L002	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
46L_Month6_15_S1076	Bacteria	Proteobacteria	Gammaproteobacteria	
46L_Month6_16_S1084	Bacteria	Proteobacteria	Gammaproteobacteria	
46L_Month6_21_S1124	Bacteria	Proteobacteria	Gammaproteobacteria	
46L_Month6_2_S1067	Bacteria	Proteobacteria	Gammaproteobacteria	
46L_Month6_6_S1099	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
46N_Month6_14_S1066	Bacteria	Proteobacteria	Gammaproteobacteria	
46N_Month6_15_S1074	Bacteria	Proteobacteria	Gammaproteobacteria	
46N_Month6_19_S1106	Bacteria	Proteobacteria	Gammaproteobacteria	
46R_Day7_10_S363_L002	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
46R_Day7_15_S308_L002	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
46R_Day7_3_S307_L002	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
46R_Day7_5_S323_L002	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
46R_Day7_6_S331_L002	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
46R_Month6_13_S1062	Bacteria	Proteobacteria	Gammaproteobacteria	
46R_Month6_14_S107	Bacteria	Proteobacteria	Gammaproteobacteria	
46R_Month6_15_S1078	Bacteria	Proteobacteria	Gammaproteobacteria	
46R_Month6_16_S1086	Bacteria	Proteobacteria	Gammaproteobacteria	
46R_Month6_17_S1094	Bacteria	Proteobacteria	Gammaproteobacteria	
46R_Month6_18_S1102	Bacteria	Proteobacteria	Gammaproteobacteria	
46R_Month6_19_S111	Bacteria	Proteobacteria	Gammaproteobacteria	

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	superkingdom	phylum	class	order
46R_Month6_20_S1118	Bacteria	Proteobacteria	Gammaproteobacteria	
46R_Month6_21_S1126	Bacteria	Proteobacteria	Gammaproteobacteria	
46R_Month6_22_S1134	Bacteria	Proteobacteria	Gammaproteobacteria	
46R_Month6_23_S1142	Bacteria	Proteobacteria	Gammaproteobacteria	
46R_Month6_24_S115	Bacteria	Proteobacteria	Gammaproteobacteria	

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	family	genus	short-id	long-sample-id
41LL_Month3_6_S1469			41LL_Month3_6	41LL_Month3_6_S1469
41L_Month3_19_S1476			41L_Month3_19	41L_Month3_19_S1476
41L_Month3_20_S1484	Enterobacteriaceae	Escherichia	41L_Month3_20	41L_Month3_20_S1484
42L_Month3_22_S1598	Enterobacteriaceae	Escherichia	42L_Month3_22	42L_Month3_22_S1598
42R_Month3_22_S16	Enterobacteriaceae	Escherichia	42R_Month3_22	42R_Month3_22_S16
42R_Month3_6_S1567	Enterobacteriaceae	Escherichia	42R_Month3_6	42R_Month3_6_S1567
43B_Month6_15_S984			43B_Month6_15	43B_Month6_15_S984
43R_Month6_18_S1006			43R_Month6_18	43R_Month6_18_S1006
44B_Month6_22_S1792	Enterobacteriaceae	Escherichia	44B_Month6_22	44B_Month6_22_S1792
44N_Month1_11_S189	Enterobacteriaceae	Escherichia	44N_Month1_11	44N_Month1_11_S189
44R_Month6_7_S1765	Enterobacteriaceae	Escherichia	44R_Month6_7	44R_Month6_7_S1765
45L_Day7_11_S1411	Enterobacteriaceae	Escherichia	45L_Day7_11	45L_Day7_11_S1411
45L_Month3_2_S1627	Pseudomonadaceae	Pseudomonas	45L_Month3_2	45L_Month3_2_S1627
45L_Month3_4_S164	Pseudomonadaceae	Pseudomonas	45L_Month3_4	45L_Month3_4_S164
45L_Month3_6_S1659	Pseudomonadaceae	Pseudomonas	45L_Month3_6	45L_Month3_6_S1659
45N_Month3_13_S1618	Pseudomonadaceae	Pseudomonas	45N_Month3_13	45N_Month3_13_S1618
45N_Month3_14_S1626	Pseudomonadaceae	Pseudomonas	45N_Month3_14	45N_Month3_14_S1626
45N_Month3_16_S1642	Pseudomonadaceae	Pseudomonas	45N_Month3_16	45N_Month3_16_S1642
45N_Month3_17_S165	Pseudomonadaceae	Pseudomonas	45N_Month3_17	45N_Month3_17_S165
45N_Month3_20_S1674	Enterobacteriaceae	Escherichia	45N_Month3_20	45N_Month3_20_S1674
45N_Month3_7_S1665	Enterobacteriaceae	Escherichia	45N_Month3_7	45N_Month3_7_S1665
46L_Day7_10_S361_L002	Pseudomonadaceae	Pseudomonas	46L_Day7_10	46L_Day7_10_S361
46L_Day7_15_S306_L002	Pseudomonadaceae	Pseudomonas	46L_Day7_15	46L_Day7_15_S306

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	family	genus	short-id	long-sample-id
46L_Day7_18_S330_L002	Pseudomonadaceae	Pseudomonas	46L_Day7_18	46L_Day7_18_S330
46L_Day7_4_S313_L002	Pseudomonadaceae	Pseudomonas	46L_Day7_4	46L_Day7_4_S313
46L_Day7_7_S337_L002	Pseudomonadaceae	Pseudomonas	46L_Day7_7	46L_Day7_7_S337
46L_Month6_15_S1076			46L_Month6_15	46L_Month6_15_S1076
46L_Month6_16_S1084			46L_Month6_16	46L_Month6_16_S1084
46L_Month6_21_S1124			46L_Month6_21	46L_Month6_21_S1124
46L_Month6_2_S1067			46L_Month6_2	46L_Month6_2_S1067
46L_Month6_6_S1099	Pseudomonadaceae	Pseudomonas	46L_Month6_6	46L_Month6_6_S1099
46N_Month6_14_S1066			46N_Month6_14	46N_Month6_14_S1066
46N_Month6_15_S1074			46N_Month6_15	46N_Month6_15_S1074
46N_Month6_19_S1106			46N_Month6_19	46N_Month6_19_S1106
46R_Day7_10_S363_L002	Pseudomonadaceae	Pseudomonas	46R_Day7_10	46R_Day7_10_S363
46R_Day7_15_S308_L002	Pseudomonadaceae	Pseudomonas	46R_Day7_15	46R_Day7_15_S308
46R_Day7_3_S307_L002	Pseudomonadaceae	Pseudomonas	46R_Day7_3	46R_Day7_3_S307
46R_Day7_5_S323_L002	Pseudomonadaceae	Pseudomonas	46R_Day7_5	46R_Day7_5_S323
46R_Day7_6_S331_L002	Pseudomonadaceae	Pseudomonas	46R_Day7_6	46R_Day7_6_S331
46R_Month6_13_S1062			46R_Month6_13	46R_Month6_13_S1062
46R_Month6_14_S107			46R_Month6_14	46R_Month6_14_S107
46R_Month6_15_S1078			46R_Month6_15	46R_Month6_15_S1078
46R_Month6_16_S1086			46R_Month6_16	46R_Month6_16_S1086
46R_Month6_17_S1094			46R_Month6_17	46R_Month6_17_S1094
46R_Month6_18_S1102			46R_Month6_18	46R_Month6_18_S1102
46R_Month6_19_S111			46R_Month6_19	46R_Month6_19_S111

A.D.2.S2. Native isolates that did not pass quality control. *Continued*

sample-id	family	genus	short-id	long-sample-id
46R_Month6_20_S1118			46R_Month6_20	46R_Month6_20_S1118
46R_Month6_21_S1126			46R_Month6_21	46R_Month6_21_S1126
46R_Month6_22_S1134			46R_Month6_22	46R_Month6_22_S1134
46R_Month6_23_S1142			46R_Month6_23	46R_Month6_23_S1142
46R_Month6_24_S115			46R_Month6_24	46R_Month6_24_S115