1-1-2016

The Impact of Tiny Organisms: Microbial Communities and Disease States

Christel Sjöland Chehoud
University of Pennsylvania, cchehoud@gmail.com

The Impact of Tiny Organisms: Microbial Communities and Disease States

Abstract
In the last decade, primarily through the use of sequencing, much has been learned about the trillions of microorganisms that reside in human hosts. These microorganisms play a wide range of roles including helping our immune systems develop, digesting our food, and protecting us from the invasion of pathogenic organisms. My thesis focuses on the characterization of fungal, viral, and bacterial communities in humans, investigating the use of defined microbial communities to cure diseases in animal models, and examining the effects of human microbiome modifications through fecal microbiota transfers. In the first part of this thesis, I use deep sequencing of ribosomal RNA gene tags to characterize the composition of the bacterial, fungal, and archaeal microbiota in pediatric patients with Inflammatory Bowel Disease and healthy controls. Archaeal reads were rare in the pediatric samples, whereas an abundant amount of fungal reads was recovered. Pediatric IBD was found to be associated with reduced diversity in both fungal and bacterial gut microbiota, and specific Candida taxa were increased in abundance in the IBD samples. I, then, describe my use of a variety of experimental and computational methods to study the viral communities of immune-compromised lung transplant recipients. Anelloviruses, circular, single-stranded DNA viruses, were found in all lung samples but were 56 times more abundant in samples from lung transplant recipients as compared to healthy controls or HIV+ subjects. In the third part of this thesis, I describe the use of defined microbial communities in mice, and its ability to reduce the production of ammonia long term and mitigate hepatic encephalopathy. This was shown to be true in both mice on a normal protein diet or a low protein diet. Last, I investigate the transfer of viral communities between humans through FMT and characterize features associated with efficient transmission. A case series where feces from a single donor were used to treat three children with ulcerative colitis was used for the analysis. Ultimately this work showed that multiple viral lineages do transfer between human individuals through fecal microbiota transplants, but in this case series none of the viruses were known to infect human cells. In this thesis, I elucidate numerous roles for the microbiome in pediatric patients with IBD and lung-transplant recipients, show exciting new finding about engineering the microbiota to help with hyperammonia, and finally investigate a possible limitation about using microbial communities as therapeutics. Together this body of work provides insights into the assemblage of tiny organisms that live within us, constantly contributing.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Cell & Molecular Biology

First Advisor
Frederic D. Bushman

THE IMPACT OF TINY ORGANISMS: MICROBIAL COMMUNITIES AND DISEASE STATES

Christel Sjöland Chehoud

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

_____

Frederic D. Bushman, Ph.D.

Professor, Department of Microbiology

Co-Supervisor of Dissertation

_____

Gary D. Wu, M.D.

Professor, Department of Gastroenterology

Graduate Group Chairperson

_____

Daniel S. Kessler, PhD.

Associate Professor of Cell and Developmental Biology

Dissertation Committee

Paul Bates, Ph.D., Professor, Department of Microbiology

Ronald G. Collman, M.D., Professor, Department of Medicine

Elizabeth Grice, Ph.D., Assistant Professor, Department of Dermatology

Mark Goulian, Ph.D. Professor, Department of Biology

THE IMPACT OF TINY ORGANISMS: MICROBIAL COMMUNITIES AND DISEASE STATES

DEDICATION

To my family, my secret weapon.

ACKNOWLEDGMENT

I would like to extend my deepest gratitude to all those who supported me throughout my years in graduate school. I am grateful to my mentors, Rick Bushman and Gary Wu. Rick's wisdom, kindness, and enthusiasm for all things related to science have been pivotal in my development as a scientist and as a person. Gary's ability to find insight in any dataset has taught me the importance of thinking critically. Elizabeth Grice, my "third" advisor, has always eagerly provided encouragement and guidance. Paul Bates has supported me through my MVP years, always willing to share his endless knowledge on science-related matters. I am indebted to Ron Collman, Mark Goulian, Jim Lewis and Hognzhe Li for their feedback and advice, and to Sara Cherry for her rigorousness which helped me pass my prelims.

This work would not have been possible without the supportive and collaborative environment of the Bushman Lab. I thank Aubrey Bailey, Kyle Bittinger, Erik Clarke, Serena Dollive, Anatoly Drygra, Nirav Malani, Sam Minot, Scott Sherrill-Mix, and Rohini Sinha for their computational insights. I am appreciative of Laurie Zimmerman and Arwa Abbas's artistic talents. I am grateful to my wonderful baymates, Alexandra Bryson and Jacque Young. The analysis could not have been done without the efforts of the experimental technicians in our lab, Stephanie Grunberg, Chris Hoffmann, Young Hwang, Abby Lauder, Alice Laughlin, and Frances Male, and rotation students, Amy Davis, Elizabeth Loy, and Seth Zost. I am also thankful for invaluable insight and advice from Fabiana Hoffmann, Brendan Kelly, Chris Nobles, and Vesa Turkki.

Members of the Wu lab, including Lindsey Albenberg, Lillian Chau, Elliot Friedman, Colleen Judge, Sue Keilbaugh, Josie Ni, David Shen and Sarah Smith, have taught me everything I know about gastroenterology and provided me with a rich source of collaboration. I would like to acknowledge members of Grice lab, Geoffrey Hannigan, Adam SanMiguel, and Amanda Tyldsley, who were there for me during my first rotation and remained my friends throughout graduate school.

Emily Roberts prepared me for the immunology portion of my prelim and has been a great motivator in science and in life. I am grateful to my GCB friends, Eric Chen and Yuchao Jiang, and, particularly, Varun Aggarwala who has been a source of support and advice since the very beginning of my graduate school years.

I am grateful to my Whitehall High School friends, Alex Beers, Hector Moreno, Kaitlin Reed, Liz Rice, and Jill Richards, teachers, Mr. Adams and Mr. Stout, and counselor, Ms. McGill. I am indebted to my Princeton mentors, David Botstein, Tullis Onstott, and Bess Ward, and Integrated Science friends. My Broad mentors, Shawna Young, Eboney Smith, Bruce Birren, Brian Haas, and Dirk Gevers, introduced me to the microbiome back in 2008 and sparked this 8 year adventure. I am additionally grateful to my colleagues at Second Genome, particularly Todd DeSantis, Janet Warrington and Justin Kuczynski.

Finally, I am thankful for the unconditional love, encouragement and support of my family, both nearby and around the world in Brazil, Canada, Lebanon, Sweden, and Venezuela. My parents, Albert and Georgette, sister, Grace, and husband, Erik, have been my biggest blessing during my graduate years, and throughout my life.

ABSTRACT

THE IMPACT OF TINY ORGANISMS: MICROBIAL COMMUNITIES AND DISEASE STATES

Christel Sjöland Chehoud

Frederic D. Bushman

Gary D. Wu

In the last decade, primarily through the use of sequencing, much has been learned about the trillions of microorganisms that reside in human hosts. These microorganisms play a wide range of roles including helping our immune systems develop, digesting our food, and protecting us from the invasion of pathogenic organisms. My thesis focuses on the characterization of fungal, viral, and bacterial communities in humans, investigating the use of defined microbial communities to cure diseases in animal models, and examining the effects of human microbiome modifications through fecal microbiota transfers. In the first part of this thesis, I use deep sequencing of ribosomal RNA gene tags to characterize the composition of the bacterial, fungal, and archaeal microbiota in pediatric patients with Inflammatory Bowel Disease and healthy controls. Archaeal reads were rare in the pediatric samples, whereas an abundant amount of fungal reads was recovered. Pediatric IBD was found to be associated with reduced diversity in both fungal and bacterial gut microbiota, and specific *Candida* taxa were increased in abundance in the IBD samples. I, then, describe my use of a variety of experimental and computational methods to study the viral communities of immune-compromised lung transplant recipients. Anelloviruses, circular, single-stranded DNA viruses, were found in all lung samples but were 56 times more abundant in samples from lung transplant recipients as compared to healthy controls or HIV+ subjects. In the third part of this thesis, I describe the use of defined microbial communities in mice, and its ability to reduce the production of ammonia long term and mitigate hepatic encephalopathy. This was shown to be true in both mice on a normal protein diet or a low protein diet. Last, I investigate the transfer of viral communities between humans through FMT

and characterize features associated with efficient transmission. A case series where feces from a single donor were used to treat three children with ulcerative colitis was used for the analysis. Ultimately this work showed that multiple viral lineages do transfer between human individuals through fecal microbiota transplants, but in this case series none of the viruses were known to infect human cells. In this thesis, I elucidate numerous roles for the microbiome in pediatric patients with IBD and lung-transplant recipients, show exciting new finding about engineering the microbiota to help with hyperammonia, and finally investigate a possible limitation about using microbial communities as therapeutics. Together this body of work provides insights into the assemblage of tiny organisms that live within us, constantly contributing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: Introduction

The term “microbiome” was coined by Joshua Lederberg to describe the ecological community of commensal, symbiotic, and pathogenic microorganisms that live in or on human bodies (1). These trillions of microorganisms represent an ensemble of bacteria, archaea, eukaryotes, and viruses. These organisms play important roles in human development, physiology, immunity, disease, and nutrition. Many of these finding have been obtained over the last decade, in part, thanks to the development of culture-independent sequencing technology that allows us to investigate our microbial companions in an unbiased fashion. In this introductory chapter, I provide an overview of sequencing-based microbiome research, the tools used in the field, and the use of human stool for therapy of human disease.

Human Microbiome Project

The National Institute of Health launched the Human Microbiome Project (HMP) in 2008 (2). This $148 million endeavor sought to investigate the microbial diversity of 5 body sites: mouth, nose, skin, vagina, and gut, in healthy American adults. As part of this initiative, they also set out to sequence reference bacterial genomes, develop analytical and visualization tools, and standardized protocols for the entire scientific community to use. The first phase of HMP’s work was meant to, and has, served as the foundation for subsequent human microbiome work, particularly when examining differences in microbial communities between healthy and diseased individuals. The core questions of the HMP are: (1) is there a core human microbiome that is shared by all humans; (2) is there a core set of microbial functions that are shared by all humans; (3) to what extent does human phenotype or behavior impact the microbiome; (4) are changes in the microbiome causing disease; and (5) can disease be treated by altering the microbiome (3) . These questions have multi-faceted answers that we are still trying to answer today.

Analytical Tools

Carl Woese revolutionized the field of microbiology when he described using the sequence of the small subunit ribosomal RNA gene (rRNA) for phylogenetic identification of organisms (4). The 16S rRNA, which is ubiquitously present in bacteria and archaea, and its 18S rRNA counterpart in eukaryotes, became the main tool used by microbiologists to explore microbial diversity and identify microorganisms throughout the body. DNA is typically isolated from stool, the main surrogate for the gastrointestinal tract's microbiome, or other body sites including skin, vagina, and nose, and upper and lower respiratory tracts. 16S rRNA gene sequences, typically 1300 to 1600 base pairs in length, contain both conserved regions, which can be used for primer landing sites, as well as variable regions, which can be used for species-level identification (5, 6). This allows us to sequence a small part of a genome but identify an entire organism.

16S rRNA gene sequencing is the most commonly used method to identify bacteria when using amplicon-based sequencing. To identify archaeal sequences, 16S rRNA and rpoB genes are typically used, while for fungal sequences, 18S rRNA and Internal Transcribed Spacers (ITS) genes are commonly used. Since there is no gene ubiquitously present in all viruses, we cannot perform targeted gene sequencing for viruses.

To perform amplicon-based sequencing, PCR reactions containing primers directed against conserved regions of the targeted gene are used to generate amplicons that span both the conserved and variable regions of the gene. Depending on the capacity of the sequencing technology, the amplicons can span a different number of variable regions in the targeted gene. The amplicons from each sample, which contain a unique barcode to distinguish the samples, are pooled, sequenced, and computationally analyzed.

Computational analysis of targeted sequence data involves several stages. First, the reads are grouped into operational taxonomic units (OTUs) based on sequence identity. A representative sequence is used to represent each OTU, typically the most abundant sequence

or the longest sequence. OTUs can be defined at different sequencing similarity thresholds; for example, a 97% OTU signifies that all of its member sequences are at least 97% similar based on sequence identity. The representative OTU sequences are then aligned to reference sequences and annotated using databases such as Greengenes or Ribosomal Database Project (7-10). The alignments of the representative OTU sequences are also used to build a phylogenetic tree. Popular distance metrics, such as Weighted and Unweighted Unifrac (11, 12), rely on the phylogenetic distance between representative sequences when assessing differences in samples' microbial compositions. Distances, based on each sample's quantity of sequences in each OTU, are calculated and visualized using hierarchical clustering and ordinations (13).

As sequencing technology continues to improve, efforts have moved from targeted amplification of select genes to shotgun sequencing of entire microbial communities. Metagenomics refers to the untargeted sequencing and analysis of DNA from a mixture of organisms. This is particularly useful since it can provide insight into the community's gene content and corresponding pathways and metabolites. This technique involves randomly sheering DNA fragments and sequencing the resulting fragments (14). These resulting reads can be queried independently or assembled into contigs. The reads or contigs can be compared to databases of known genes including KEGG and COG (15-17). Programs such as metaphlan2 utilize unique clade-specific marker genes to match reads to specific bacteria, archaea, eukaryotes and viruses (18). The resulting microbial community composition of each sample can be visualized with heatmaps, to assess abundance or proportional abundance of the microorganism present, or compared using hierarchical clustering and ordinations.

A major limitation to these methods is that we can only identify organisms that have been previously identified and placed into databases. A way to circumvent this is to use the contigs, or assembled genomes from reads. This is particularly common in the case of viruses. For viral analysis, we typically use experimental techniques to filter out human and bacterial cells prior to DNA extraction. After we have isolated DNA, we check for contaminating human and bacterial

DNA by quantifying beta-tubulin and 16S rRNA genes respectively. This ensure that there are at most low levels of human and bacteria DNA, so that the sample is mostly viral DNA (19). Then, after we have performed the sequencing, we take the resulting reads and assemble them into contigs through traditional overlap consensus methods (20), de Bruijn graph assembly (21, 22) or self-organizing map binning (23-25). Using the assembled contigs from all of the samples, we can map the reads from each sample to these contigs. This will result in a matrix of counts of matches; this can be analyzed and visualized as before with heatmaps, ordinations, and clustering methods. We refer to these contigs as “putative full or partial viral genomes”. In most cases, they do not have taxonomic annotations, but we can almost always predict open reading frames (ORFs) in the genomes (26, 27).

Culture-independent sequencing technology has allowed us to profile the bacterial, fungal, archaeal, and viral communities within healthy individuals and patients with different diseases. New tools need to be developed to help us continue our shift from characterization to testing hypotheses and finding causal links between the microbiome and disease. In the next section, I describe microbiome therapeutic development, in the form of fecal microbiota transplants, for several disorders in the clinical trials stage.

Translating Microbiome Science into the Clinic

Fecal microbiota transplantation (FMT) is the infusion of a fecal suspension from a healthy individual into the gastrointestinal tract of a sick person. This procedure is thought to restore a healthy, diverse microbial gut community and treat diseases with altered commensal bacterial communities. FMTs are best known for their high efficacy in the treatment of recurrent *Clostridium difficile* infections but have shown promise for many other disorders including antibiotic-associated diarrhea, inflammatory bowel diseases, irritable bowel syndrome, metabolic syndrome, autoimmune diseases, allergic diseases, pulmonary-intestinal-associated inflammatory diseases, and genitourinary tract-intestinal-associated inflammatory diseases (28). Up until recently, FMTs were exclusively performed through enemas, nasogastric tubes,

colonoscopy, endoscopy, or sigmoidoscopy. Industry has now developed orally-administrable capsules filled with fecal suspension or freeze-dried preparations, to allow more convenient delivery for patients (29, 30).

History of FMT

Literary evidence for FMTs dates back to 4th century China. Ge Hong, a traditional Chinese medical doctor, described positive results from the use of human fecal suspension by mouth for patients who had food poisoning or severe diarrhea (29, 31). In 16th century China, Li Shizhen describes using fermented fecal solution, fresh fecal suspension, dry feces, and infant feces to treat severe diarrhea, fever, pain, vomiting, and constipation (32). A 17th century Swedish text also describes veterinarians transplanting microorganisms from healthy to sick animals (33, 34). Since the 1950s, FMT have been used in the United States sporadically without much regulation (35-37). Since the early 2010s, interest in FMTs has dramatically increased with many studies assessing its efficacy for a wide variety of diseases (38); this has resulted in several FDA regulation and enforcement policies surrounding FMTs.

FMT and *C. difficile*

The first rigorous randomized controlled trial utilizing FMTs was published in January 2013 and showed 93.8% efficacy rate for FMTs to treat recurring *C. difficile* infections (37). The spore-forming *C. difficile* bacterium becomes more abundant after long-term, continued use of antibiotics. *C. difficile* produces toxins which cause the patient to suffer from diarrhea and colonic disease, and can lead to death (39). Restoring a person's microbiome with diverse microorganisms through an FMT provides an environment where other bacteria can better outcompete *C. difficile*, effectively reducing it back to healthy levels (40).

IBS and IBD

Given the immense success of FMT for treating *C. difficile*, FMT have been investigated for several other diseases including refractory diarrhea-dominant irritable bowel syndrome (41, 42) and Inflammatory Bowel Diseases (IBD). IBD is a chronic relapsing inflammatory disorder of the intestine and includes both Ulcerative Colitis (UC) and Crohn's Disease (CD). Many studies have implicated the intestinal microbiota in the pathophysiology of IBD, suggesting that bacterial dysbiosis in a susceptible host may lead to inflammation through a complex interplay of factors (40, 43-45). Since mouse models have shown bacteria can activate T-regulatory cells and directly contribute to the reduction in inflammation (45-48), it is thought that FMT can be used for IBD to modulate the gut microbiome and improve the existing pathologic inflammatory state. Studies for FMT and IBD have had mixed results (49). Large scale, phase 3 clinical trials are currently underway for both CD and UC (50, 51).

Obesity, Type 2 Diabetes, and Non-Alcoholic Liver Disease

Multiple descriptive analyses have suggested an association between obesity and the gut microbial community in both human and rodent models (52-57). Along with early work showing that cecal content transplants could alter adiposity in murine models (52), the seminal contribution to this area came from Jeff Gordon's lab where colonization of germ-free mice with a stool from obese mice resulted in greater increase in total body fat than colonization with stool from lean mice (54, 57). Follow-up to this work has further shown causal links between the microbiome and obesity and have highlighted the potential contribution of microbiome-modulated immune responses to the development of obesity and its associated complications in both rodent models and large-scale human studies (53, 58-65).

Non-alcoholic fatty liver disease (NAFLD), one of the most common causes of chronic liver disease worldwide (66), is a known co-morbidity of obesity and type 2 diabetes mellitus (T2D) (67). Small intestinal bacterial overgrowths of Gram-negative organisms have been shown to promote insulin resistance, increase endogenous ethanol production and induce choline deficiency, all factors implicated in obesity, T2D, and NAFLD (68). Several clinical trials are

currently underway where subjects receive FMT, via capsules or nasoduodenal tube, from lean, metabolically healthy donors, and the resulting effects on body weight, insulin sensitivity, and/or amount of fat in the liver are studied (69-71). Given FMT's success in treating refractory *C. difficile* infections, clinical trials are currently underway to investigate FMTs' efficacy in treating patients with IBD, IBS, obesity, type 2 diabetes, and NAFLD.

Thesis Outline

Sequencing has greatly increased our ability to query entire communities of microbial DNA in any part of the human body and investigate how they change with different diseases. In the first part of this thesis, I investigate the bacteria, archaea and fungi in healthy children and children with IBD using next-generation sequencing technology. I assess overall diversity and use the abundance of the microorganisms to predict healthy versus IBD. I also investigate correlations between the microorganisms. This was all done to better understand the relationship between the fungi and bacteria in the context of IBD.

In the second part of this thesis, I deploy novel computational methods to investigate DNA viruses within the respiratory tract. Metagenomic analysis of allograft bronchoalveolar lavage from lung-transplant recipients was compared to healthy and HIV+ subjects. The viral communities were compared to the bacterial members and the patient metadata.

Chapters 4 and 5 describe the use of rodent models to assess the efficacy of engineered microbial communities in reducing long-term production of ammonia and mitigating hyperammonemia-associated neurotoxicity and encephalopathy. In the intestine, bacterial urease converts host-derived urea to ammonia and carbon dioxide; thus, we investigated durability of an engineered community lacking urease capabilities, both in mice on a normal protein diet or a low protein diet, and the long-term effects the new community had on fecal ammonia levels, mortality, and neurobehavioral deficits in acute and chronic liver injury models.

In the last part of thesis, I ask an important question concerning viral transfer during FMT. Using a sample set of 3 subjects who each received between 22 and 30 FMTs, I search for viruses present in the donor but not in any of the recipients' pre-FMT samples. I further develop the viral analysis pipeline discussed in Chapter 3 and investigate similarity between viruses transferred to assess if certain viral families or viral genes make viruses more likely to transfer. This work is important when assessing the safety of FMT, in particular given its increased use.

Overall, I leveraged recent advances in next-generation sequencing technology to conduct metagenomic analyses of microorganisms in several different diseases. I discuss the characterization of these microorganisms and how these microorganisms can be modulated to alleviate disease. I end by describing the viral communities that may play a role in modulating human gut bacterial communities.

References

1. McCray JLaAT. 'Ome Sweet' Omics-- A Genealogical Treasury of Words. The Scientist; 2001:8-9
2. Group NHW, Peterson J, Garges S, et al. The NIH Human Microbiome Project. Genome Res. 2009;19:2317-2323
3. NIH HMP Roadmap 2009. Available at: http://nihroadmap.nih.gov/hmp
4. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A. 1977;74:5088-5090
5. Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. J Clin Microbiol. 2007;45:2761-2764
6. Patel JB. 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. Mol Diagn. 2001;6:313-321
7. Cole JR, Wang Q, Fish JA, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 2014;42:D633-642
8. DeSantis TZ, Dubosarskiy I, Murray SR, et al. Comprehensive aligned sequence construction for automated design of effective probes (CASCADE-P) using 16S rDNA. Bioinformatics. 2003;19:1461-1468
9. DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72:5069-5072
10. Wang Q, Garrity GM, Tiedje JM, et al. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73:5261-5267
11. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol. 2005;71:8228-8235
12. Lozupone C, Lladser ME, Knights D, et al. UniFrac: an effective distance metric for microbial community comparison. ISME J. 2011;5:169-172
13. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7:335-336
14. Venter JC, Remington K, Heidelberg JF, et al. Environmental genome shotgun sequencing of the Sargasso Sea. Science. 2004;304:66-74

15. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27-30
16. Kanehisa M, Sato Y, Kawashima M, et al. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44:D457-462
17. Tatusov RL, Galperin MY, Natale DA, et al. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000;28:33-36
18. Segata N, Waldron L, Ballarini A, et al. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods. 2012;9:811-814
19. Thurber RV, Haynes M, Breitbart M, et al. Laboratory procedures to generate viral metagenomes. Nat Protoc. 2009;4:470-483
20. Treangen TJ, Sommer DD, Angly FE, et al. Next generation sequence assembly with AMOS. Curr Protoc Bioinformatics. 2011;Chapter 11:Unit 11 18
21. Simpson JT, Durbin R. Efficient de novo assembly of large genomes using compressed data structures. Genome Res. 2012;22:549-556
22. Peng Y, Leung HC, Yiu SM, et al. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics. 2012;28:1420-1428
23. Dick GJ, Andersson AF, Baker BJ, et al. Community-wide analysis of microbial genome sequence signatures. Genome Biol. 2009;10:R85
24. Nagarajan N, Pop M. Sequence assembly demystified. Nat Rev Genet. 2013;14:157-167
25. Vazquez-Castellanos JF, Garcia-Lopez R, Perez-Brocal V, et al. Comparison of different assembly and annotation tools on analysis of simulated viral metagenomic communities in the gut. BMC Genomics. 2014;15:37
26. Delcher AL, Bratke KA, Powers EC, et al. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics. 2007;23:673-679
27. Salzberg SL, Delcher AL, Kasif S, et al. Microbial gene identification using interpolated Markov models. Nucleic Acids Res. 1998;26:544-548
28. Borody TJ, Khoruts A. Fecal microbiota transplantation and emerging applications. Nat Rev Gastroenterol Hepatol. 2012;9:88-96
29. Zhang F, Luo W, Shi Y, et al. Should we standardize the 1,700-year-old fecal microbiota transplantation? Am J Gastroenterol. 2012;107:1755; author reply p 1755-1756
30. Youngster I, Russell GH, Pindar C, et al. Oral, capsulized, frozen fecal microbiota transplantation for relapsing Clostridium difficile infection. JAMA. 2014;312:1772-1778
31. Fang ZHBJ. Ge H (Dongjin Dynasty). Tianjin: Tianjin Science & Technology Press; 2000
32. Mu BCG. Li S (Ming Dynasty). Beijing: Huaxia Press; 2011
33. Brag S, Hansen HJ. Treatment of ruminal indigestion according to popular belief in Sweden. Rev Sci Tech. 1994;13:529-535
34. DePeters EJ, George LW. Rumen transfaunation. Immunol Lett. 2014;162:69-76
35. Eiseman B, Silen W, Bascom GS, et al. Fecal enema as an adjunct in the treatment of pseudomembranous enterocolitis. Surgery. 1958;44:854-859
36. Khoruts A, Dicksved J, Jansson JK, et al. Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea. J Clin Gastroenterol. 2010;44:354-360
37. van Nood E, Vrieze A, Nieuwdorp M, et al. Duodenal infusion of donor feces for recurrent Clostridium difficile. N Engl J Med. 2013;368:407-415
38. Rohlke F, Stollman N. Fecal microbiota transplantation in relapsing Clostridium difficile infection. Therap Adv Gastroenterol. 2012;5:403-420
39. Bartlett JG, Chang TW, Gurwith M, et al. Antibiotic-associated pseudomembranous colitis due to toxin-producing clostridia. N Engl J Med. 1978;298:531-534
40. Bowman KA, Broussard EK, Surawicz CM. Fecal microbiota transplantation: current clinical efficacy and future prospects. Clin Exp Gastroenterol. 2015;8:285-291
41. Aroniadis O. Fecal Microbiota Transplantation for the Treatment of Diarrhea-Predominant Irritable Bowel Syndrome. 2014. Available
42. Collins SM. A role for the gut microbiota in IBS. Nat Rev Gastroenterol Hepatol. 2014;11:497-505

43. Gevers D, Kugathasan S, Denson LA, et al. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe. 2014;15:382-392
44. Lewis JD, Chen EZ, Baldassano RN, et al. Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease. Cell Host Microbe. 2015;18:489-500
45. Ott SJ, Plamondon S, Hart A, et al. Dynamics of the mucosa-associated flora in ulcerative colitis patients during remission and clinical relapse. J Clin Microbiol. 2008;46:3510-3513
46. Atarashi K, Tanoue T, Shima T, et al. Induction of colonic regulatory T cells by indigenous Clostridium species. Science. 2011;331:337-341
47. Damman CJ, Miller SI, Surawicz CM, et al. The microbiome and inflammatory bowel disease: is there a therapeutic role for fecal microbiota transplantation? Am J Gastroenterol. 2012;107:1452-1459
48. Round JL, Lee SM, Li J, et al. The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. Science. 2011;332:974-977
49. Colman RJ, Rubin DT. Fecal microbiota transplantation as therapy for inflammatory bowel disease: a systematic review and meta-analysis. J Crohns Colitis. 2014;8:1569-1581
50. Russell G. Fecal Microbiota Transplant (FMT) in Pediatric Active Crohn's Colitis. 2014.
51. Zhang F. Standardized Fecal Microbiota Transplantation for Ulcerative Colitis. 2013.
52. Backhed F, Ding H, Wang T, et al. The gut microbiota as an environmental factor that regulates fat storage. Proc Natl Acad Sci U S A. 2004;101:15718-15723
53. Harley IT, Karp CL. Obesity and the gut microbiome: Striving for causality. Mol Metab. 2012;1:21-31
54. Ley RE, Backhed F, Turnbaugh P, et al. Obesity alters gut microbial ecology. Proc Natl Acad Sci U S A. 2005;102:11070-11075
55. Ley RE, Turnbaugh PJ, Klein S, et al. Microbial ecology: human gut microbes associated with obesity. Nature. 2006;444:1022-1023
56. Samuel BS, Gordon JI. A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism. Proc Natl Acad Sci U S A. 2006;103:10011-10016
57. Turnbaugh PJ, Ley RE, Mahowald MA, et al. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444:1027-1031
58. Arumugam M, Raes J, Pelletier E, et al. Enterotypes of the human gut microbiome. Nature. 2011;473:174-180
59. Cani PD, Amar J, Iglesias MA, et al. Metabolic endotoxemia initiates obesity and insulin resistance. Diabetes. 2007;56:1761-1772
60. Cani PD, Bibiloni R, Knauf C, et al. Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet-induced obesity and diabetes in mice. Diabetes. 2008;57:1470-1481
61. Cani PD, Possemiers S, Van de Wiele T, et al. Changes in gut microbiota control inflammation in obese mice through a mechanism involving GLP-2-driven improvement of gut permeability. Gut. 2009;58:1091-1103
62. Henao-Mejia J, Elinav E, Jin C, et al. Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. Nature. 2012;482:179-185
63. Turnbaugh PJ, Hamady M, Yatsunenko T, et al. A core gut microbiome in obese and lean twins. Nature. 2009;457:480-484
64. Vijay-Kumar M, Aitken JD, Carvalho FA, et al. Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. Science. 2010;328:228-231
65. Wu GD, Chen J, Hoffmann C, et al. Linking long-term dietary patterns with gut microbial enterotypes. Science. 2011;334:105-108
66. Milic S, Stimac D. Nonalcoholic fatty liver disease/steatohepatitis: epidemiology, pathogenesis, clinical presentation and treatment. Dig Dis. 2012;30:158-162
67. Vanni E, Bugianesi E, Kotronen A, et al. From the metabolic syndrome to NAFLD or vice versa? Dig Liver Dis. 2010;42:320-330
68. Abu-Shanab A, Quigley EM. The role of the gut microbiota in nonalcoholic fatty liver disease. Nat Rev Gastroenterol Hepatol. 2010;7:691-701

69. Yu EW. Fecal Microbiota Transplant for Obesity and Metabolism. 2015.
70. Silverman M. Transplantation of Microbes for Treatment of Metabolic Syndrome & NAFLD (FMT). 2015.
71. Zhang F. Fecal Microbiota Transplantation on Type 2 Diabetes Mellitus. 2013.

CHAPTER 2: Fungal Signature in the Gut Microbiota of Pediatric Patients with Inflammatory Bowel Disease

The contents of this chapter have been published as:

> **Chehoud C***, Albenberg L*, Judge C, Hoffmann C, Grunberg S, Bittinger K, Baldassano R, Lewis J, Bushman F, Wu G. 2015. *Fungal Signature in the Gut Microbiota of Pediatric Patients with IBD.* Inflammatory Bowel Diseases Journal (PMID: 26083617).

Abstract

Inflammatory bowel disease (IBD) involves dysregulation of mucosal immunity in response to environmental factors such as the gut microbiota. The bacterial microbiota is often altered in IBD, but the connection to disease is not fully clarified, and gut fungi have recently been suggested to play a role as well. In this study, we compared microbes from all three domains of life--bacteria, archaea, and eukaryota--in pediatric patients with IBD and healthy controls. A stool sample was collected from patients with IBD (n=34) or health control subjects (n=90), and bacterial, archaeal, and fungal communities were characterized by deep sequencing of rRNA gene segments specific to each domain. IBD patients (Crohn's disease or ulcerative colitis) had lower bacterial diversity and distinctive fungal communities. Two lineages annotating as *Candida* were significantly more abundant in IBD patients ($p = 0.0034$ and $p=0.00038$, respectively) while a lineage annotating as *Cladosporium* was more abundant in healthy subjects ($p=0.0025$). There were no statistically significant differences in archaea, which were rare in pediatric samples compared to those from adults. Pediatric IBD is associated with reduced diversity in both fungal and bacterial gut microbiota. Specific *Candida* taxa were increased in abundance in the IBD samples. These data emphasize the potential importance of fungal microbiota signatures as biomarkers of pediatric IBD, supporting their possible role in disease pathogenesis.

Introduction

The three domains of life--bacteria, archaea, and eukaryota--are all represented in human gut communities (1-5). Numerous disorders, both intestinal and systemic, have been associated with alterations in microbial community structure, termed “dysbiosis” (6-8). The dysbiotic bacterial microbiota has been extensively characterized for the inflammatory bowel diseases (IBD) and *Clostridium difficile* infection, although the composition of the archaea and eukaryota domains and their interaction with bacteria has not been studied in detail. The pathogenesis of IBD is associated with an inappropriate and persistent inflammatory response to the commensal gut microbiota in genetically susceptible individuals. In animal studies, dysbiosis has been associated with the development of intestinal inflammation (reviewed in (6, 8, 9)). Many of the genetic risk alleles for IBD involve regulation of innate immune responses that protect the host from bacterial invasion or that regulate the adaptive immune system (reviewed in (10, 11)). In Crohn’s disease (CD), the phylum Firmicutes is commonly reduced in proportional abundance (12-17), notably *Faecailbacterium prausnitzii* (18-20) and members of the Proteobacteria phylum such as *Enterobacteriaceae* (21, 22), including *E. coli* (7, 17, 23) are commonly increased.

The importance of fungi in the human gut to health or disease is not fully understood (2, 24). Fungi, such as *Aspergillus*, *Histoplasma*, and *Cryptococcus* are known to cause pathogenic intestinal infections. Fungi are also part of the commensal gut microbiota. Increases in gut *Candida* species are hypothesized to cause systemic yeast infections in immunocompromised patients via translocation through the intestinal barrier (25). Fungal-bacterial interactions are just beginning to be described (1, 5, 6, 26-28), and this may be important in IBD (2).

Several lines of evidence link fungi and IBD. Innate immune receptor activation by fungi may augment the development of colitis. The main pattern recognition receptors for fungi include Dectin-1, Dectin-2, DC-SIGN, mannose receptor, and mannose-binding lectin (29). Dectin-1 is a C-type lectin receptor, which recognizes β-glucans in the fungal cell wall. A recent study (30) demonstrated that mice lacking Dectin-1 had increased susceptibility to chemically induced colitis

due to their altered responses to indigenous fungi. A polymorphism in the gene encoding Dectin-1 (*CLEC7A*) was subsequently associated with a severe form of ulcerative colitis (UC) in humans (30). There may also be a relationship between high dietary concentrations of yeast and increased disease activity in patients with CD (31). Treatment with fluconazole, an antifungal, may reduce inflammation in animal models of colitis and in patients with IBD (32), and elevated levels of *Saccharomyces cerevisiae* antibodies are a biomarker for CD (33). Finally, preliminary studies using denaturing gradient gel electrophoresis (DGGE) suggest that there may be alterations in gut fungal populations associated with IBD (25, 34).

Alterations in the archaeal microbiota may also be relevent to IBD. Methane production from archaea has been shown to be altered in patients with IBD (35). Additionally, the syntrophic interaction between archaea and fermentative bacteria leading to the increase in short chain fatty acid (SCFA) production (36) may have an impact on mucosal immunity via to the impact of SCFA on T regulatory cells and immune tolerance (37). Studies of a modest number of subjects are beginning to suggest that fungi may be dysbiotic in IBD (34) as well as bacteria.

In this study, we use deep sequencing of rRNA gene tags to characterize the composition of the bacterial, fungal, and archaeal microbiota in pediatric patients with IBD (n=32) and healthy controls (n=90), revealing a distinctive signature in the mycobiome.

Results

Patient cohort studied

The bacterial component of the microbiome has been studied extensively in IBD, and is known to be dysbiotic (6, 8, 9, 46). Thus, in this study, we sought to assess effects on the fungal and archaeal components of the microbiome. A cross-sectional analysis was performed whereby a single stool sample was obtained from pediatric patients (3 to 21 years of age) with CD, UC, or indeterminate colitis (IC) (n=32). The diagnosis of IBD was based on endoscopic, radiologic, laboratory, and clinical findings. Clinical phenotypes of all patients and their therapies are

described in Supplemental Table 1. Microbiome sequencing results for IBD patients were compared to results for healthy pediatric and adult subjects studied previously (n=90) (40).

Reduced richness in the intestinal bacterial microbiota of patients with IBD

We first characterized the bacterial composition of the microbiota by purifying and amplifying DNA using PCR primers recognizing the V1V2 region of the16S rRNA gene, and pyrosequencing an average of 2,864 reads per sample. We analyzed fecal samples from two IC, four UC, and 26 CD patients (Sup. Table 2.1). Reads were clustered into operational taxonomic units (OTUs) at 97% sequence identity, and representative sequences were aligned to databases for taxonomic attribution. Comparison of data among the disease classes revealed few differences among groups (Sup. Figure 2.1). Most samples were dominated with Bacteroides. A few had Clostridia as the dominant taxa.

We compared a dataset from healthy adults and children to the IBD samples (40). OTU composition was compared in a pairwise fashion between samples using UniFrac, and healthy and IBD sample sets were compared to each other using PERMANOVA. We observed a clear separation between the IBD and healthy sample sets ($p<0.0001$) (Sup. Figure 2.2).

We also investigated the diversity of the IBD samples and healthy controls, since previously several studies have reported lower diversity in IBD (15). We assessed diversity by calculating the Shannon Index for each sample, and found that the IBD samples had significantly lower diversity compared to either healthy adult or pediatric control subjects (p-value <0.05) (Sup. Figure 2.3).

Comparison of archaea in patients with IBD and healthy controls

We compared the archaeal microbiota composition in the 32 IBD samples with 90 healthy controls by amplification with PCR primers selective for archaea, then pyrosequencing and alignment to databases. Only three IBD samples contained detectable Methanobrevibacter (Sup. Figure 2.4), the major archaeal lineage in healthy adults (5). Six samples had a few reads (<5

reads each) mostly matching "Unclassified Archaea". Querying those reads against NCBI's nucleotide database did not provide clearer taxonomic resolution. We validated these findings by quantitative PCR targeting Archaea, which confirmed that only three IBD samples, Subject004_CD, Subject006_CD, and Subject037_CD, had detectable archaea. There was a lower recovery of archaeal sequences from pediatric IBD samples compared to healthy adults ($p<1X10^{-6}$), but pediatric IBD patients were not significantly different from healthy children ($p=0.128$) (Sup. Figure 2.5). This indicates that children have lower archaeal colonization than do adults, and that alterations in the archaeal colonization of the gut are not related to IBD.

A distinctive fungal signature in the microbiota of pediatric patients with IBD

The fungal representation in the fecal samples was characterized by amplification of stool DNA with PCR primers targeting the fungal rRNA ITS region, followed by pyrosequencing using the 454/Roche platform, formation into OTUs at 95% sequence identity, and alignment to databases. The ITS amplicon has been reported to have biases in recovery of certain lineages (47), but it does successfully capture a broad range of taxa (5, 26, 28). Thirty-two IBD subjects were compared to 90 healthy adults and children (40). IBD samples had significantly lower fungal diversity than healthy children (p-value <0.05) (Figure 2.1).

Ordination analysis based on the presence or absence of fungal species revealed a separation between healthy and IBD pediatric subjects ($p=0.004$) (Figure 2.2). Samples from healthy adults overlapped with both pediatric groups (data not shown). Ordination based on the abundance (rather than presence/absence) of each fungal species also showed separation between healthy and IBD pediatric subjects ($p=0.044$) (Sup. Figure 2.6). A focused analysis of CD only, in which the IC and UC patients were excluded, also revealed significant separation between healthy and IBD pediatric subjects (Sup. Figure 2.7 $p< .0001$).

Next, we investigated whether fungal representation differed in IBD and healthy control samples (Sup. Figure 2.3). The most commonly observed order was Saccharomyetales, which contains the genus Candida. IBD samples averaged 2,675 Candida sequences (proportionally

72.9%) while healthy controls averaged 1,320 reads (proportionally 32.9%) (p=0.0107 based on read counts). Healthy adults had slightly higher, non-significant average proportions of Candida than did pediatric healthy subjects (1428 reads (37%) vs. 1055 reads (23%); p=0.535 based on read counts).

The Candida reads were categorized in 97 different OTUs with the majority belonging to one OTU (genbank ID: KP132001) (Figure 2.4A). This OTU annotates as single species, which has previously been given multiple different names, including *Pichia jadinii*, *Candida utilis*, *Cyberlindnera jadinii, Torula yeast, and Candida guilliermondii var. nitratophila*. This OTU was significantly enriched in pediatric IBD patients (p-value<0.01 using raw reads or proportion for both pediatric IBD versus all healthy (adult and pediatric) and pediatric IBD versus pediatric healthy). A second Candida OTU (genbank ID: EF197997), annotating as *Candida parapsilosis*, was also more common in the pediatric IBD samples (Figure 2.4B, p-value <0.01 based on number of reads). A third OTU (genbank ID: KJ596320), annotating as *Cladosporium cladosporioides,* was more common in healthy subjects (pediatric or adult healthy versus pediatric IBD, p-value <0.01 based on number of reads; Figure 2.4C).

A microbial signature for pediatric IBD

Random Forest, a supervised machine-learning algorithm, was used to generate a classifier capable of sorting IBD and healthy controls based on microbial community composition. Classifiers were developed comparing pediatric IBD to pediatric healthy (Figure 2.5A), and pediatric IBD versus pooled adult and pediatric healthy controls (Figure 2.5B). Classifiers were compared that used bacterial 16S rRNA gene data only, fungal ITS sequence data only, or both.

The models successfully partitioned the pediatric samples by disease status (Figure 2.5A) and partitioned all healthy controls (children and adults) from IBD children (Figure 2.5B). For pediatric samples, the16S or ITS OTUs used alone showed a median accuracy of 90% and 83% respectively, while the combination of OTUs showed greater than 93% accuracy. The

bacterial OTUs that could best distinguish IBD patients from healthy controls annotated as *Bacteroides*, *Parabacteroides*, and *Faecalibacterium prausnitzii*. For fungi, the classifier was dominated by OTUs annotating as Pichia jadinii/Candida utilis. The full list of distinguishing taxa is presented in Supplemental Tables 2.3-2.6.

Bacterial and fungal correlations in IBD

Two-sided, Pearson tests were used to measure correlation between bacterial and fungal OTUs across IBD samples. Of the 128 bacterial OTUs tested, 33 showed FDR-corrected, significant (p-value <0.05) correlation with at least one fungal OTU (Figure 2.6). The most commonly found fungal species in IBD subjects, *Candida* OTU KP132001, did not correlate strongly with any bacterial species. Many more significant correlations between fungal and bacterial OTUs were detected in healthy subjects (Sup. Figure 2.8). Evidently fungal and bacterial species co-vary, but this was not prominently associated with the lineages implicated as important in IBD in these subjects.

Discussion

Here, we analyzed the composition and diversity of the fungal, bacterial and archaeal microbiota in pediatric IBD patients compared to healthy controls. In contrast to the many studies of the bacterial component of the gut microbiota, there is relatively little known about the fungal and archaeal microbiota and their roles in IBD. Fungi and archaea are known to be normal components of the gut microbiota (1, 4, 5, 48). In this study, we recovered abundant fungal reads from both IBD patients and healthy controls. In contrast, archaea were rare in the pediatric samples. The potential importance of fungi in IBD is well recognized. Antibodies to baker's yeast (*Saccharomyces cerevisiae*), termed ASCA, are detected more frequently in patients with CD than in healthy controls or in patients with UC (29-69% of patients with CD (33) (49-51)). The fungal antigen recognized is thought to be phosphopeptidomannan of the *Saccharomyces cerevisiae* cell wall (52). ASCA positivity may precede the development of IBD (53), and ASCA are found more commonly in healthy relatives of patients with CD (54). The difference in ASCA

prevalence raises the possibility of a differential fungal microbiota in patients with CD versus healthy individuals, and our data documents such a difference.

Although the ASCA epitope is associated with *S. cerevisiae*, it is also associated with *Candida albicans* (55) and potentially other related yeasts. Candida colonization has been associated with multiple diseases of the gastrointestinal tract including CD, UC, esophagitis, oral mucositis, and even gastric ulcers (56). Candida strains colonize gastric ulcers, particularly when the ulcers are large or perforated (56) and candida esophagitis responds to antifungal therapy demonstrating the role of Candida colonization in the pathogenesis of inflammatory processes. Increased colonization with Candida in IBD might be a consequence of mucosal or immune system alterations as well as the use of antibiotics(28). The health consequences are unknown but may be adverse.

We found three fungi with differential abundance in IBD patients versus healthy controls. Fungal phylogeny is in a state of transition, so species level attributions are not fully secure. For example, until recently, sexual forms (teleomorphs) were commonly placed in separate genera from asexual forms (anamorphs) of the same fungal organisms (57). We found one Candida OTU (genbank ID: KP132001) that was more abundant in patients with IBD, which has been associated with five named species, including *Pichia jadinii*, *Candida utilis*, *Cyberlindnera jadinii*, *Torulopsis utilis*, and *Torula utilis* (58), illustrating the challenges of current fungal taxonomy. A common name for this group is Torula yeast. In its inactive form, Torula yeast is widely used as a flavoring which imparts a savory taste to soups, sauces, and snack products (59). Possibly the pediatric subjects with IBD were eating diets more enriched in Torula yeasts, though there is no universal diet for patients with IBD, so this seems unlikely. Patients with IBD frequently identify dietary components that cause increased symptoms and so follow restricted diets when the disease is active, but most subjects studied here had relatively low disease activity scores and so were not expected to be on specialized diets. Also, a review of dietary records kept by the patients in our study did not reveal an obvious bias towards the consumption of foods rich in

Torula yeast (data not shown). The second explanation is that there is truly increased Torula yeast colonization in patients with IBD, which at present seems more likely. Members of the fungal genus Pichia can be pathogens in humans, especially in immunocompromised hosts (60), Pichia species may also lead to enteritis in animals (61, 62), and pathogenesis by Candida species is well described. Further study is needed to clarify whether colonization is related to disease pathogenesis or whether it is a consequence of gut inflammation or immune suppressive therapy.

We found a second Candida OTU (genbank ID: EF197997) that was more abundant in patients with IBD, and a *Cladosporium cladosporioides* OTU (genbank ID: KJ596320) that was more abundant in healthy controls. All of the IBD patients in our cohort were undergoing various therapies, so it is not clear whether these taxonomic changes were associated with the IBD itself or were a consequence of treatment, for example immunosuppression (63). Other Candida yeasts, in contrast, are used as probiotics (64), raising the possibility that the yeast lineages present contribute a mixture of positive and negative influences.

Studies in animal models suggest a functional role for fungi in intestinal inflammation. Mice lacking the Dectin-1 receptor, a pattern recognition receptor known to identify fungi, were more susceptible to colonic inflammation in a dextran sodium sulfate (DSS) colitis model (30). Administration of *Candida tropicalis* resulted in a more severe DSS-induced colitis in this model, and treatment with fluconazole, an anti-fungal agent, reduced disease(30). Another PRR, SIGNR3, was recently found to be involved in recognizing fungal commensals. Mice lacking SIGNR3 also exhibit increased intestinal inflammation upon exposure to DSS (65). Additional investigation is needed in humans to determine the mechanisms responsible for the observed alterations in the fungal microbiota in patients with IBD and the role they may play in disease pathogenesis via pattern recognition receptors.

To probe distinctive lineages in health and disease, and to develop noninvasive biomarkers, we developed a Random Forest classifier to sort patients with IBD versus healthy

controls. Using a combination of bacterial and fungal OTUs allowed greater than 93% accuracy in distinguishing samples. The currently-available biomarkers focus on the evaluation of patients already diagnosed with IBD such as differentiating CD from UC, differentiating quiescent from active disease, and predicting disease course (66). Drug levels and anti-drug antibodies are also used as biomarkers (66). Unfortunately, the biomarkers currently used to evaluate patients with symptoms of IBD, such as C - reactive protein, erythrocyte sedimentation rate, and fecal calprotectin, are not specific to IBD. Additionally, methods are needed to establish early diagnosis of at-risk individuals, such as asymptomatic family members of patients with IBD. It will be useful to evaluate the performance of the assay targeting both bacteria and fungi described here in these settings.

In a recently described analysis of the archaeal microbiome of healthy human subjects, about 45% of subjects were found to have archaea present in the stool (5). The vast majority of sequences annotated as *Methanobrevibacter* (5). In our study, we found only 3 pediatric patients with IBD with detectable *Methanobrevibacter*. The previous literature has been inconsistent, with some studies reporting increased prevalence of methanogens or decreased methane production in patients with IBD (67-70) and other studies demonstrating increased methane production in patients with IBD (71). The low number of pediatric IBD subjects found to have archaeal species from the stool in our study is consistent with a report of methane production increase with aging (72). Thus our data suggest that archaeal colonization may be more characteristic of the adult gut microbiota.

In conclusion, we report that pediatric IBD was associated with decreased fungal diversity and also altered taxonomic composition. We found an increased proportion of two Candida OTUs in IBD patients. The fungal microbiota may be important in IBD given the increased prevalence of ASCA and the more recent work suggesting a relationship between fungal pattern recognition receptors and IBD in mice. We also report a compositional signature elucidated in the Random Forrest analysis that may be useful in diagnosing IBD patients. Thus, these findings 1)

clarify the existence and nature of fungal dysbiosis in pediatric IBD, 2) pose the question of whether specific yeast lineages may be more pro-inflammatory or colonize associated with inflammation, 3) motivate the investigation of fungal blooms in response to treatment of IBD, and 4) provide a signature of dysbiosis in pediatric patients potentially useful in individualized molecular diagnosis.

Methods

Subjects

All subjects studied were from the Philadelphia area and <21 years of age (Supplemental Table 1). Subjects provided written consent or, when appropriate, assent. Subjects were excluded if they had been treated with antibiotics or probiotics within the two weeks prior to sample collection, if there was presence of an ostomy, or if they had been treated with supplemental nutrition (TPN, enteral nutritional therapy, or other nutritional supplements) accounting for more than 50% of the total caloric intake for the one week prior to sample collection. Pediatric Crohn's Disease Activity Index (PCDAI) (38) or Pediatric Ulcerative Colitis Activity Index (PUCAI) (39) were calculated as described. Stool samples were collected between February and September 2012 and kept frozen at -80° C until they were processed for DNA extraction. The sequence samples obtained from the healthy pediatric and adult subjects and used as controls are described in (5, 40). Briefly, healthy volunteers were screened to be free from any chronic disease, to have a normal bowel movement frequency, and to have a body mass index (BMI) between 18.5 and 35 kg/m^2. One stool sample was provided per subject and kept frozen at -80° C until processed for DNA extraction.

DNA isolation

DNA was isolated from approximately 200 mg of stool using the PSP Spin Stool DNA Plus Kit (Cat #10381102, Invitek, Berlin, Germany) as per the manufacturer's instructions.

16S rDNA gene and ITS1 region gene sequencing

Isolated DNA was quantified using the Picogreen system and 50 ng of DNA was amplified in each PCR reaction. Primers were barcoded to label each sample as described previously (41, 42). PCR reactions were carried out in triplicate using Accuprime (Invitrogen, Carlsbad, CA, USA). Each reaction contained 50 nanograms of DNA and 10 pM of each primer. Primers annealing to the V1V2 region of the 16S bacterial gene were used for amplification. The PCR protocol for 16S amplicons was described previously (43). The development of the archaeal specific 16S rDNA primers and ITS1 fungal primers and PCR cycling conditions are described in [42]. Amplified 16S rDNA and ITS1 fragments were purified using a 1:1 volume of Agencourt AmPure XP beads (Beckman-Colter, Brea, CA, USA). The purified products from the stool samples were pooled in equal amounts and analyzed by pyrosequencing using the Roche/454 Genome Sequencer Junior. DNA pools were separated by amplicon type. DNA-free water was subjected to the same amplification and purification procedure, and no DNA product was observed in agarose gels. Sequences of oligonucleotides used in this study are in Supplemental Table 2.

Sequence Analysis

Sequence data was processed using QIIME (44), augmented by the R package QIIMER (http://cran.r-project.org/web/packages/qiimer). Taxonomy was assigned to the sequences using Ribosomal Database Project (RDP) for 16S (45) and the UNITE database for ITS. Further information on methods can be found in the Supplemental Information.

We fit a Random Forest model to predict health versus disease (IBD) using 16S bacterial OTUs, ITS fungal OTUs, or a combination of the two. One thousand bootstrapped iterations were performed to obtain an estimate of the misclassification rate. The classification error rate was compared to guessing, e.g. if the dataset consisted of 32 subjects with IBD and 68 healthy subjects, a sample would be randomly classified as a sample from a subject with IBD 32% of the time and as a sample from a healthy subject 68% of the time.

Bacterial OTUs with greater than 100 sequences across all IBD samples (128 OTUs) were correlated with ITS fungal OTUs with greater than 100 sequences across all IBD samples (27 OTUs). P-values were calculated using a two-sided Pearson correlation test. Bacterial OTUs with at least one correlation to a fungal OTU producing a Bonferroni-corrected p-value less than 0.05 were displayed on the final heatmap.

Figures

Figure 2.1. Diversity of fungal communities is decreased in IBD patients compared to healthy subjects. The Shannon diversity index was calculated based on the OTU-level classification tables. The boxplots show the distribution of diversity values for: (1) all healthy subjects, (2) only the adult healthy subjects, (3) only the pediatric healthy subjects, and (4) the pediatric IBD subjects. Each black dot represents a different subject. *** p<0.0005 on Wilcox test.

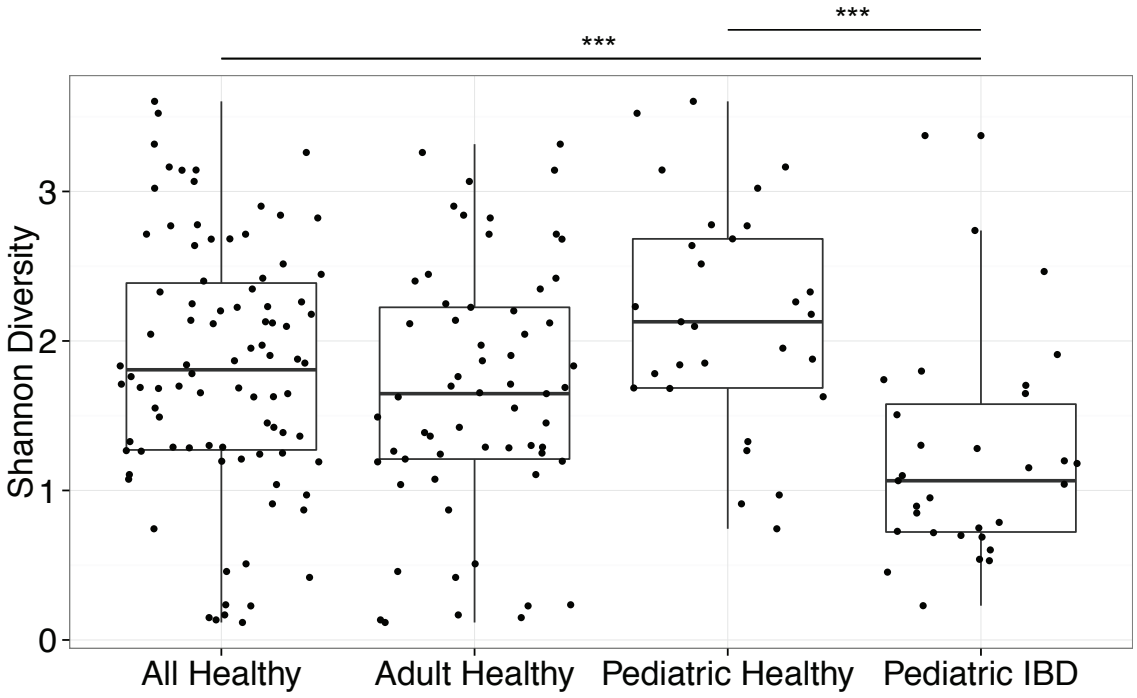


Figure 2.2. Comparison of pediatric healthy and pediatric IBD subjects' fungal community composition using principal coordinate ordination. Principal coordinate analysis was used to depict the relatedness of fungal communities based on presence or absence. The axes represent the two most discriminating axes using the binary Jaccard index distance metric. Pediatric healthy subjects are depicted in cyan and pediatric IBD subjects are depicted in lavender. The two groups clustered separately (p=0.004).

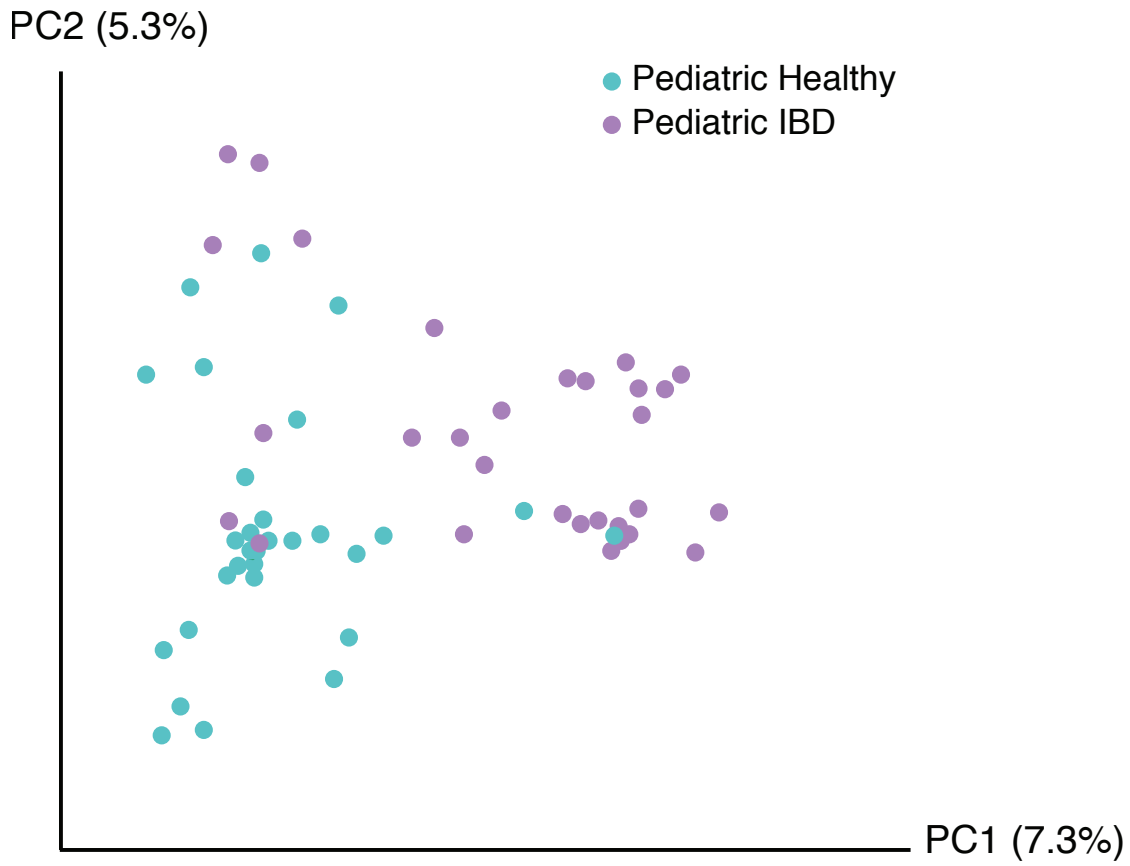


Figure 2.3. Taxonomic heatmap of fungal community members in healthy and IBD subjects. Proportions of fungal OTUs in adult healthy, pediatric healthy, and IBD subjects. The color bar on the top indicates the health status of each subject. Each column indicates a different subject. The color bar on the right side indicates the average relative abundances of these genera in each subject.

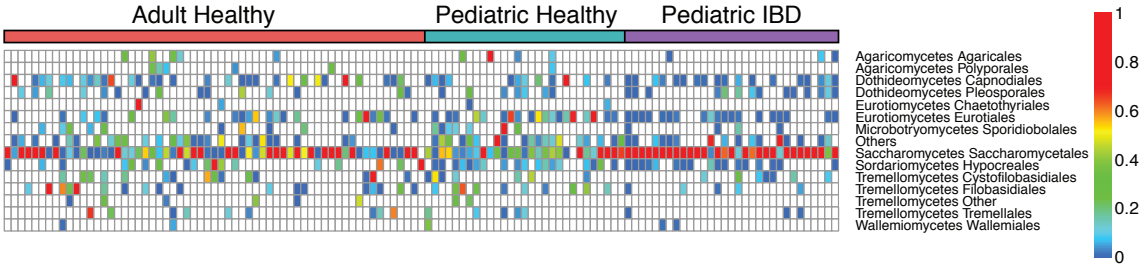


Figure 2.4. Abundance of selected fungal OTUs. Barcharts showing the number of reads from three selected fungal OTUs: (A) Candida (accession KP132001), (B) Candida (accession EF197997) and (C) Cladosporium (accession KJ596320). Adult healthy subjects, pediatric healthy subjects, and pediatric IBD subjects are shown in red, blue, and purple respectively. A negative control sample (NC) that was processed alongside the samples is shown at the right in green. No reads matching any of the three lineages mentioned above were found in the negative control sample.

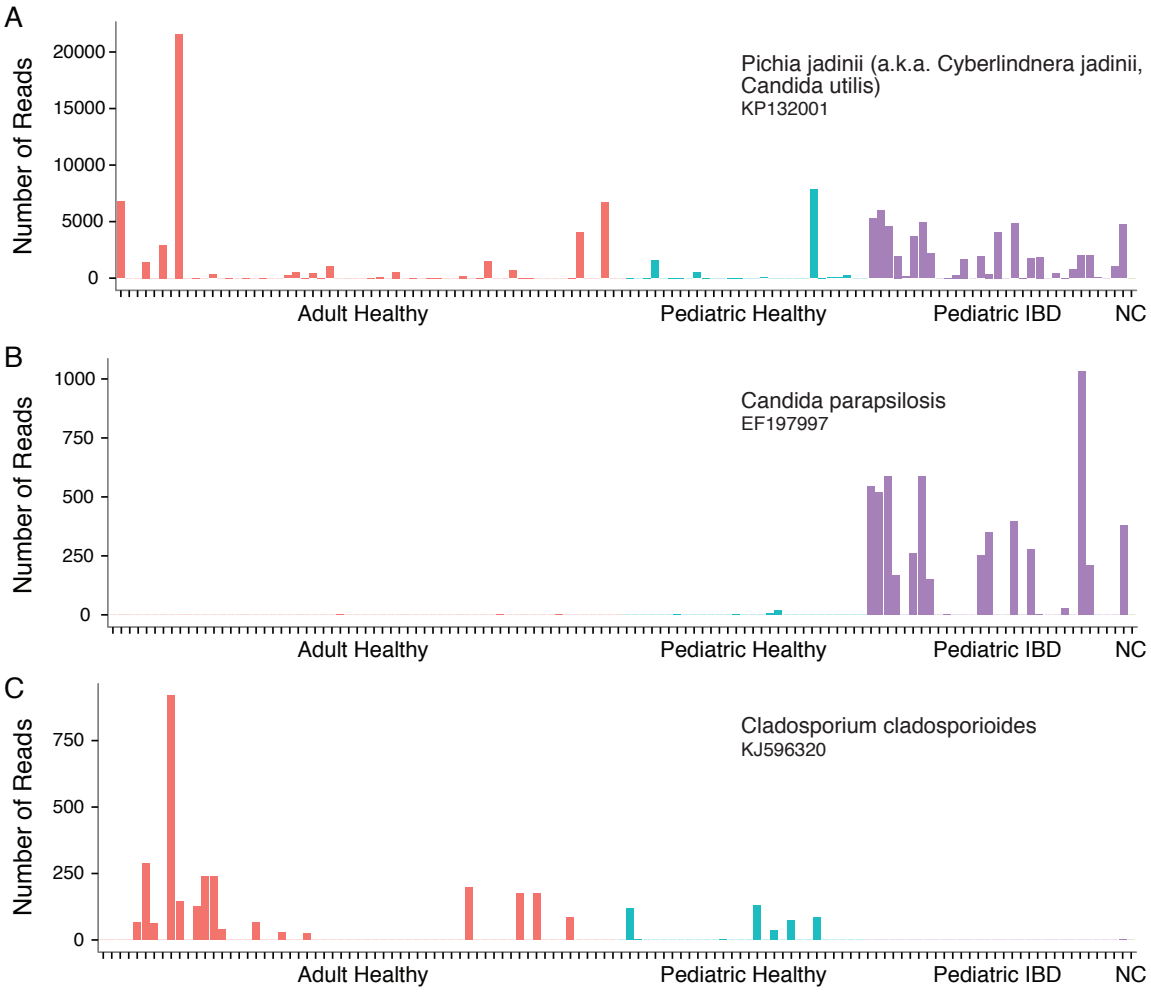


Figure 2.5. Random Forest Classification Accuracy for Healthy and IBD subjects. A random forest classifier was used to group samples into either healthy or IBD categories. Random forest accuracy (red) was compared to random guessing (blue) with misclassification rate indicated on the y-axis. The classifier was run using 16S bacterial OTUs, ITS fungal OTUs and the combination of both bacterial and fungal OTUs. Results were also compared using (A) pediatric healthy and pediatric IBD subjects or (B) both adult and pediatric healthy subjects and IBD subjects.

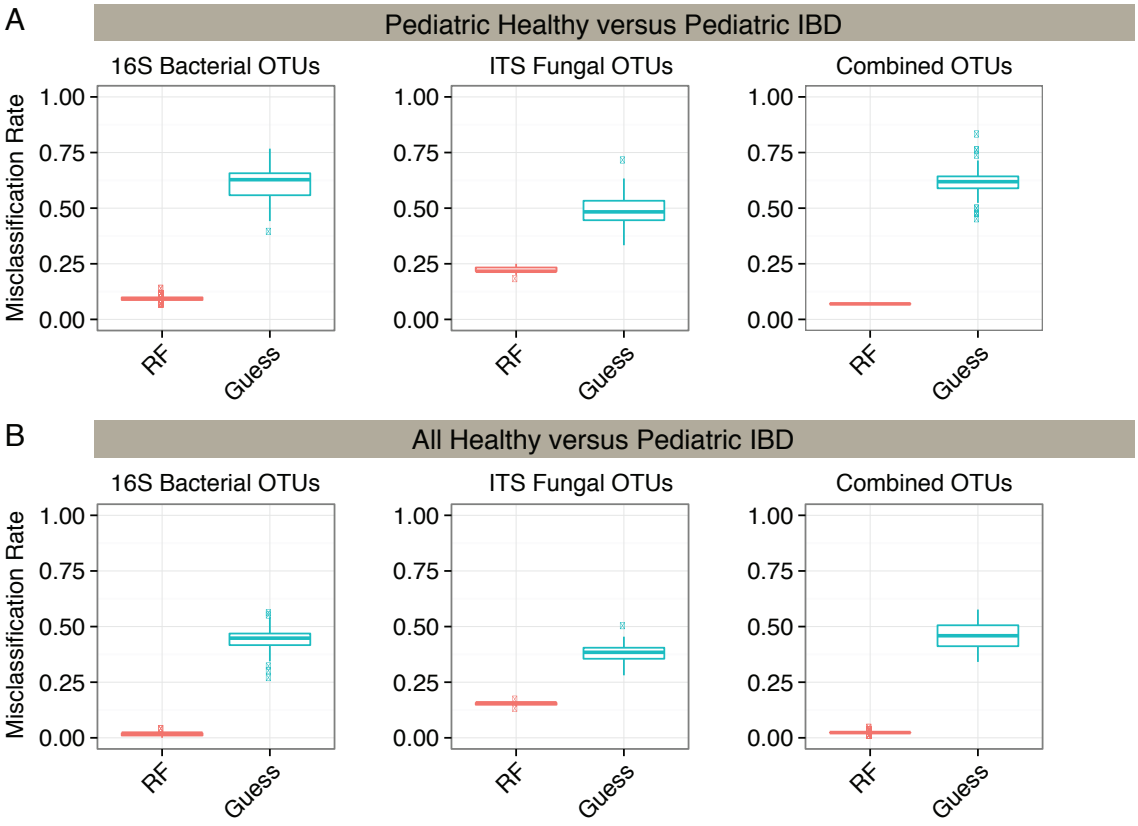
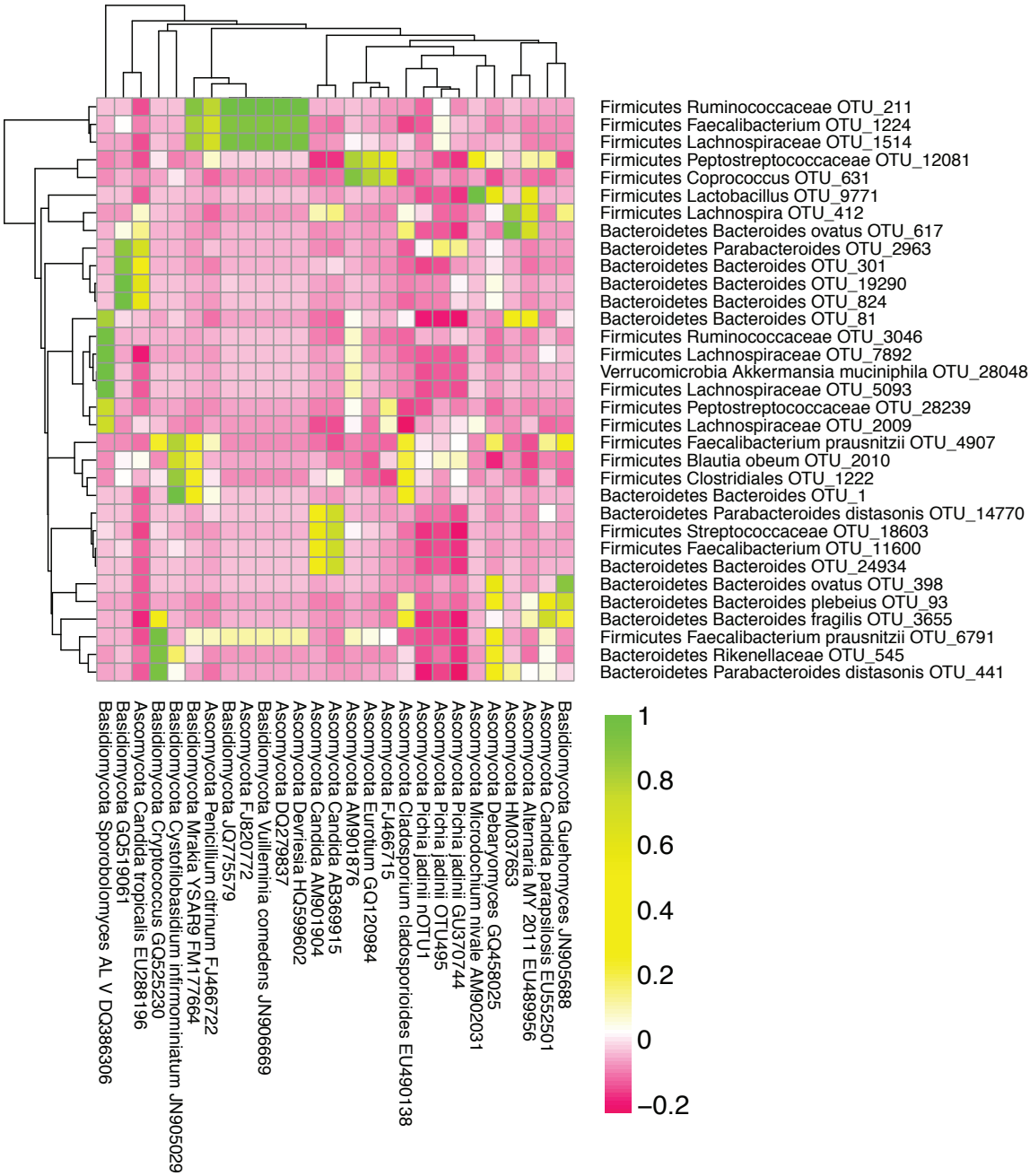


Figure 2.6. Correlations between Bacterial and Fungal OTUs in pediatric IBD. The Pearson correlation coefficient between fungal and bacterial OTUs in pediatric IBD patients was calculated. OTUs were included in this heatmap if they had greater than 100 sequences in pediatric IBD patients. OTUs were included if they significantly correlated with at least one other OTU (two-sided correlation, where the p-value exceeded 0.05 after Bonferroni correction).



References

1. Arumugam M, Raes J, Pelletier E, et al. Enterotypes of the human gut microbiome. Nature. 2011;473:174-180
2. Richard ML, Lamas B, Liguori G, et al. Gut Fungal Microbiota: The Yin and Yang of Inflammatory Bowel Disease. Inflamm Bowel Dis. 2014
3. Hansen EE, Lozupone CA, Rey FE, et al. Pan-genome of the dominant human gut-associated archaeon, Methanobrevibacter smithii, studied in twins. Proc Natl Acad Sci U S A. 2011;108 Suppl 1:4599-4606
4. Samuel BS, Hansen EE, Manchester JK, et al. Genomic and metabolic adaptations of Methanobrevibacter smithii to the human gut. Proc Natl Acad Sci U S A. 2007;104:10643-10648
5. Hoffmann C, Dollive S, Grunberg S, et al. Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. PLoS One. 2013;8:e66019
6. Huttenhower C, Kostic AD, Xavier RJ. Inflammatory bowel disease as a model for translating the microbiome. Immunity. 2014;40:843-854
7. Sartor RB. Therapeutic correction of bacterial dysbiosis discovered by molecular techniques. Proc Natl Acad Sci U S A. 2008;105:16413-16414
8. Walters WA, Xu Z, Knight R. Meta-analyses of human gut microbes associated with obesity and IBD. FEBS letters. 2014;588:4223-4233
9. Sartor RB. Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis. Nat Clin Pract Gastroenterol Hepatol. 2006;3:390-407
10. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. Nature. 2011;474:307-317
11. Ek WE, D'Amato M, Halfvarson J. The history of genetics in inflammatory bowel disease. Annals of gastroenterology : quarterly publication of the Hellenic Society of Gastroenterology. 2014;27:294-303
12. Vanderploeg R, Panaccione R, Ghosh S, et al. Influences of intestinal bacteria in human inflammatory bowel disease. Infect Dis Clin North Am. 2010;24:977-993, ix
13. Van de Merwe JP, Schroder AM, Wensinck F, et al. The obligate anaerobic faecal flora of patients with Crohn's disease and their first-degree relatives. Scand J Gastroenterol. 1988;23:1125-1131
14. Walker AW, Sanderson JD, Churcher C, et al. High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. BMC Microbiol. 2011;11:7
15. Manichanh C, Rigottier-Gois L, Bonnaud E, et al. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. Gut. 2006;55:205-211
16. Gophna U, Sommerfeld K, Gophna S, et al. Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis. J Clin Microbiol. 2006;44:4136-4141
17. Frank DN, St Amand AL, Feldman RA, et al. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci U S A. 2007;104:13780-13785
18. Martinez-Medina M, Aldeguer X, Gonzalez-Huix F, et al. 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. 2006;12:1136-1145
19. Prescott NJ, Fisher SA, Franke A, et al. A nonsynonymous SNP in ATG16L1 predisposes to ileal Crohn's disease and is independent of CARD15 and IBD5. Gastroenterology. 2007;132:1665-1671
20. Swidsinski A, Loening-Baucke V, Vaneechoutte M, et al. Active Crohn's disease and ulcerative colitis can be specifically diagnosed and monitored based on the biostructure of the fecal flora. Inflamm Bowel Dis. 2008;14:147-161
21. Seksik P, Rigottier-Gois L, Gramet G, et al. Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. Gut. 2003;52:237-242

22. Baumgart M, Dogan B, Rishniw M, et al. Culture independent analysis of ileal mucosa reveals a selective increase in invasive Escherichia coli of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. ISME J. 2007;1:403-418
23. Mangin I, Bonnet R, Seksik P, et al. Molecular inventory of faecal microflora in patients with Crohn's disease. FEMS Microbiol Ecol. 2004;50:25-36
24. Kohler JR, Casadevall A, Perfect J. The Spectrum of Fungi That Infects Humans. Cold Spring Harbor perspectives in medicine. 2014;5
25. Ott SJ, Kuhbacher T, Musfeldt M, et al. Fungi and inflammatory bowel diseases: Alterations of composition and diversity. Scand J Gastroenterol. 2008;43:831-841
26. Dollive S, Peterfreund GL, Sherrill-Mix S, et al. A tool kit for quantifying eukaryotic rRNA gene sequences from human microbiome samples. Genome biology. 2012;13:R60
27. Bittinger K, Charlson ES, Loy E, et al. Improved characterization of medically relevant fungi in the human respiratory tract using next-generation sequencing. Genome biology. 2014;15:487
28. Dollive S, Chen YY, Grunberg S, et al. Fungi of the Murine Gut: Episodic Variation and Proliferation during Antibiotic Treatment. PLoS One. 2013;8:e71806
29. Wang ZK, Yang YS, Stefka AT, et al. Review article: fungal microbiota and digestive diseases. Aliment Pharmacol Ther. 2014;39:751-766
30. Iliev ID, Funari VA, Taylor KD, et al. Interactions between commensal fungi and the C-type lectin receptor Dectin-1 influence colitis. Science. 2012;336:1314-1317
31. Barclay GR, McKenzie H, Pennington J, et al. The effect of dietary yeast on the activity of stable chronic Crohn's disease. Scand J Gastroenterol. 1992;27:196-200
32. Zwolinska-Wcislo M, Brzozowski T, Budak A, et al. Effect of Candida colonization on human ulcerative colitis and the healing of inflammatory changes of the colon in the experimental model of colitis ulcerosa. J Physiol Pharmacol. 2009;60:107-118
33. Prideaux L, De Cruz P, Ng SC, et al. Serological antibodies in inflammatory bowel disease: a systematic review. Inflamm Bowel Dis. 2012;18:1340-1355
34. Li Q, Wang C, Tang C, et al. Dysbiosis of gut fungal microbiota is associated with mucosal inflammation in Crohn's disease. Journal of clinical gastroenterology. 2014;48:513-523
35. Carbonero F, Benefiel AC, Gaskins HR. Contributions of the microbial hydrogen economy to colonic homeostasis. Nature reviews Gastroenterology & hepatology. 2012;9:504-518
36. Samuel BS, Gordon JI. A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism. Proc Natl Acad Sci U S A. 2006;103:10011-10016
37. Furusawa Y, Obata Y, Fukuda S, et al. Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. Nature. 2013
38. Hyams JS, Ferry GD, Mandel FS, et al. Development and validation of a pediatric Crohn's disease activity index. Journal of pediatric gastroenterology and nutrition. 1991;12:439-447
39. Turner D, Otley AR, Mack D, et al. Development, validation, and evaluation of a pediatric ulcerative colitis activity index: a prospective multicenter study. Gastroenterology. 2007;133:423-432
40. Wu GD, Chen J, Hoffmann C, et al. Linking long-term dietary patterns with gut microbial enterotypes. Science. 2011;334:105-108
41. McKenna P, Hoffmann C, Minkah N, et al. The macaque gut microbiome in health, lentiviral infection, and chronic enterocolitis. PLoS pathogens. 2008;4:e20
42. Hoffmann C, Minkah N, Leipzig J, et al. DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. Nucleic Acids Res. 2007;35:e91
43. Wu GD, Lewis JD, Hoffmann C, et al. Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. BMC Microbiol. 2010;10:206
44. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7:335-336
45. Cole JR, Wang Q, Cardenas E, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 2009;37:D141-145
46. Peterson DA, Frank DN, Pace NR, et al. Metagenomic approaches for defining the pathogenesis of inflammatory bowel diseases. Cell Host Microbe. 2008;3:417-427

47. Bellemain E, Carlsen T, Brochmann C, et al. ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. BMC Microbiol. 2010;10:189
48. Li J, Jia H, Cai X, et al. An integrated catalog of reference genes in the human gut microbiome. Nature biotechnology. 2014;32:834-841
49. Ruemmele FM, Seidman EG. Cytokine--intestinal epithelial cell interactions: implications for immune mediated bowel disorders. Zhonghua Min Guo Xiao Er Ke Yi Xue Hui Za Zhi. 1998;39:1-8
50. Quinton JF, Sendid B, Reumaux D, et al. Anti-Saccharomyces cerevisiae mannan antibodies combined with antineutrophil cytoplasmic autoantibodies in inflammatory bowel disease: prevalence and diagnostic role. Gut. 1998;42:788-791
51. Peeters M, Joossens S, Vermeire S, et al. Diagnostic value of anti-Saccharomyces cerevisiae and antineutrophil cytoplasmic autoantibodies in inflammatory bowel disease. Am J Gastroenterol. 2001;96:730-734
52. Peyrin-Biroulet L, Standaert-Vitse A, Branche J, et al. IBD serological panels: facts and perspectives. Inflamm Bowel Dis. 2007;13:1561-1566
53. Israeli E, Grotto I, Gilburd B, et al. Anti-Saccharomyces cerevisiae and antineutrophil cytoplasmic antibodies as predictors of inflammatory bowel disease. Gut. 2005;54:1232-1236
54. Standaert-Vitse A, Sendid B, Joossens M, et al. Candida albicans colonization and ASCA in familial Crohn's disease. Am J Gastroenterol. 2009;104:1745-1753
55. Standaert-Vitse A, Jouault T, Vandewalle P, et al. Candida albicans is an immunogen for anti-Saccharomyces cerevisiae antibody markers of Crohn's disease. Gastroenterology. 2006;130:1764-1775
56. Kumamoto CA. Inflammation and gastrointestinal Candida colonization. Curr Opin Microbiol. 2011;14:386-391
57. Hawksworth DL. A new dawn for the naming of fungi: impacts of decisions made in Melbourne in July 2011 on the future publication and regulation of fungal names. IMA fungus. 2011;2:155-162
58. Available at: http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=4903. Accessed 11/3/2014, 2014
59. Festring D, Hofmann T. Discovery of N(2)-(1-carboxyethyl)guanosine 5'-monophosphate as an umami-enhancing maillard-modified nucleotide in yeast extracts. J Agric Food Chem. 2010;58:10614-10622
60. Moses A, Maayan S, Shvil Y, et al. Hansenula anomala infections in children: from asymptomatic colonization to tissue invasion. Pediatr Infect Dis J. 1991;10:400-402
61. Suchodolski JS, Morris EK, Allenspach K, et al. Prevalence and identification of fungal DNA in the small intestine of healthy dogs and dogs with chronic enteropathies. Vet Microbiol. 2008;132:379-388
62. Pettoello-Mantovani M, Nocerino A, Polonelli L, et al. Hansenula anomala killer toxin induces secretion and severe acute injury in the rat intestine. Gastroenterology. 1995;109:1900-1906
63. Dave M, Purohit T, Razonable R, et al. Opportunistic infections due to inflammatory bowel disease therapy. Inflamm Bowel Dis. 2014;20:196-212
64. Sanders ME, Guarner F, Guerrant R, et al. An update on the use and investigation of probiotics in health and disease. Gut. 2013;62:787-796
65. Eriksson M, Johannssen T, von Smolinski D, et al. The C-Type Lectin Receptor SIGNR3 Binds to Fungi Present in Commensal Microbiota and Influences Immune Regulation in Experimental Colitis. Front Immunol. 2013;4:196
66. Lewis JD. The utility of biomarkers in the diagnosis and therapy of inflammatory bowel disease. Gastroenterology. 2011;140:1817-1826 e1812
67. Levine J, Ellis CJ, Furne JK, et al. Fecal hydrogen sulfide production in ulcerative colitis. Am J Gastroenterol. 1998;93:83-87
68. Pimentel M, Mayer AG, Park S, et al. Methane production during lactulose breath test is associated with gastrointestinal disease presentation. Digestive diseases and sciences. 2003;48:86-92

69. Rana SV, Sharma S, Malik A, et al. Small intestinal bacterial overgrowth and orocecal transit time in patients of inflammatory bowel disease. Digestive diseases and sciences. 2013;58:2594-2598
70. Scanlan PD, Shanahan F, Marchesi JR. Human methanogen diversity and incidence in healthy and diseased colonic groups using mcrA gene analysis. BMC Microbiol. 2008;8:79
71. Eadala P, Matthews SB, Waud JP, et al. Association of lactose sensitivity with inflammatory bowel disease--demonstrated by analysis of genetic polymorphism, breath gases and symptoms. Aliment Pharmacol Ther. 2011;34:735-746
72. Peled Y, Gilat T, Liberman E, et al. The development of methane production in childhood and adolescence. Journal of pediatric gastroenterology and nutrition. 1985;4:575-579

Acknowledgements

This work was supported by Project UH2DK083981; R01 GM103591; the Penn Center for Molecular Studies in Digestive and Liver Disease including the Molecular Biology/Gene Expression Core (P30 DK050306); The Joint Penn-CHOP Center for Digestive, Liver, and Pancreatic Medicine; S10RR024525; UL1RR024134, and K24-DK078228; and the University of Pennsylvania Center for AIDS Research (CFAR) P30 AI 045008. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center for Research Resources, National Institutes of Health, or Pennsylvania Department of Health. Sequences were submitted to the Sequence Read Archive with accession numbers SRP050217.

Contribution

I performed the analysis and built all of the figures. I worked with LA, GW, FB to generate the analysis strategy and author the manuscript.

Supplemental Information

Supplemental Methods

Sequence analysis

Sequence data was processed using QIIME [1]. Sequences were quality trimmed and assigned to their respective sample based on their barcodes. Sequences were binned into de novo Operational Taxonomic Units (OTUs) using CDHIT [2]. with a 97% minimum sequence

identity threshold for 16S and a 95.2% minimum sequence identity threshold for ITS. The most abundant sequence from each OTU was selected as the representative sequence for that OTU.

For the ITS analysis, spacer sequences of lengths 130 to 1000 nt were used for the analysis. In a previous study [3], a longer minimum was used, which resulted in loss of the unusually short Cyberlindnera jadinii (aka Pichia jadinii and Candida utilis), which is only 142 nt in length. We also note that length variation in the ITS region may have resulted in under-estimation of the abundance of Saccharomyces cerevisiae, which has a relatively long ITS region (368 nt).

Taxonomy was assigned to the representative sequences using Ribosomal Database Project (RDP) for 16S [4] and the UNITE database for the ITS. The bacterial sequences were then NAST aligned using the Greengenes' reference database [5] and used to build a phylogenetic tree using the using the FastTree algorithm [6]. Bacterial community distances were calculated between all pair of samples using Unifrac [7]. UniFrac distances are based on the fraction of branch length shared in a phylogenetic tree between two samples' microbial communities. Weighted UniFrac incorporate the relative abundances of each OTU [8]. For the fungal community distances, Jaccard and abundance-weighted Jaccard indices were calculated.

Principal Coordinate Analysis (PCoA) based on UniFrac distances was used to compare samples. Heatmaps that show the taxonomic distribution of each sample's sequences were created using Qiimer (http://cran.r-project.org/web/packages/qiimer). Diversity was assessed using the Shannon diversity metric. LEFSE [9] was used to identify taxa that differed between IBD and healthy samples.

Supplemental Methods References

1. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;**7**:335-6.
2. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 2006;**22**:1658-9.
3. Hoffmann C, Dollive S, Grunberg S, et al. Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. PLoS One. 2013;8:e66019

4. Cole JR, Wang Q, Cardenas E, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 2009;**37**:D141-5.
5. DeSantis TZ, Jr., Hugenholtz P, Keller K, et al. NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic Acids Res 2006;**34**:W394-9.
6. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 2009;**26**:1641-50.
7. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 2005;**71**:8228-35.
8. Lozupone CA, Hamady M, Kelley ST, et al. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Appl Environ Microbiol 2007;**73**:1576-85.

Supplemental Figures and Tables

Supplemental figures and tables are available online: *Chehoud, Albenberg* et al, IBD 2015.

CHAPTER 3: Blooms of Anelloviruses in the Respiratory Tract of Lung Transplant Recipients

The contents of this chapter have been published as:

> Young J, **Chehoud C**, Bittinger K, Bailey A, Diamond J, Cantu E, Haas A, Abbas A, Frye L, Christie J, Bushman F, Collman R. 2014. *Blooms of Anelloviruses in the Respiratory Tract of Lung Transplant Recipients.* American Journal of Transplantation (PMID: 25403800).

Abstract

Few studies have examined the lung virome in health and disease. Outcomes of lung transplantation are known to be influenced by several recognized respiratory viruses, but global understanding of the virome of the transplanted lung is incomplete. To define the DNA virome within the respiratory tract following lung transplantation we carried out metagenomic analysis of allograft bronchoalveolar lavage (BAL), and compared to healthy and HIV+ subjects. Viral concentrates were purified from BAL and analyzed by shotgun DNA sequencing. All of the BAL samples contained reads mapping to anelloviruses, with high proportions in lung transplant samples. Anellovirus populations in transplant recipients were complex, with multiple concurrent variants. Q-PCR quantification revealed that anellovirus sequences were 56-fold more abundant in BAL from lung transplant recipients compared with healthy controls or HIV+ subjects ($p<0.0001$). Anellovirus sequences were also more abundant in upper respiratory tract specimens from lung transplant recipients than controls ($p=0.006$). Comparison to metagenomic data on bacterial populations showed that high anellovirus loads correlated with dysbiotic bacterial communities in allograft BAL ($p=0.00816$). Thus the respiratory tracts of lung transplant recipients contain high levels and complex populations of anelloviruses, warranting studies of anellovirus lung infection and transplant outcome.

Introduction

Little is known about the virome of the human respiratory tract as a whole, though infections by individual viruses are well characterized. For the case of lung transplantation, viral infection is a major complicating factor impacting graft survival rates (1-4). Respiratory infections with known viruses can cause direct lung injury or increase risk of graft failure, as in the case of cytomegalovirus and community acquired respiratory viruses (1, 5). Intense interest has thus focused on viruses in the respiratory tract and transplantation outcome.

Today it is possible to characterize large viral populations using high throughput metagenomic sequencing (6-8), which has identified both well-recognized and little-studied viruses living in association with humans. Only a few studies have applied metagenomic approaches to understand viruses of the lower respiratory tract (8, 9), and none in lung transplantation.

Anelloviruses are circular, nonenveloped, negative-sense, single-stranded DNA viruses that commonly colonize humans and show increased abundance in blood after hematopoietic and solid organ transplantation (10-12). The anellovirus family consists of Torque Teno viruses (TTVs), Torque Teno Midi Viruses (TTMDV), Torque Teno Mini Viruses (TTMV), and Small Anelloviruses (SAVs), each of which has multiple subtypes (12, 13). Their small genomes (2.3-3.8 kb) consist of three to four open reading frames and a highly conserved untranslated region (UTR) (12)_ENREF_8. These viruses are ubiquitous in the human population and have not yet been causally linked with any human disease (12, 13). Diverse types of anelloviruses have been found in various organs, tissues, and cell types (12, 14-16).

In the respiratory tract, TTV was recently identified in bronchoalveolar lavage fluid from 28% of individuals with acute exacerbations of idiopathic pulmonary fibrosis (IPF), but not those with stable IPF, and in a quarter of individuals with acute lung injury (17). In the upper respiratory tract, elevated levels of TTV have been found in nasal secretions from children with respiratory diseases, and correlated with disease severity (15, 16). In HIV-infected patients, plasma

concentrations of anelloviruses increased during progression to AIDS (18) and decreased following therapy (19). TTV viremia was reported to increase following autologous hematopoietic stem cell transplantation, and then return to baseline levels following immune reconstitution (10). Recently, TTV levels in blood were shown to increase in association with immunosuppression following lung and heart transplantation (11). The association between TTV levels and immune status has led some authors to propose that anelloviruses genome copy numbers may serve as an empirical measure of successful immune suppression (10, 11, 19).

We report here the first study to apply viral metagenomics to understand the lung virome in lung transplantation. We first used Illumina metagenomic sequencing to characterize lung DNA viromes from lung transplant recipients, along with another immunologically impaired group, HIV-positive individuals. This showed that anellovirus sequences were prominent in BAL from transplant recipients, and revealed the presence of complex populations with multiple concurrent variants. Based on these metagenomic data, we then quantified anellovirus levels within the lungs and upper respiratory tracts of lung transplant recipients, HIV+ subjects, and healthy individuals using Q-PCR, demonstrating high levels of anellovirus DNA in BAL and OW of lung transplant recipients. Our findings demonstrate metagenomic detection and genetic characterization of the allograft virome, which is then followed by broader quantitative analysis. We also compared metagenomic data on bacterial populations, and found that high levels of anelloviruses correlated with dysbiotic bacterial communities, revealing covariation among microbiome constituents.

Results

BAL and OW samples were collected from 1) lung transplant recipients undergoing post-transplant bronchoscopy, 2) HIV+ individuals without lung disease or respiratory symptoms, and 3) healthy control subjects (20, 21) (Sup. Tables 3.1, 3.2, and 3.3). VLPs were purified from BAL and OW, followed by concentration and treatment with nucleases to eliminate non-encapsidated nucleic acids (Figure 3.1). DNA was then purified from each sample. To verify selective reduction

in human and bacterial cells, two BAL samples were analyzed using Q-PCR to quantify bacterial 16S rRNA gene copies or human beta-tubulin gene copies. Copies of bacterial 16S DNA decreased from 3000 copies/ng in unfractionated BAL to 80 copies/ng after viral particle purification (38-fold reduction). Human beta-tubulin DNA was detected at 925 copies/ng in unfractionated BAL but became undetectable after purification (>900-fold reduction).

Metagenomic sequencing of lung transplant and HIV+ samples

We first applied metagenomic analysis to viral DNA from six lung transplant and three HIV+ BAL samples. DNA from viral preps was subjected to whole genome amplification and then shotgun sequenced using the Illumina MiSeq platform (2 x 250 bp reads) (Table 3.1). Up to 3.4 million reads per sample were generated, for a total of 5.42×10^9 bp of sequence data. Reads were filtered and trimmed to remove low quality sequences, and human reads were identified and removed. Reads were then aligned using BLAST to the NCBI viral database (Table 3.1 and Sup. Table 3.4).

Of reads assigned to reference viruses, the majority (>68%) mapped to anelloviruses, which included Torque Teno viruses, Torque Teno Mini viruses, Torque Teno Midi viruses, and Small Anelloviruses (Table 3.1, Figure 3.2A). Based on the closest match to reference viruses, a wide variety of anelloviruses were present within the lungs of lung transplant subjects, while substantially fewer anellovirus reads were detected in lungs of HIV+ individuals (Figure 3.2A). Furthermore, diverse anelloviruses were present in single individuals, especially in the lung transplant recipients, as evidenced by reads aligning to many different TTVs, TTMVs, TTMVDs, and SAVs within single samples (Table 3.1).

Other eukaryotic viruses were detected within the samples, but at much lower levels. These included *Epstein-Barr virus (Human Herpesvirus 4*) (2 reads in one sample), *Human Herpesvirus 7* (2 reads in one sample, 36 reads in another sample), and *Human Papillomavirus* (1 read) (Sup. Table 3.4). To determine whether use of the 0.22 micron filter step during virus

purification may have resulted in depletion of some of the larger viruses, we compared filtration and centrifugation (Sup. Table 3.6). This showed that filtration resulted in reduced recovery of herpesviruses, but was more efficient at capturing small viruses such as human papillomaviruses and anelloviruses.

A small percentage of reads (0.75% of total reads for all samples combined) aligned to bacteriophage genomes, predominantly phages of *Enterobacteria*, *Salmonella, Pseudomonas*, *Streptococcus* and *Yersinia*. When bacterophages were compared with bacterial 16S sequences from these subjects that were previously analyzed (31), we found representation of several of these bacterial lineages (*Streptococcus*, *Pseudomonas*, *Staphylococcus*, and *Enterobacteria*), consistent with possible phage/host pairs in our samples.

An average of 81% of the reads did not match reference viruses in the NCBI viral database, a high proportion similar to results of other viral metagenomic studies (9, 32). We surmise that many of these may correspond to DNA phage sequences, because phage are hyperabundant globally and poorly represented in sequence databases (33). Unidentified reads could also correspond to novel mammalian viruses or low quality reads not removed by our filtering strategy.

Genetic structure of lung anellovirus populations

Anellovirus reads were assembled into contigs for more complete analysis of anelloviruses community structure, using iterative deBruijn graph assembly (23). To validate correct contig assembly, we designed specific primer sets for three of the contigs, then PCR amplified 500-700 base pair regions, sequenced using the Sanger method, and assembled using the conventional overlap method. Assembled contigs were then aligned to the NCBI nt database (Figure 3.2B). Selected regions within the anellovirus contigs were 87-93% identical to the reference sequences, though highly divergent regions were also detected. Anellovirus open reading frames (ORF) 1 and 2 were identified in the three anellovirus contigs. ORF1 was

previously shown to encode the capsid gene and contains a hypervariable region, which may evolve rapidly to evade the host immune response (12). The ORF2 encoded protein was reported to suppress the NF-kB pathway, thereby potentially regulating the host innate and adaptive immune response (12, 34).

To compare anelloviruses populations within and between individuals, a phylogenetic tree was generated using ORF1 sequences from reference genomes and contigs assembled from the Illumina and Sanger sequence data. As shown in Figure 3.3, multiple different anellovirus ORF1s were found within individual lung transplant patient allografts, with subjects exhibiting as many as 17 distinct contigs. In contrast, richness of anelloviruses was lower in BAL of HIV+ individuals, even though sequences were detected in all of them. Anellovirus contigs were mostly different between individuals (Figure 3.3).

Quantification of Anellovirus DNA in lung and upper respiratory tract samples using quantitative PCR

Given diverse anellovirus populations found in some of the transplant samples, we sought to quantify absolute anellovirus DNA levels in a larger group of subjects using Q-PCR. We used the universal PCR primers targeting the anellovirus highly conserved UTR (22) to generate a Q-PCR assay. These primers recognize TTVs, TTMVs, TTMVDs, and SAVs. We analyzed a total of 53 BAL samples from lung transplant recipients (n=38), HIV+ subjects (n=3), and healthy controls (n=12). To assess the upper respiratory tract, we also analyzed 24 OW samples from lung transplant recipients (n=12) and healthy controls (n=12).

As shown in Figure 3.4, anellovirus copy numbers in lung ranged widely among transplant recipients, with a median of 15,133 copies/ml (range=314 to $2.3x10^7$ copies/ml). These levels were far higher than in lungs of healthy or HIV+ subjects (medians of 271 and 191 copies/ml, respectively; $p<0.0002$ for comparison to lung transplant samples; Mann-Whitney test).

Quantities of anelloviruses in the upper respiratory tract were also measured in OW collected from lung transplant recipients and healthy subjects. As was seen in lung, anellovirus genome copies were significantly more abundant in the upper respiratory tract of transplant recipients compared with healthy controls (Figure 3.5) (median=16,700 in transplant recipients, range=634 to 632,759 copies/ml compared to median=1,033 in healthy; p=0.0055). When quantities of anellovirus were compared between lung and upper respiratory tract within the same treatment group, healthy subjects had significantly lower titers in their BAL compared to oropharyngeal wash (p=0.0068, Mann-Whitney test). However, in lung transplant recipients, anellovirus genome copies were higher in both sample types and not significantly different in lung compared to the upper respiratory tract (p=0.23, Figure 3.6).

Comparison of anellovirus DNA levels to transplant subject metadata

We then asked whether anellovirus levels in the transplant recipients' lungs were correlated with clinical variables that might explain the high variability. We queried patient age, gender, time since transplant, number of immunosuppressive drugs, tacrolimus levels, prednisone dosing, target immunosuppression range, surgical type (single vs. bilateral lung transplant), use of azithromycin, cytomegalovirus status (donor and recipient), BAL bacterial culture results and bacterial load in BAL based on 16S rRNA gene Q-PCR, acute rejection grade, and pre-transplant diagnosis (suppurative vs. nonsuppurative lung disease). No significant associations were detected (Sup. Table 3.5). Though power was limited by the relatively small sample sizes, these data suggest that anellovirus replication in the lung allograft is linked to factors other than these clinical variables.

Comparison of lung anellovirus DNA levels to the bacterial microbiome

Finally, we asked whether there was a relationship between anellovirus DNA levels and composition of the bacterial microbiome in allografts. Previously, DNA was purified from whole

BAL, PCR ampified using primers that recognize conserved regions of the bacterial 16S rRNA gene, and products analyzed by deep sequencing (20). That analysis showed that lung transplant bacterial communities differed in composition from healthy subjects, and transplant recipients were commonly colonized by oral bacteria, known pathogenic bacteria, and sometimes by unexpected bacterial lineages. Absolute levels of bacteria were also higher than in healthy controls (40-100-fold). Here we compared the anellovirus titers with bacterial community composition among transplant recipients, and found a significant correlation (Figure 3.7; weighted UniFrac, p=0.032). However, no specific bacterial lineages were significantly linked to anellovirus levels among transplant subjects. We then asked if the relationship between anellovirus and 16S bacterial composition might be linked to the degree to which the bacterial communities diverged from healthy controls. This analysis revealed a significant correlation between transplant BAL anellovirus levels and a metric measuring the divergence between the corresponding bacterial community and those of healthy subjects (Sup. Figure 3.1; p=0.008). Thus, high anellovirus levels in transplant recipients are also associated with aberrant bacterial microbiota.

Discussion

Here we used metagenomic sequencing to analyze DNA viruses present in the respiratory tract of lung transplant recipients and HIV-infected individuals. The metagenomic data then enabled targeted analysis of anellovirus levels in a larger cohort of lung transplant and control subjects. This is the first study to apply viral metagenomics to the lung transplant allograft. Multiple viruses infecting animal cells were detected by sequencing, including a remarkable abundance and complexity of anelloviruses, as well as much lower levels of papilloma viruses and herpes viruses. Bacteriophage were also detectable, as were unidentified sequences likely corresponding at least in part to phages. Our most notable finding was that the anelloviruses, including TTVs, TTMVDs, TTMVs, and SAVs, were greatly increased in abundance in the lungs and the upper respiratory tract of lung transplant recipients compared with healthy and HIV+ subjects. As this work was being completed, Quake and colleagues reported that anellovirus sequences were increased in abundance in blood in organ transplant recipients (11). Our data

reported here show that after lung transplantation, anellovirus sequences are abundant in the lung allograft itself.

The shotgun metagenomic sequence data enabled assembly of near-complete genome sequences for some of the most abundant anelloviruses, allowing the genetic nature of anelloviruses present to be analyzed in detail. A notable feature of transplant recipients was the great diversity of anelloviruses represented within single individuals. Quantifying the exact number of different forms is challenging and dependent on the metric chosen. Taking each independent contig into account, there were up to 247 contigs identified as anelloviruses within one sample (Tx-24). However, this may be an overestimate because some contigs are only partial genomes (sizes ranged from 250-3600 base pairs), though this is hard to assess definitively because replication-competent subviral molecules have been reported (35). Analysis based on viral ORF1 regions showed that up to 17 different anelloviruses were present within a single transplant recipient (Tx-24), indicating remarkable diversity. Some of the anelloviruses identified in these subjects were distant in sequence from database genomes. Such high divergence was confirmed in the three genomes reconstructed by Sanger sequencing, and so is not a result of error in Illumina sequence determination or assembly. The origin of this extreme diversity is unclear, but may be due in part to the known high mutation rates of single-stranded DNA viruses (36) combined with high virus levels likely reflecting high rates of virus replication in lung.

We found large numbers of bacteriophage sequences in our shotgun metagenomic data, likely due to the abundant bacterial populations in lung transplant subjects (20), but relatively few mammalian DNA viruses besides anelloviruses. This is likely in large part related to the near-universal use of herpesviruses prophylaxis in transplant subjects, and relatively modest level of immune dysfunction based on CD4+ T cell counts in our HIV+ subjects. In addition, samples for metagenomic analysis were subject to multiple displacement amplification, which favors small circular genomes such as anelloviruses. An additional technical factor may be that the filtration

step used could have depleted large viruses or viral particles bound to other macromolecules, although reconstruction experiments suggest that the losses were modest. In addition, a large proportion of sequences did not match any database genomes, which is concordant with findings in other metagenomic virome studies (6, 9, 32). While many of these are likely novel bacterophages, it is possible that novel mammalian viruses may exist within those sequences as well. One limitation of our sample set is the relatively short elapsed time since transplant (median of 5 months), and it will be useful to investigate whether additional viral lineages emerge over longer times.

Anellovirus genome copy numbers in lung from Q-PCR data were remarkably different among transplant subjects, varying across nearly 5 orders of magnitude. However, comparison to standard patient metadata did not show significant correlations with any of the clinical parameters queried. A recent study of anellovirus in serum following organ transplantation reported that anellovirus genome levels correlated with the extent of immunosuppression, suggesting that anellovirus DNA in blood might serve as an assay for the overall level of immunosuppression (11). In contrast, our subjects were receiving relatively homogenous immunosuppression regimens (Table S1), and we found no association with any measure of immunosuppression, nor with time since transplant (Table S5). Thus, the absence of correlations seen here between lung anellovirus levels and clinical variables suggests that anellovirus replication in lung allografts may be regulated by factors that have yet to be identified.

We compared anellovirus DNA quantification to previously determined data on the structure of bacterial communities, and identified a novel form of covariation. Among transplant recipients, BAL samples with high anellovirus levels differed from those with lower levels based on the bacterial 16S data. Furthermore, samples with highest levels of anellovirus were more divergent from the bacterial communities of healthy controls, indicating bacterial dysbiosis. This suggests that the anellovirus and bacterial communities may both be blooming in response to some common property of the host, such as loss of immune control. If so, the immune impairment

would not be reflected by standard clinical measures, which did not correlate with features of anelloviral DNA levels. The correlation between anellovirus titers and bacterial dysbiosis was imperfect, however, indicating that there are likely both shared and unique factors influencing viral and bacterial microbiota, respectively.

In summary, our results employing shotgun metagenomic sequencing revealed robust anellovirus populations in lungs of lung transplant recipients, which was then confirmed and quantified by targeted Q-PCR. Long-term lung transplant outcomes have been linked to prior infection with several other viruses (1, 2, 4). While our subjects were sampled at relatively early time points (median 5 months post-transplant), early post-transplant colonization with other microbial agents has been associated with BOS at later time points (37, 38). Thus, the finding of unexpected high levels of anellovirus replication within lung allografts and marked inter-subject variability suggest that longitudinal studies are warranted to determine whether levels of anelloviruses in the lung allograft are associated with, and may play a role in, transplant outcomes.

Methods

Sample collection

Bronchoalveolar lavage (BAL) was carried out within the lung allograft on lung transplant recipients, most of whom were undergoing routine clinical surveillance bronchoscopy during the first year post-transplant (Sup. Table 3.1) as described previously (20). Oropharyngeal wash (OW) to sample the upper respiratory tract was collected as previously described (20). One group of control samples were obtained from healthy volunteers who underwent research bronchoscopy using the same procedure (Sup. Table 3.2). Samples from three HIV+ subjects not on antiretroviral therapy (CD4 T cell counts of 301, 321 and 682; Sup. Table 3.3) and a second set from healthy volunteers (Table S2) were obtained by bronchoscopy using a previously described two-scope procedure (21). The University of Pennsylvania IRB approved all procedures (protocols # 812748 and #810851), and subjects gave written informed consent.

Virus-Like Particle (VLP) purification

Virus-like particles were purified from 1-5 milliliters of BAL or OW, depending upon availability. 10mM MgSO4 and 10mM dithiothreitol were added to the BAL or OW and filtered through a 0.22 μm filter (Millipore). The filtrate was concentrated using a 100kDa molecular weight cut-off filter (Amicon 20, Millipore), resuspended in 1 ml Buffer SM_ENREF_21, and reconcentrated. The concentrate was treated with DNase I and RNase (Roche) at 37°C for 15 min to eliminate non-encapsidated nucleic acids, then the enzymes were deactivated at 70°C for 5 min.

Nucleic acid extraction and metagenomic sequencing

Nucleic acids were extracted from virus particle preps using Qiagen's All Prep nucleic acid isolation kit. Six lung transplant and three HIV+ BAL samples that had large volumes of BAL and detailed patient metadata available were used for metagenomic sequencing. Whole genome amplification was performed on these samples using the GenomiPhi V2 Amplification Kit (GE Healthcare). Libraries for sequencing were made using Illumina's Nextera XT DNA Sample Preparation Kit with 1 ng of input DNA, generating paired-end fragments. Metagenomic sequencing was performed on an Illumina MiSeq instrument.

Quantitative PCR

Q-PCR was carried out on non-Genomiphi amplified DNA using primers targeting the UTR of anelloviruses from a previously-described assay (22) that we adapted to a SYBR green-based method as described in Supplemental Methods. All samples were run in duplicate or triplicate and values averaged. The detection limit was 1.4 copies per reaction.

Bioinformatics pipeline

Paired-end reads from the MiSeq instrument were quality-trimmed and processed through BMTagger to remove human sequences. Non-human reads were then analyzed using

BLAST against the NCBI viral database. Reads were assembled into contigs by iterative deBruijn graph assembly using IDBA-UD (23). To verify contig assembly, contig-specific PCR primers were synthesized and amplification products subjected to Sanger sequencing. Dot plots were generated using Gepard with a word size of 10 nucleotides (24). Anellovirus ORF1 amino acid sequences were identified from our contigs and aligned along with 49 reference TTV, TTMDV, and TTMV sequences from Genbank. The phylogenetic tree was built using FastTree v2.1.3 (25). Further details of bioinformatic analysis are provided in Supplemental Methods.

Analysis of 16S reads from (20) was undertaken using QIIME v1.8 (26). Operational taxonomic units were selected at 97% sequence similarity with UCLUST (27). Following this step, the seed sequences from the clustering algorithm were used as representatives for each cluster. Taxonomic assignments against the GreenGenes 16S database v13_8 (28) were generated using the default consensus method in QIIME. A phylogenetic tree was estimated by FastTree v2.1.3 (25) using PyNAST-aligned representative sequences (29). Pairwise UniFrac distances were then calculated using the implementation in QIIME.

Statistical analyses

Mann-Whitney tests, paired Wilcoxon signed rank tests, t-tests, and Spearman correlation tests were performed in R (30).

Table and Figures

Table 3.1. Anelloviruses identified through metagenomic sequencing of BAL samples from lung transplant recipients and HIV+ individuals.

	HIV+			Lung Transplant Recipients					
	HUP1A03	**HUP1B03**	**HUP1B07**	**Tx-24**	**Tx-34**	**Tx-38**	**Tx-49**	**Tx-51**	**Tx-52**
Total Reads (Unfiltered)	1,134,270	3,480,764	649,864	644,536	869,518	1,236,626	1,169,756	171,892	71,018
Total Reads after Quality Trimming	1,134,256	3,480,750	649,852	644,526	869,504	1,236,620	1,169,742	171,884	71,010
% Human Reads	0%	0%	0%	1%	0%	26%	29%	21%	14%
% Reads Matching to Viral DB	4%	4%	5%	21%	23%	1%	10%	7%	1%
Small anellovirus 1	0	0	0	1	2	0	0	0	0
Small anellovirus 2	0	0	0	14	0	0	0	0	0
Torque teno felis virus	0	0	0	0	0	1	0	0	0
Torque teno midi virus 1	0	0	0	20	1	0	0	0	0
Torque teno midi virus 2	0	0	0	13	1	0	0	0	0
Torque teno mini virus 2	0	0	0	0	4	0	0	0	0
Torque teno mini virus 3	0	0	0	0	0	5	2	0	0
Torque teno mini virus 4	0	0	0	0	6	0	0	0	0
Torque teno mini virus 5	0	0	0	0	0	7	0	0	18
Torque teno mini virus 7	0	0	0	5	0	9	0	0	6
Torque teno mini virus 8	0	0	0	0	3	2	0	0	0
Torque teno virus 1	4	6	3	3124	10901	44	1051	128	2
Torque teno virus 10	8	10	3	6984	10537	97	4828	513	1
Torque teno virus 12	4	13	0	9946	222	21	2695	98	4
Torque teno virus 14	1	4	0	480	15	290	10819	4235	0
Torque teno virus 15	19	46	19	44356	75859	115	7173	564	3
Torque teno virus 16	7	18	5	10860	5918	166	19424	175	0
Torque teno virus 19	7	12	5	3102	12143	78	14768	233	2
Torque teno virus 2	1	6	0	262	7	82	66	501	0
Torque teno virus 27	2	0	1	1381	1046	46	1036	57	1
Torque teno virus 28	3	3	1	211	80	11	161	6	0
Torque teno virus 3	5	33	1	15693	59	79	19995	1202	0
Torque teno virus 4	2	1	0	853	9	0	1114	661	0
Torque teno virus 6	19	24	5	22153	13397	184	10487	127	6
Torque teno virus 7	4	48	2	8714	28	20	1950	22	0
Torque teno virus 8	20	28	79	6834	69854	868	17068	2440	24
Torque teno virus	0	1	0	1076	3	24	970	0	0

Figure 3.1. Overview of experimental design. BAL and OW samples (1-5 ml) were filtered through a 0.22 μM filter, concentrated on an Amicon 10 kDa MWCO filter, washed, and treated with DNase and RNase to yield purified virus-like particles. Select samples were checked for viral purity by quantifying the bacterial 16S rRNA and human β-tubulin genes using Q-PCR. Anelloviruses were quantified in all samples by Q-PCR. Select lung transplant and HIV+ samples were whole-genome amplified by multiple displacement amplification (MDA) and shotgun sequenced using the Illumina MiSeq platform. Bioinformatic analysis of the virome reads consisted of removing human reads (BMTagger) and aligning to the NCBI viral database. Reads were assembled into contigs, aligned to the NCBI viral database and analyzed.

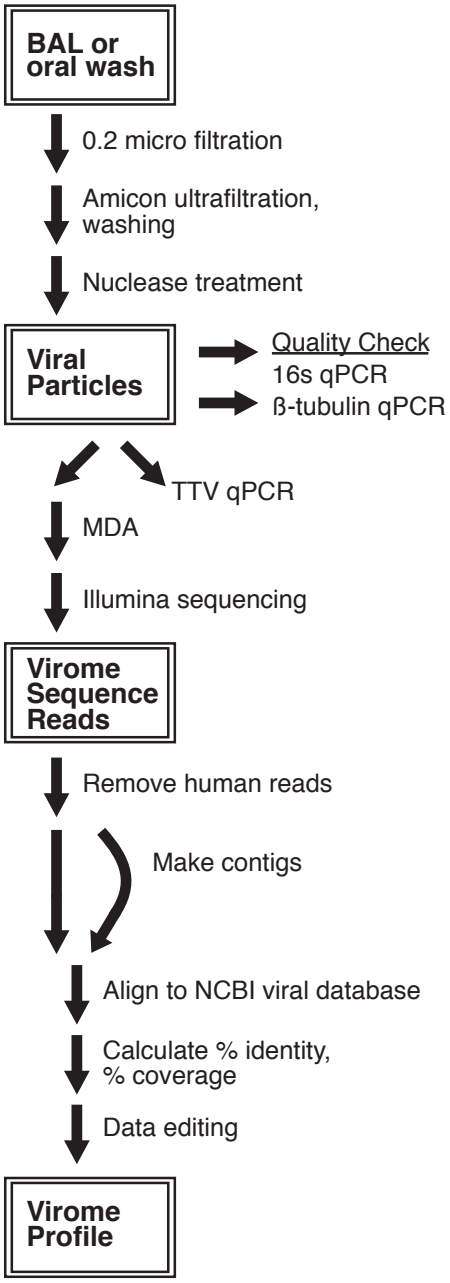


Figure 3.2. Anelloviruses in BAL of HIV+ individuals and lung transplant recipients. A) Distribution of reads aligning to anelloviruses. Metagenomic sequencing reads were searched against the NCBI viral database and the number of combined reads aligning to each anellovirus calculated for lung transplant recipients (n=6) and HIV+ subjects (n=3). B) Presence of multiple anelloviruses present within the lungs of lung transplant recipient Tx-24 verified by Sanger sequencing. For each of the three analyzed by Sanger sequencing (x-axis), similarity was scored versus their closest matching NCBI reference sequence (y-axis). Open reading frames were assigned using MacVector and are illustrated on the genetic maps under the dot plots.

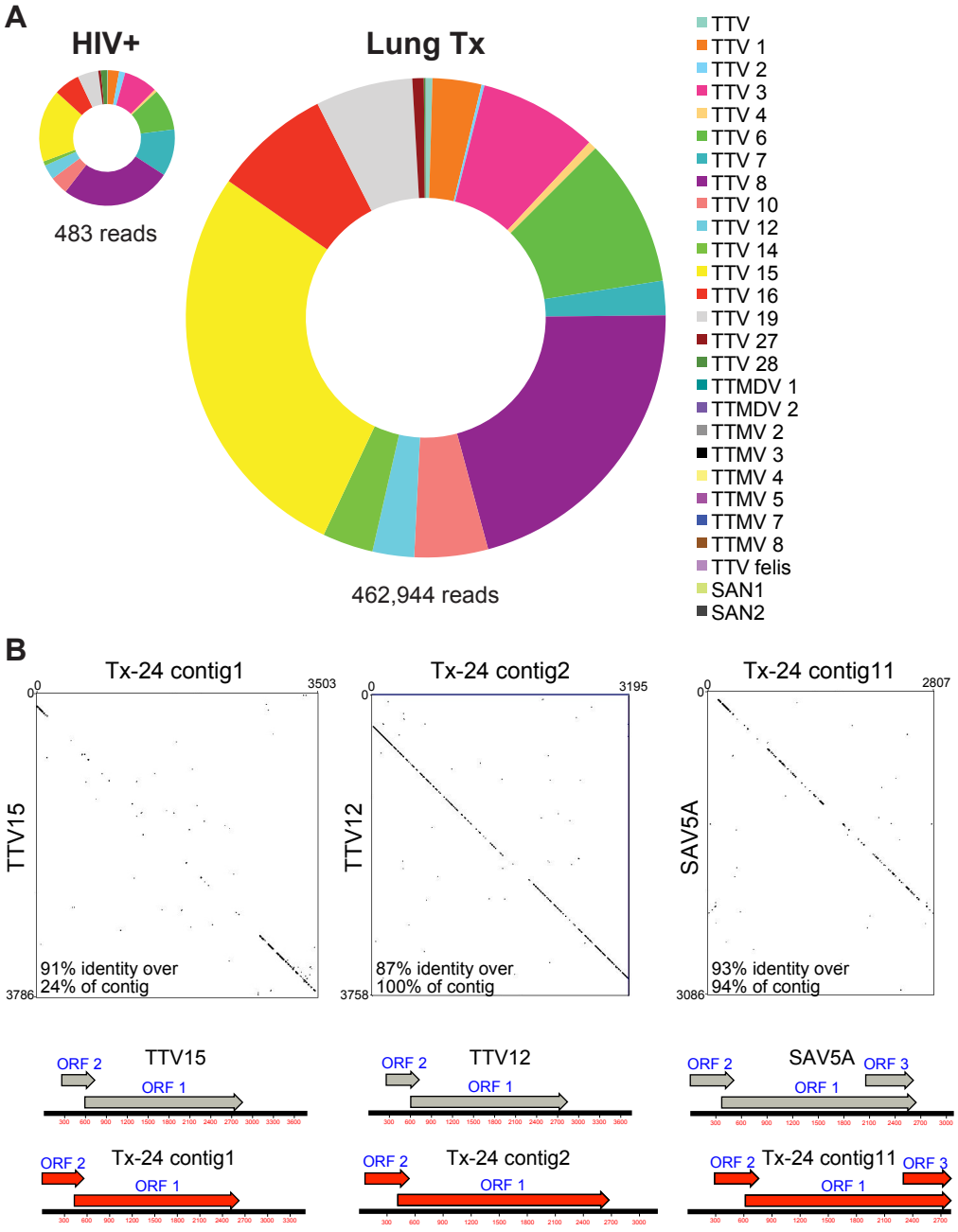


Figure 3.3. Diversity of anellovirus ORF1 sequences in samples from each individual studied. ORF1 amino acid sequences from anellovirus contigs and Genbank reference sequences were aligned and trimmed to generate the phylogenetic tree shown. The names of the reference sequences include the TTV strain name followed by the Genbank identification (gi) number. Clades are designated as groups, in the outer ring, as described by Okamoto et al (12, 22). The BAL sample of origin of each ORF1 sequence is designated by the color code (key at right). The Shimodaira-Hasegawa (SH) score was calculated using FastTree to estimate the reliability of each split compared to alternate topologies (25). Local SH support values over 0.9 are labeled as circles on the nodes of the tree. The scale bar represents the number of amino acid substitutions per position.

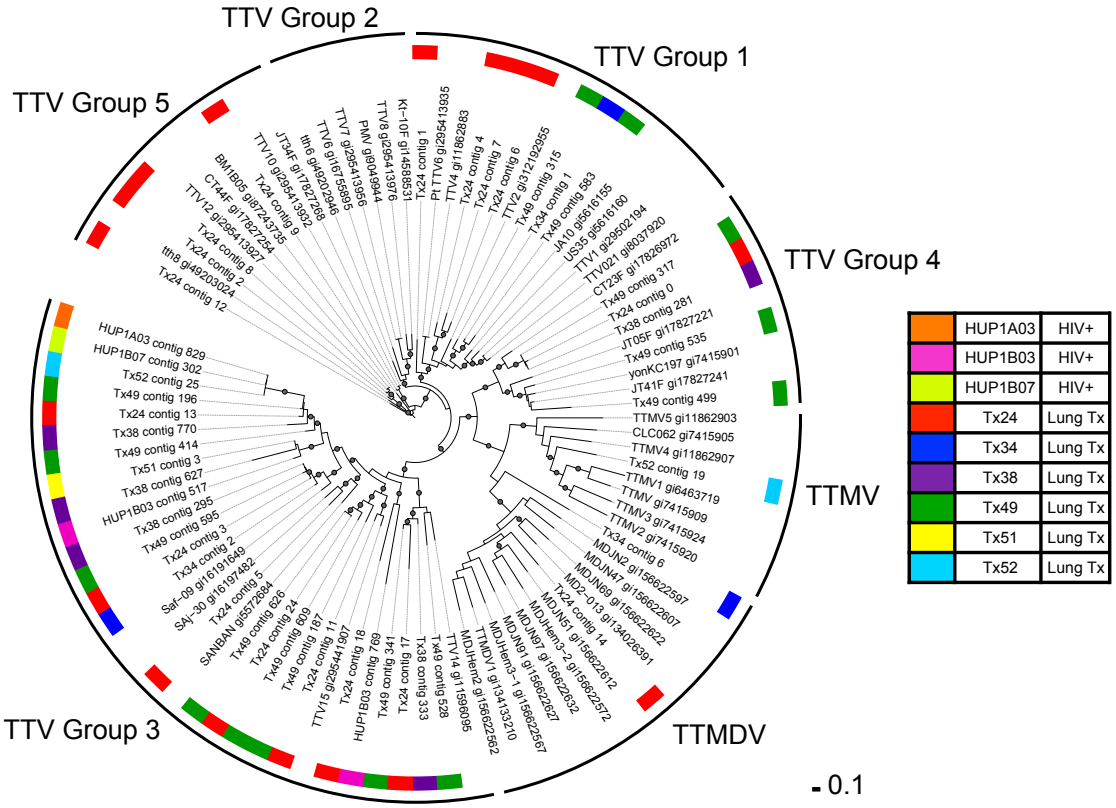


Figure 3.4. Abundance of anelloviruses in BAL samples. Anelloviruses were quantified by Q-PCR in BAL from lung transplant recipients, healthy individuals, and HIV+ individuals. Boxes represent the middle two quartiles for each group and whiskers are placed at the minimum and maximum values. Quantities were higher in lung transplant recipients compared with healthy and HIV+ individuals as determined by the Mann-Whitney test with Bonferroni correction: $p<0.0002$. All samples quantified were above the Q-PCR detection limit (1.4 copies per reaction).

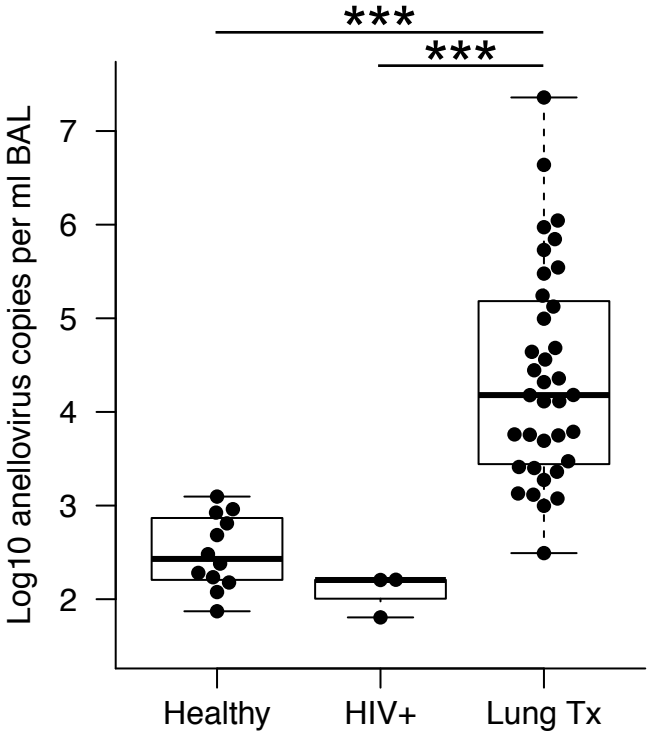


Figure 3.5. Abundance of anelloviruses in OW samples. Anelloviruses were quantified by Q-PCR in oropharyngeal wash from healthy individuals and lung transplant recipients. Anellovirus quantities were higher in the upper respiratory tract of lung transplant recipients compared with healthy individuals. Mann-Whitney test: p=0.0055. All samples quantified were above the Q-PCR detection limit.

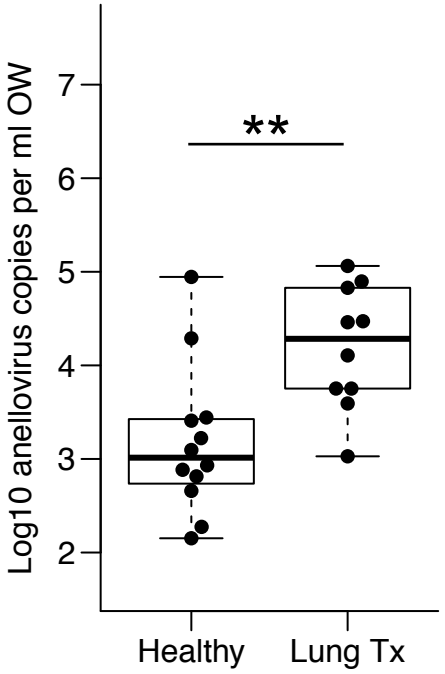


Figure 3.6. Comparison of anellovirus quantities in lung and upper respiratory tract within individuals. Paired Wilcoxon signed rank tests were performed on anellovirus copies from BAL and OW in A) healthy control subjects and B) lung transplant recipients. Anellovirus DNA copies were lower in the lung compared with the upper respiratory tract of healthy controls but not in lung transplant recipients (p-values: 0.0068 and 0.23, respectively).

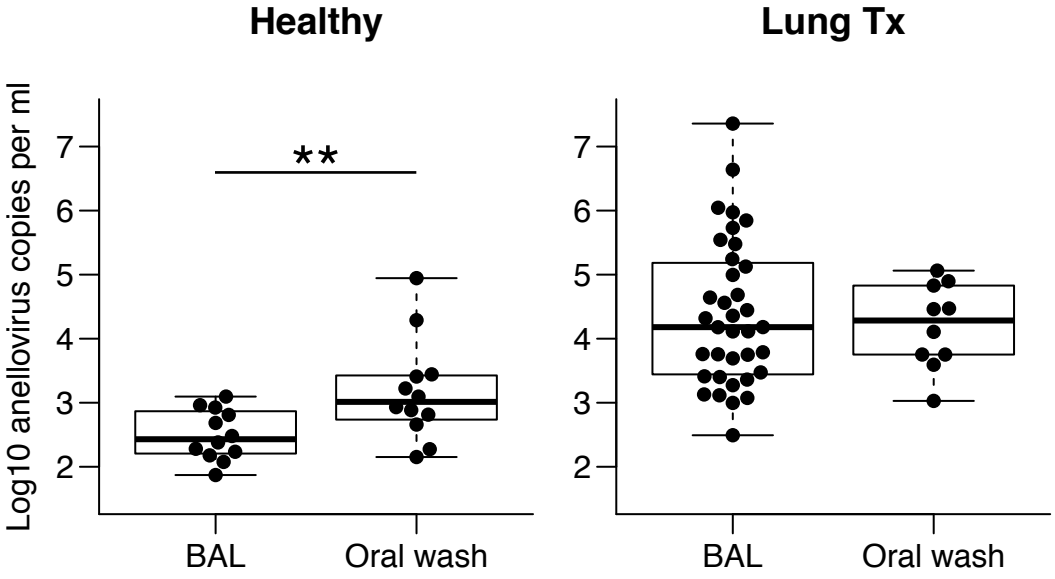
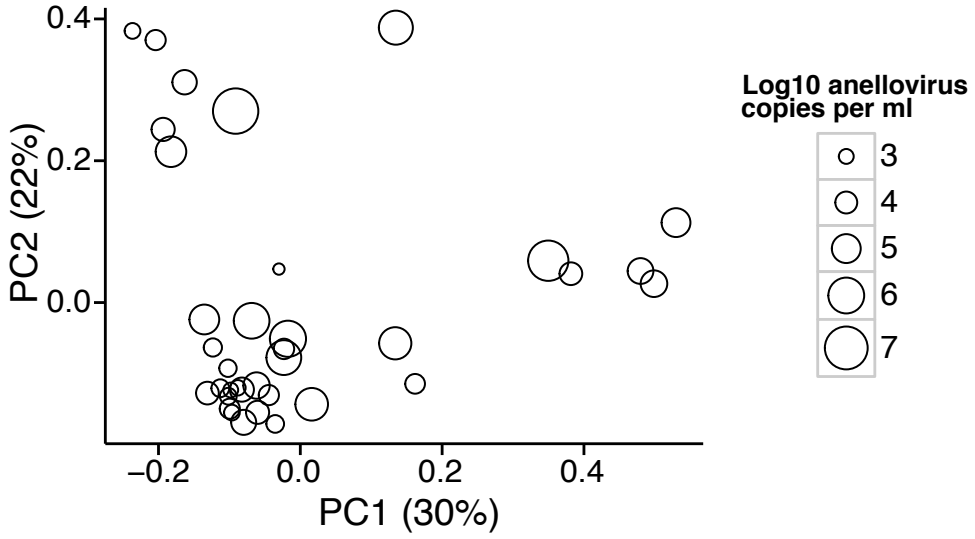


Figure 3.7. Ordination based on composition of bacterial 16S sequence analysis, showing the relationship to anellovirus DNA copy numbers. Each circle on the plot shows data for a single transplant recipient BAL sample. For each sample, bacterial 16S rRNA gene tags were subject to deep sequencing (median 7,925 reads per sample; data published in (20)) and analyzed using Weighted UniFrac, which generated a set of pairwise distances among samples. The distances were then analyzed using Principal Coordinate Analysis (PCoA), and the samples plotted along the first two principal coordinates. For each sample, the anellovirus titer is summarized by the size of the open circle. Among these transplant recipient BALs there is a significant relationship between anellovirus DNA level and 16S bacterial community composition, which is reflected in the observation that samples with low anellovirus titers tend to appear on the lower left side of the PCoA plot (p=0.032; ADONIS test).



References

1. Burguete SR, Maselli DJ, Fernandez JF, Levine SM. Lung transplant infection. *Respirology* 2013;18:22-38.
2. Vu DL, Bridevaux PO, Aubert JD, Soccal PM, Kaiser L. Respiratory viruses in lung transplant recipients: A critical review and pooled analysis of clinical studies. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* 2011;11:1071-1078.
3. Garbino J, Gerbase MW, Wunderli W, Kolarova L, Nicod LP, Rochat T, Kaiser L. Respiratory viruses and severe lower respiratory tract complications in hospitalized patients. *Chest* 2004;125:1033-1039.
4. Clark NM, Lynch JP, 3rd, Sayah D, Belperio JA, Fishbein MC, Weigt SS. DNA viral infections complicating lung transplantation. *Seminars in respiratory and critical care medicine* 2013;34:380-404.
5. Speich R, van der Bij W. Epidemiology and management of infections after lung transplantation. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2001;33 Suppl 1:S58-65.
6. Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, Bushman FD. Rapid evolution of the human gut virome. *Proceedings of the National Academy of Sciences of the United States of America* 2013;110:12450-12455.
7. Breitbart M, Haynes M, Kelley S, Angly F, Edwards RA, Felts B, Mahaffy JM, Mueller J, Nulton J, Rayhawk S, Rodriguez-Brito B, Salamon P, Rohwer F. Viral diversity and dynamics in an infant gut. *Research in microbiology* 2008;159:367-373.
8. Willner D, Furlan M, Haynes M, Schmieder R, Angly FE, Silva J, Tammadoni S, Nosrat B, Conrad D, Rohwer F. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PloS one* 2009;4:e7370.
9. Willner D, Haynes MR, Furlan M, Hanson N, Kirby B, Lim YW, Rainey PB, Schmieder R, Youle M, Conrad D, Rohwer F. Case studies of the spatial heterogeneity of DNA viruses in the cystic fibrosis lung. *American journal of respiratory cell and molecular biology* 2012;46:127-131.
10. Focosi D, Maggi F, Albani M, Macera L, Ricci V, Gragnani S, Di Beo S, Ghimenti M, Antonelli G, Bendinelli M, Pistello M, Ceccherini-Nelli L, Petrini M. Torquetenovirus viremia kinetics after autologous stem cell transplantation are predictable and may serve as a surrogate marker of functional immune reconstitution. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology* 2010;47:189-192.
11. De Vlaminck I, Khush KK, Strehl C, Kohli B, Luikart H, Neff NF, Okamoto J, Snyder TM, Cornfield DN, Nicolls MR, Weill D, Bernstein D, Valantine HA, Quake SR. Temporal response of the human virome to immunosuppression and antiviral therapy. *Cell* 2013;155:1178-1187.
12. De Villiers E-M, Zur Hausen H. Tt viruses: The still elusive human pathogens. Springer; 2009.
13. Hino S, Miyata H. Torque teno virus (ttv): Current status. *Reviews in medical virology* 2007;17:45-57.
14. Maggi F, Fornai C, Zaccaro L, Morrica A, Vatteroni ML, Isola P, Marchi S, Ricchiuti A, Pistello M, Bendinelli M. Tt virus (ttv) loads associated with different peripheral blood cell types and evidence for ttv replication in activated mononuclear cells. *Journal of medical virology* 2001;64:190-194.
15. Maggi F, Pifferi M, Fornai C, Andreoli E, Tempestini E, Vatteroni M, Presciuttini S, Marchi S, Pietrobelli A, Boner A, Pistello M, Bendinelli M. Tt virus in the nasal secretions of children with acute respiratory diseases: Relations to viremia and disease severity. *Journal of virology* 2003;77:2418-2425.
16. Maggi F, Pifferi M, Tempestini E, Fornai C, Lanini L, Andreoli E, Vatteroni M, Presciuttini S, Pietrobelli A, Boner A, Pistello M, Bendinelli M. Tt virus loads and lymphocyte subpopulations in children with acute respiratory diseases. *Journal of virology* 2003;77:9081-9083.
17. Wootton SC, Kim DS, Kondoh Y, Chen E, Lee JS, Song JW, Huh JW, Taniguchi H, Chiu C, Boushey H, Lancaster LH, Wolters PJ, DeRisi J, Ganem D, Collard HR. Viral infection in acute

exacerbation of idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine* 2011;183:1698-1702.
18. Thom K, Petrik J. Progression towards aids leads to increased torque teno virus and torque teno minivirus titers in tissues of hiv infected individuals. *Journal of medical virology* 2007;79:1-7.
19. Madsen CD, Eugen-Olsen J, Kirk O, Parner J, Kaae Christensen J, Brasholt MS, Ole Nielsen J, Krogsgaard K. Ttv viral load as a marker for immune reconstitution after initiation of haart in hiv-infected patients. *HIV clinical trials* 2002;3:287-295.
20. Charlson ES, Diamond JM, Bittinger K, Fitzgerald AS, Yadav A, Haas AR, Bushman FD, Collman RG. Lung-enriched organisms and aberrant bacterial and fungal respiratory microbiota after lung transplant. *American journal of respiratory and critical care medicine* 2012;186:536-545.
21. Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, Bushman FD, Collman RG. Topographical continuity of bacterial populations in the healthy human respiratory tract. *American journal of respiratory and critical care medicine* 2011;184:957-963.
22. Ninomiya M, Takahashi M, Nishizawa T, Shimosegawa T, Okamoto H. Development of pcr assays with nested primers specific for differential detection of three human anelloviruses and early acquisition of dual or triple infection during infancy. *Journal of clinical microbiology* 2008;46:507-514.
23. Peng Y, Leung HC, Yiu SM, Chin FY. Idba-ud: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 2012;28:1420-1428.
24. Krumsiek J, Arnold R, Rattei T. Gepard: A rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 2007;23:1026-1028.
25. Price MN, Dehal PS, Arkin AP. Fasttree 2--approximately maximum-likelihood trees for large alignments. *PloS one* 2010;5:e9490.
26. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. Qiime allows analysis of high-throughput community sequencing data. *Nature methods* 2010;7:335-336.
27. Edgar RC. Search and clustering orders of magnitude faster than blast. *Bioinformatics* 2010;26:2460-2461.
28. 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. *The ISME journal* 2012;6:610-618.
29. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. Pynast: A flexible tool for aligning sequences to a template alignment. *Bioinformatics* 2010;26:266-267.
30. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria.: R Foundation for Statistical Computing; 2011.
31. Charlson ES, Bittinger K, Chen J, Diamond JM, Li H, Collman RG, Bushman FD. Assessing bacterial populations in the lung by replicate analysis of samples from the upper and lower respiratory tracts. *PloS one* 2012;7:e42786.
32. Ogilvie LA, Bowler LD, Caplin J, Dedi C, Diston D, Cheek E, Taylor H, Ebdon JE, Jones BV. Genome signature-based dissection of human gut metagenomes to extract subliminal viral sequences. *Nature communications* 2013;4:2420.
33. Reyes A, Semenkovich NP, Whiteson K, Rohwer F, Gordon JI. Going viral: Next-generation sequencing applied to phage populations in the human gut. *Nature reviews Microbiology* 2012;10:607-617.
34. Zheng H, Ye L, Fang X, Li B, Wang Y, Xiang X, Kong L, Wang W, Zeng Y, Ye L, Wu Z, She Y, Zhou X. Torque teno virus (sanban isolate) orf2 protein suppresses nf-kappab pathways via interaction with ikappab kinases. *Journal of virology* 2007;81:11917-11924.
35. de Villiers EM, Borkosky SS, Kimmel R, Gunst K, Fei JW. The diversity of torque teno viruses: In vitro replication leads to the formation of additional replication-competent subviral molecules. *Journal of virology* 2011;85:7284-7295.

36. Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: Patterns and determinants. *Nat Rev Genet* 2008;9:267-276.
37. Weigt SS, Elashoff RM, Huang C, Ardehali A, Gregson AL, Kubak B, Fishbein MC, Saggar R, Keane MP, Saggar R, Lynch JP, 3rd, Zisman DA, Ross DJ, Belperio JA. Aspergillus colonization of the lung allograft is a risk factor for bronchiolitis obliterans syndrome. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* 2009;9:1903-1911.
38. Vos R, Vanaudenaerde BM, Geudens N, Dupont LJ, Van Raemdonck DE, Verleden GM. Pseudomonal airway colonisation: Risk factor for bronchiolitis obliterans syndrome after lung transplantation? *The European respiratory journal* 2008;31:1037-1045.

Acknowledgements

We are grateful to members of the Bushman and Collman laboratory for help and suggestions. We are grateful to the research subjects who volunteered for this study and the clinicians who assisted with specimen collection; A. Fitzgerald for project management; W. Russell and D. Frame for critical study assistance. This work was supported by NIH awards U01 HL098957 and R01 HL113252. JCY was supported by NIH T32 AI007632. We also acknowledge assistance and support from the Penn Center for AIDS Research (P30-AI045008), and the Penn DNA Sequencing Facility.

Contribution

I was the primary contributor to the viral analysis pipeline used in this publication. JY, KB, and I used the viral analysis pipeline to generate the results. I built figures 3.2A, 3.2B and 3.3 and helped JY with the statistical analysis. I provided input regarding viral analysis strategy and assisted in the writing of the relevant methods.

Supplemental Information

Supplemental Methods

Quantitative PCR

Q-PCR was carried out on non-Genomiphi amplified DNA isolated from BAL and OW. TTV primers NG779 (ACWKMCGAATGGCTGAGTTT) and NG781 (CCCKWGCCCGARTTGCCCCT) targeting the UTR of anelloviruses were used (1), and

the assay adapted to a quantitative SYBR green-based Q-PCR. Briefly, a 126 bp fragment from TTV was amplified by end-point PCR, TA-cloned into the pCR4 TOPO vector (Invitrogen), quantified by PicoGreen, and used to generate a standard curve. 10µl SYBR green FAST master mix (Applied Biosystems), 50µM of each primer, and 10µl standard curve or DNA (diluted 1:10) were added per well. PCR cycle parameters were as follows: 95°C for 90 seconds, and forty cycles of 94°C for 15 seconds, and 68°C for 1 minute. Melt curve analysis of amplification products showed a single transition suggestive of formation of a single DNA product. All samples were run in duplicate or triplicate and values averaged. The detection limit was 1.4 copies per reaction.

Bioinformatics pipeline

Paired-end reads from the MiSeq instrument were quality-trimmed and processed through BMTagger to remove human sequences. Non-human reads were then analyzed using BLAST against the NCBI viral database (downloaded April 2, 2013 from https://www.ncbi.nlm.nih.gov/genomes/Genomes Home.cgi) and the percent of each virus genome covered by reads was calculated. The percentage of paired reads matching the same virus was also calculated. Subsequently, reads were aligned to selected individual viruses one at a time. The alignments were performed using Bowtie2 (2) and visualized using IGV (3). A large number of reads (170,651 across all samples) showed a best BLAST hit from the NCBI viral database to a 50 base region within the RNA plant virus *Physalis mottle virus*. This sequence covered less than 1% of the genome, and also shared 100% nucleotide identity with DNA cloning vectors, so those sequences were removed manually from the analysis.

Reads were assembled into contigs by iterative deBruijn graph assembly using IDBA-UD (4). Sanger sequencing was performed on selected contigs following PCR amplification using contig-specific primers to verify sequences. Dot plots were generated using Gepard with a word size of 10 nucleotides (5). Anellovirus ORF1 amino acid sequences were identified from our contigs and aligned along with 49 reference TTV, TTMDV, and TTMV sequences from Genbank.

Sequences were aligned using MUSCLE v.3.8.31 (6) and trimmed to include only columns with 80% amino acid coverage. Subsequently, the 10% most variable columns were removed. The phylogenetic tree was built using FastTree v. 2.1.3 (7) which calculates local support values using the Shimodaira-Hasegawa (SH) test to estimate the reliability of each split compared to alternate topologies (7-10). The groups, color bars corresponding to the samples, and SH- indices were incorporated using iTOL (11). Groups were chosen based on clades determined by SH-indices (>90%) and a guide tree previously published by Okamoto et al (12).

Supplemental Methods References

1. Ninomiya M, Takahashi M, Nishizawa T, Shimosegawa T, Okamoto H. Development of pcr assays with nested primers specific for differential detection of three human anelloviruses and early acquisition of dual or triple infection during infancy. *Journal of clinical microbiology* 2008;46:507-514.
2. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* 2009;10:R25. 3. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nature biotechnology* 2011;29:24-26.
4. Peng Y, Leung HC, Yiu SM, Chin FY. Idba-ud: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 2012;28:1420-1428. 5. Krumsiek J, Arnold R, Rattei T. Gepard: A rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 2007;23:1026-1028.
6. Edgar RC. Muscle: Multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 2004;32:1792-1797. 7. Price MN, Dehal PS, Arkin AP. Fasttree 2--approximately maximum-likelihood trees for large alignments. *PloS one* 2010;5:e9490.
8. Shimodaira HH, M. Multiple comparisons of log-liklihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 1999;16:1114-1116. 9. Guindon S, Delsuc F, Dufayard JF, Gascuel O. Estimating maximum likelihood phylogenies with phyml. *Methods in molecular biology* 2009;537:113-137.
10. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of phyml 3.0. *Systematic biology* 2010;59:307-321. 11. Letunic I, Bork P. Interactive tree of life (itol): An online tool for phylogenetic tree display and annotation. *Bioinformatics* 2007;23:127-128.
12. De Villiers E-M, Zur Hausen H. Tt viruses: The still elusive human pathogens. Springer; 2009.

Supplemental Figures and Tables

Supplemental figures and tables are available online: *Young* et al, AJT 2014.

CHAPTER 4: Engineering the Gut Microbiota to Treat Hyperammonemia. Journal of Clinical Investigation

The contents of this chapter have been published as:

> Shen T*, Albenberg L*, Bittinger K, **Chehoud C**, Chen Y, Judge C, Wang L, Sheng M, Lin A, Wilkins B, Lewis J, Daikhin Y, Nissim I, Yudkoff M, Bushman F, Wu G. 2015. *Engineering the Gut Microbiota to Treat Hyperammonemia.* Journal of Clinical Investigation (PMID: 26098218).

Abstract

Increasing evidence indicates that the gut microbiota can be altered to ameliorate or prevent disease states, and engineering the gut microbiota to therapeutically modulate host metabolism is an emerging goal of microbiome research. In the intestine, bacterial urease converts host-derived urea to ammonia and carbon dioxide, contributing to hyperammonemia-associated neurotoxicity and encephalopathy in patients with liver disease. Here, we engineered murine gut microbiota to reduce urease activity. Animals were depleted of their preexisting gut microbiota and then inoculated with altered Schaedler flora (ASF), a defined consortium of 8 bacteria with minimal urease gene content. This protocol resulted in establishment of a persistent new community that promoted a long-term reduction in fecal urease activity and ammonia production. Moreover, in a murine model of hepatic injury, ASF transplantation was associated with decreased morbidity and mortality. These results provide proof of concept that inoculation of a prepared host with a defined gut microbiota can lead to durable metabolic changes with therapeutic utility.

Introduction

Dysbiosis, an abnormal and pathogenic state of the human microbiome, has been implicated in inflammatory bowel diseases, atherosclerosis, obesity, diabetes, colon cancer, and other diseases (1, 2). Fecal microbiota transplantation (FMT) is highly effective in the treatment of refractory *Clostridium difficile* infection (CDI), providing proof of principle that a human disease can be treated by engineering the gut microbiota (3), and further studies indicate that a healthy microbiota can prevent disease acquisition (4).

Bacteria residing in the human gut produce urease, the activity of which is beneficial in healthy hosts but pathogenic in hosts with liver disease. Urea produced by the liver as a waste product is both excreted in urine and transported into the colon, where it is hydrolyzed by bacterial urease into carbon dioxide and ammonia. Ammonia is then (a) utilized by the microbiota for protein synthesis, (b) reabsorbed by the host, where it is incorporated into the nitrogen pool by hepatic metabolism, or (c) excreted in the feces. Mammalian genomes do not encode urease genes, so ammonia production results from bacterial urease activity acting on host-produced urea (5, 6). Ammonia is also largely responsible for the alkaline pH of the colonic luminal environment, acting to buffer the short-chain fatty acids also produced by the microbiota. Systemic ammonia levels are elevated in patients with liver injury, chronic liver disease, or urea cycle defects, whose hepatic abnormalities prevent the normal processing of ammonia delivered to the liver from the intestinal tract. Circulating ammonia is correlated with damage to the CNS in patients with chronic liver disease or inborn errors of metabolism, resulting in hepatic encephalopathy (HE) (7, 8).

Current treatments for HE and hyperammonemia are inadequate (8). Antibiotics (ABX) traditionally used to treat HE, including aminoglycosides and metronidazole, are limited by side effects and concerns for safety including ototoxicity, nephrotoxicity, and peripheral neuropathy (9, 10). Although rifaximin, a minimally absorbed ABX, has shown efficacy in the treatment and prevention of HE (11, 12), potential development of antimicrobial resistance with long-term use remains a concern. Lactulose is used to acidify feces and sequester ammonia as ammonium, but

lactulose is poorly tolerated, resulting in poor adherence (13). In a mouse model of thioacetamide-induced liver injury, a lactobacillus pro- biotic has been reported to reduce ammonia levels and mortality (14), but these benefits have not been extended to human studies (15, 16). While promising, studies of probiotic therapies in humans described to date have suffered from methodological limitations, did not document long-term effects, and showed consistently minimal effects on outcome, motivating efforts to engineer more resilient and effective bacterial communities.

Here, we show that a synthetic microbial community lacking urease activity can be installed in the gut to reduce the production of ammonia long term and thereby mitigate HE in a mouse model. For proof-of-concept studies, we used altered Schaedler flora (ASF), which consists of 8 murine gut commensal bacterial strains that were assembled in the 1970s and standardized by the National Cancer Institute in 1978 (17). The strains were originally selected on the basis of their persistence from generation to generation in germ-free mice and their ability to restore a cecal morphology that is comparable to that of conventional mice. The ASF community is innocuous, known to have a beneficial effect of inducing immune tolerance (18), and is used by commercial mouse vendors to enhance the health of immunodeficient mouse strains. We found that ASF is low in urease activity, making it useful for the studies of metabolic engineering described here. Mixed results have been reported regarding the transfer of conventional microbiota between rodents, with some studies reporting successful transfer with repetitive inoculation (19), but others demonstrating that the use of ABX prevented transfer (20). We developed methods for purging the gut microbiota from normally colonized mice that were then transplanted with ASF by gavage. Transplantation of ASF was monitored longitudinally using deep sequencing of DNA from fecal pellets, revealing highly efficient colonization in properly prepared hosts. Over a 4-week monitoring period, ASF was partially displaced by selective colonization with environmental Firmicutes, reaching a new steady state, but fecal ammonia levels remained low. The engineered gut community was tested in thioacetamide (TAA) models of acute and chronic liver injury (Figure 4.1A), in which we demonstrate that it reduced fecal

ammonia levels, mortality, and neurobehavioral deficits. These results show that transplantation of a defined minimal microbial community can alter metabolism in a predetermined fashion by establishing a new gut microbiome, resulting in therapeutic benefits in disease models.

Results

Transplantation of ASF into previously colonized mice

The original ASF was composed of 8 bacterial strains. Over time, ASF has been maintained in laboratory mice by fecal-oral transmission associated with cohousing in gnotobiotic isolators. Thus, the composition of the ASF donor material used here was first characterized by DNA isolation and shotgun metagenomic sequencing (16.9 Gb). Alignment to draft genome sequences of ASF strains (21) documented the presence of 7 of the 8 original strains in our samples (Figure 4.1B). *Parabacteroides* (ASF519) was the predominant lineage present in pellets. Additional ASF strains consisted of roughly even proportions of Clostridia (ASF356, ASF492, ASF500, ASF502) and *Mucispirillum schaedleri* (ASF457), accompanied by low levels of *Lactobacilli* (ASF361). No high-quality read pairs mapped concordantly to strain ASF360. In addition to ASF strains, 11% of reads mapped to the mouse genome, and a small fraction of alignments mapped to probable artifacts including cloning vectors, metazoans, and additional bacteria. We conclude that the bacteria in our donor material were mostly or entirely ASF strains.

To assess the persistence of ASF in mice housed under non-sterile conditions, sequential fecal pellets were collected from 10 ASF-colonized mice that were transferred to conventional SPF housing, DNA was purified from pellets, and the abundance and types of bacteria present were assessed by quantitative PCR (qPCR) and deep sequencing of 16S rRNA V1V2 gene segments (Sup. Figure 4.1). Copy numbers of 16S sequences per gram of stool showed roughly similar abundance for ASF and conventionally colonized mice. Six of the eight ASF strains were resolved at the depth of sequencing performed.

ASF519 (*Parabacteroides*) accounted for approximately 75% of sequence reads at the time of transfer of the mice. After approximately 2 months under nonsterile conditions, ASF519 remained the dominant taxon, and non-ASF taxa accounted for almost half of the sequence reads. Evidently, ASF lineages persisted but did not entirely exclude other bacteria.

We reasoned that reduction of the endogenous microbiota would promote the transfer of ASF, so we investigated the response of conventional microbiota to the oral delivery of 2 non-absorbable ABX, neomycin and vancomycin. The 16S rRNA gene copy number was reduced by approximately 4 logs within 72 hours of oral ABX initiation ($P < 0.0001$ for days 0, 1, and 2 compared with the mean of days 5–15), then returned to baseline 5 days after discontinuing ABX (Figure 4.1C; $P = 0.36$ for comparison of ABX versus control at day 21), paralleling previous studies (22, 23).

Fecal slurries obtained from ASF-colonized mice (Taconic) were then gavaged into conventionally housed recipients for 7 days following a 72-hour pretreatment with oral ABX and a 12-hour intestinal purge using polyethylene glycol (PEG) (pretreatment with ABX and PEG is henceforth termed “prepared”). PEG was used in our gut-cleansing protocol, because the use of a purgative will likely be necessary to reduce bacterial load in the human intestinal tract due to high biomass. For comparison, ASF transplants were carried out on conventional mice without pretreatment and on germ-free mice. Mice that were treated with ABX and PEG and subsequently transplanted showed normal numbers of bacteria by 16S qPCR copy numbers within 10 days (data not shown).

Longitudinal fecal samples were analyzed by deep sequencing of 16S rRNA gene tags, then the proportions of bacterial lineages detected were plotted as heatmaps, in which each row shows a bacterial lineage and each column a fecal specimen (Figure 4.2). Conventional mice that were not pretreated showed no increase in ASF lineages, despite ASF gavage (Figure 4.2, Conventional + ASF gavage). A few preexisting lineages were present that were indistinguishable from ASF using the V1V2 16S sequence window, but these did not increase in abundance after

ASF gavage. A second group of mice (discussed above) were colonized with ASF from birth and then moved to the non-sterile facility. These mice had high levels of ASF519 and detectable levels of 4 other ASF strains (Figure 4.2, ASF-colonized).

These groups were then compared with (a) germ-free mice gavaged with ASF (referred to herein as germ-free/ASF mice); and (b) conventional mice pretreated with ABX and PEG and gavaged with ASF (referred to herein as prepared/ASF mice) (Figure 4.2, Germ-free + ASF gavage and Prepared host + ASF gavage). Prior to ASF gavage, both the germ-free animals and ABX-treated animals showed high levels of *Lactococcus*, a lineage found at high levels in mouse chow (22), indicating that endogenous gut bacteria were mostly or entirely absent. Gavage of these animals with ASF resulted in establishment of ASF lineages that persisted for the duration of the sampling period. The communities were again dominated by ASF519.

Longitudinal evolution of the transplanted ASF community

Persistence of the transplanted ASF community was quantified using segmental regression, plotting time against the proportion of ASF in each sample (Figure 4.3A). Prepared/ASF mice and germ-free/ ASF mice transferred to non-sterile conditions were compared and found to behave similarly. Initially, ASF comprised the majority of the community. After transfer, the proportion of ASF slowly declined, constituting approximately 45% of the gut community after 34 days (Davies' test for a change in the slope $P < 0.001$; 95% CI for the breakpoint was 27–41 days by segmented linear regression). After 34 days, the proportion of ASF strains did not decrease for the duration of the experiment (120 days; $P = 0.86$). Comparisons of community membership using unweighted UniFrac analysis of the 16S rRNA gene-sequencing data showed that non-ASF lineages had colonized by day 14, and members persisted for the duration of the study (Figure 4.3B), although the final abundance of non-ASF lineages was not achieved until about day 30 (Figure 4.3A). Diversity of this new stable state measured by the Shannon index approached that of the starting community (Figure 4.3C). ASF519 (*Parabacteroides*) was the main taxon persisting in gavage from both germ-free/ASF and

prepared/ASF hosts after 6 weeks, constituting approximately 40% and 50%, respectively, in each group (Figure 4.2). As in humans, bacteria belonging to the *Bacteroides* genus were dominant taxa in conventionally housed mice (24, 25).

The new lineages appearing after transplantation were specific (Figure 4.4). Members of the Bacteroidetes phylum did not recolonize after ASF transplantation, suggesting that ASF519, a member of the closely related *Parabacteroides* genus, may have occupied niches available to this group. Several Firmicutes lineages did colonize over time, including lineages annotated as *Oscillibacter*, *Clostridium*, and "other" Firmicutes (mostly of the order Clostridiales). Thus, the community achieved a new steady state containing both ASF and environmentally acquired lineages.

ASF has minimal urease gene content and activity

An analysis of the complete ASF genomes showed a minimal presence of urease genes. No urease genes were identified in the predominant *Parabacteroides* ASF519. Two ASF members contained urease genes, ASF492 and ASF361, but both were minor members of the community after transplantation (Figure 4.2).

We then characterized urease activity, which we found readily detectable in pellets from conventionally housed mice, but undetectable in pellets from mice colonized with ASF (Figure 4.5A). Similar results were obtained with pellets from mice treated with oral ABX that reduced the bacterial load by 10,000-fold (Figure 4.5A). Intravenous delivery of 13C-urea to quantify urease activity in vivo through the production of 13CO2 in a breath test revealed minimal hydrolysis in ASF-colonized mice (Figure 4.5B).

Transplantation of ASF into prepared hosts led to a reduction in fecal urease activity lasting for at least 80 days (Figure 4.5C). The novel community assembled after 30 days had low urease activity, even though new lineages became established in addition to ASF (Figure 4.2). Analysis of representative genomic sequences from these newly established genera showed that

Oscillibacter, *Dorea*, *Enterococcus*, and *Roseburia* do not encode urease genes. We found that *Clostridium* and the group annotated as “other” Firmicutes (mostly Clostridiales) were mixed, with some representatives encoding urease genes, while others did not. It is unknown whether the newly proliferating organisms lacked urease genes, or whether they encoded urease genes but expressed them at low levels. Thus, the net effect of ASF colonization, together with the establishment of additional lineages, was the achievement of a new steady state with low urease activity. The new community persisted over the long term, as ASF-transplanted mice showed low fecal ammonia levels for over 1 year in a specific pathogen–free (SPF) housing facility (data not shown).

The resilience of this new community state following dietary stress was evaluated by placing mice transplanted with ASF on a low-protein diet similar to that used for patients with hyperammonemic inborn errors of metabolism (26). Fecal ammonia levels remained much lower than those achieved by a low-protein diet alone, an effect that was durable for several months (T.-C.D. Shen and G.D. Wu, unpublished observations). Another dietary stress, consumption of a non-irradiated diet, led to the displacement of ASF by other bacteria within weeks after FMT, with the predominant taxon *Parabacteroides* ASF519 being partially replaced by other taxa within the Bacteroidetes phylum (Sup. Figure 4.2). Nevertheless, the reduction in fecal ammonia levels compared with those at baseline remained significant for several months, even in mice on a non-irradiated diet (Figure 4.5D). Reductions in fecal ammonia levels have been correlated with reductions in blood ammonia (14, 27, 28), indicating that changes in colonic ammonia production and/or absorption can be associated with blood levels. The reduction in fecal ammonia may alter levels of false neurotransmitter precursors produced by the gut microbiota and/or by the host, since ammonia is a substrate for both, and this would lead to reduced formation of the biogenic amines that are hypothesized to play a role in HE (29–31). Thus, we propose that fecal ammonia is a useful biomarker for response of the host to the treatment of hyperammonemia.

ASF transplantation reduces mortality and cognitive impairment in murine models of acute and chronic liver injury

A major cause of morbidity and mortality associated with acute liver injury is the development of HE. Since hyperammonemia is associated with the development of HE in patients with impaired hepatic function (8), we asked whether the transplantation of ASF might mitigate the effects of acute hepatic injury induced by TAA (32) treatment. Prepared/ASF mice showed both a reduction in fecal ammonia levels (Sup. Figure 4.3A, Prepared + ASF) and reduced mortality in response to high-dose TAA compared with mice with conventional microbiota (Figure 4.6A). This finding was also seen in the setting of chronic liver injury in mice transplanted 3 weeks prior to the chronic delivery of TAA at low, escalating doses for 7 weeks (Figure 4.6B). Compared with control mice, prepared/ASF mice demonstrated markedly reduced mortality rates that were maintained over the 7-week period during which hepatic fibrosis developed in both groups (refs. 33, 34; data not shown), consistent with a sustained reduction in fecal ammonia for several months after ASF transplantation.

In mice, the TAA model has also been associated with neurobehavioral abnormalities resembling HE in humans (35). Using a lower dose of TAA to reduce mortality, we analyzed memory and spatial learning in a Y maze test (36) comparing prepared/ASF mice with mice transplanted with normal microbiota (referred to herein as prepared/normal microbiota mice; Figure 4.7 and Supplemental Figure 4.3B and 4.3C) as a control. The 80%–90% survival rates were similar between the 2 groups at the lower TAA dose (Sup. Figure 4.3B). Fecal ammonia levels were reduced in mice transplanted with ASF compared with those with normal microbiota (Sup. Figure 4.3C). Prepared/normal microbiota mice showed a decrease in cognitive function after TAA treatment, quantified as spontaneous alternations in a Y maze test, whereas prepared/ASF mice treated with TAA were not different from untreated controls (Figure 4.7A). Mice transplanted with either ASF or normal microbiota did not sufficiently differ in locomotor activity after TAA treatment (as quantified by the total distance traveled and number of arm

entries in the Y maze) to account for the difference in spontaneous alternations, although both groups exhibited significantly less locomotor activity compared with untreated controls (Figure 4.7B and 4.7C). To exclude the possibility that ASF transplantation directly reduced liver injury, we measured plasma alanine aminotransferase (ALT) and quantified histologic evidence of hepatocyte necrosis (Sup. Figure 4.4, A and B). Both revealed that liver damage induced by TAA was not reduced by either ASF transplantation or ABX treatment. Thus, improved survival and behavioral performance were associated with reduced ammonia levels and not improved locomotor activity or liver injury.

Discussion

The success of FMT in the treatment of CDI establishes that transplanting a resilient microbial community can alter a dysbiotic microbiota and thereby treat disease. Feces, however, contains not only bacteria but also a multitude of archaea, fungi, and viruses, so there is concern for safety (37), motivating the development of defined microbial consortia for human inoculation that have well-characterized biological properties and respond to the gut environment in predictable ways. Here, we show that a defined minimal consortium of bacteria, ASF (17, 18), can durably reprogram the composition and metabolic function of the gut microbiota when inoculated into a properly prepared host. By taking advantage of the minimal urease activity in ASF, we provide evidence that reprogramming the gut microbiota can lead to lower fecal ammonia levels and mitigate the morbidity and mortality associated with liver damage.

The endogenous gut microbiota needed to be depleted by treatment with oral ABX and PEG for efficient transfer of the ASF community into conventionally housed mice. After the gut purge, transfer was as effective as that into germ-free recipients. Tracking by sequencing suggested an orderly succession of lineages. After ASF transplantation, the predominant ASF519 strain appeared at the earliest times after gavage. *Mucispirillum schaedleri* (ASF457) and Ruminococcaceae (ASF500) also appeared early on. The last to appear was the *Clostridium* species (ASF356) on day 14 in some mice, 1 week after gavaging was complete, perhaps

indicating the need for development of a specific niche that permitted the establishment of more fastidious taxa. Most mice contained a *Lactobacillus* strain indistinguishable from ASF361 prior to transplantation, so this lineage could not be tracked with the methods used. The observed succession may be similar to the succession of bacterial taxa in human infants, as consumption of oxygen by initial gut colonizers allows the expansion of bacterial clades that are obligate anaerobes (38, 39).

By monitoring the composition of the transplanted ASF community over a 4-month period using 16S rRNA gene sequencing, we were able to assess the persistence of the community in the nonsterile SPF environment. The ASF community did not fully exclude other taxa. The community appeared to achieve a new steady state, whereby both types of transplanted hosts came to resemble the mice colonized with ASF from birth and housed long term in an SPF environment. In each of these 3 groups, *Parabacteroides* (ASF519) remained the dominant taxon. Bacteria of the *Bacteroides* genus commonly dominate the human and murine gut microbiota (24, 40–43), but new *Bacteroides* did not accumulate over time in ASF-colonized mice, suggesting that *Parabacteroides* may have excluded *Bacteroides*, reminiscent of the trade-off between *Bacteroides* and *Prevotella* in human gut (24, 44–46). The mechanism by which *Parabacteroides* excludes *Bacteroides* is unknown but, due to their taxonomic and functional similarities, may involve competition for limiting resources (47) such as has been shown for glycans and the competition between *Bacteroides* species in the colonic crypt (48). Over time, the *Parabacteroides* and environmental Firmicutes established a new steady state approximating the composition of conventional microbiota in humans and mice (40, 41, 49), possibly involving a syntrophic relationship between these lineages (50).

There was a sustained reduction in fecal urease enzymatic activity and ammonia production upon transfer of ASF into prepared mice. There was no return of urease activity several months after transfer, despite a substantial increase of non-ASF taxa. A possible lack of urease genes in many of the *Clostridium* taxa that accumulated after transplantation may explain

the persistently reduced urease activity, although we cannot exclude other mechanisms, since the regulation of urease enzymatic activity is known to be complex (6).

Production of ammonia by the gut microbiota has been implicated in host nitrogen balance (51–53), so there is a theoretical risk that a urease-free community might have an adverse effect on protein balance and growth of the host. However, we have tracked mice for over 1 year after ASF transplantation and observed no adverse effects on their body weight or mortality. ASF is acknowledged to be an innocuous bacterial consortium in mice with beneficial effects on immune tolerance (18). Nevertheless, additional safety studies will need to be performed using a humanized version of ASF in rodent models before human studies can be contemplated.

In summary, given that HE is a major contributor to morbidity and mortality in liver disease (54), transplantation of next-generation engineered communities based on ASF, coupled with improved preparation of the host as described here, represents a promising approach to more effective therapy.

Methods

Animals

Feces from CB17 SCID mice colonized with ASF (Taconic) were the source of ASF in the transplantation experiments. Mice in this study were maintained in a standard SPF barrier facility and were fed irradiated AIN-76 chow (Research Diets Inc.) containing protein as 21% of the kilocalories. For the nonirradiated diet experiment, we used Laboratory Rodent Diet 5001 (LabDiet) containing protein as 29% of the kilocalories. For microbiota transfer experiments, 0.1 g feces was diluted 10-fold in PBS. For Figures 4.2 to 4.4, germ-free recipient mice were purchased from Taconic, as were the conventionally housed Swiss Webster mice used in the first experiment. For all other experiments, we used either male or female 8-week-old C57BL6 mice (The Jackson Laboratory). Pretreated conventional mice were prepared for inoculation by oral

delivery of ABX in the drinking water (1.125 g aspartame, 0.15 g vancomycin, and 0.3 g neomycin in 300 ml sterile water) for 72 hours. During the final 12 hours, the water supply was exchanged with a 10% PEG solution (Merck), and the mice were fasted. The mice were then inoculated daily with feces by oral gavage for 5 to 7 days. Fecal pellets were collected for bacterial taxonomic and biochemical analyses at the time points indicated in the figures (Figure 4.1C, Figures 4.2- 4.4, Figure 4.5A, C, and D, Sup. Figure 4.1, Sup. Figure 4.2, and Sup. Figure 4.3A and C).

DNA isolation, qPCR, sequencing, and analysis

DNA was isolated from stool as previously described (46, 55). Bacterial 16S rRNA gene sequences were PCR amplified with primers binding to the V1V2 region (46, 55) using bar-coded primers (56, 57). Shotgun metagenomic data were collected using the TruSeq library preparation method and a HiSeq instrument (both from Illumina). Sequence reads were quality controlled and analyzed using the QIIME pipeline with default parameters (58). Sequence data sets were deposited in the NCBI's Sequence Read Archive (SRA) database (accession number SRP058968).

Urease activity and ammonia assays

Fecal ammonia levels were determined using an Ammonia Assay Kit (ab83360; Abcam). Fecal pellets were suspended in the assay buffer provided at a concentration of 1 mg/10 µl and centrifuged at 13,000 *g* for 10 minutes at room temperature to remove insoluble material. Ammonia concentration was then determined according to the kit protocol.

Fecal urease activity was measured by suspending fecal pellets in 0.5 mM HEPES buffer. After sonication and centrifugation, the supernatant was incubated for 30 minutes at 37°C with 1 µCi ^{14}C-labeled urea (19.9 mCi/mmol; catalog ARC-0150; American Radiolabeled Chemicals) in a sealed container. The air was purged into a trap containing 2.5 ml of 0.2 M benzethonium hydroxide in methanol (catalog 82156; Sigma-Aldrich), and $^{14}CO_2$ activity was

quantified by liquid scintillation counting. A standard curve was generated using purified *E*. *coli* urease (150 IU/mg; catalog 22060744-1; BioWORLD).

The urease breath test was performed on 4-hour fasted mice by placing them in a sealed glass chamber. A total of 4 cc air was withdrawn from the chamber using a gas-tight syringe after 10 minutes and injected into 12 ml (gas-helium) Exetainer glass vials (438B; Labco) to establish a baseline CO_2 level. The mouse was then injected via the tail vein with [^{13}C]urea (150 μg/g body weight) (CLM-311-0; Cambridge Isotope Labs). Blood was collected 15 minutes after injection to assess total urea and isotopic enrichment in [^{13}C]urea. Enrichment in expired air was measured at 30, 60, 120, 180, and 240 minutes after injection of the labeled urea. The $^{13}CO_2/^{12}CO_2$ ratio was measured in gas samples with a Finnigan DELTAplus isotope ratio mass spectrometer (Thermo Fisher Scientific). A commercial CO_2 source (Airgas) was used as the standard. The measurement of isotopic abundance in [^{13}C]urea in plasma involves the elimination of CO_2 from the sample and the subsequent conversion of urea to CO_2 with commercially available urease (59).

Induction of acute liver injury and hepatic fibrosis

Mice in the high-dose TAA acute liver injury model were given a single i.p. injection of TAA at 600 mg/kg. In the low-dose TAA acute liver injury model, mice were given a single dose of TAA at 300 mg/kg by i.p. injection, with subsequent performance of the Y maze neurobehavior test 48 hours after TAA injection (see *Neurobehavior test* below). In the chronic liver fibrosis model, all mice were given TAA by i.p. injection 3 times weekly, with the initial 100 mg/kg dose being decreased to 50 mg/ kg during the first week, given the rate of morality, followed by dose escalation to 100 mg/kg during the second week, 200 mg/kg during the third week, 300 mg/kg during the fourth week, and 400 mg/kg during the fifth to seventh weeks of TAA administration.

Neurobehavior test

The Y maze test was conducted to assess the rodents' memory and spatial learning (60). The maze consists of 3 identical and equally spaced arms, and the natural tendency of the mouse is to investigate a new arm of the maze rather than return to the one previously visited. A mouse with a cognitive deficit would exhibit less spontaneous alternation, defined as entering all 3 arms in 3 sequential arm entries. Control mice typically exhibit greater than 60% spontaneous alternation. In this experiment, mice were allowed to habituate to the testing room, where there were no overt visual cues, for 30 minutes prior to testing. The maze was cleaned with 70% ethanol before use and between trials to eliminate odor cues. A trial started when a mouse was released into 1 arm of the maze (same arm for each mouse). As the mouse navigated the maze, each arm entry was noted. At the end of the timed trial (8 minutes), total arm entries were summed and spontaneous alternations determined by the following formula: percentage of spontaneous alternations = [(number of alternations)/(total arm entries − 2)] × 100. Trials were video recorded as well as graded during the procedure. Image analysis software was used to measure the total distance traveled by each mouse during the trials.

Statistics

Results are expressed as the mean ± SEM. Statistical significance among 3 or more groups was assessed by ANOVA. A 2-tailed Student's *t* test and paired-sample *t* test were used for direct comparisons between 2 groups and within groups, respectively. Tukey's test was used to adjust for multiple comparisons. Kaplan-Meier survival curves were compared using the log-rank test. A *P* value of less than 0.05 was considered statistically significant.

Study approval

All animal studies were performed with the approval of the IACUC of the University of Pennsylvania.

Figures

Figure 4.1. Transfer of ASF into a previously colonized murine host. (A) Diagram of the experimental method. Ure, urease. (B) Shotgun metagenomic analysis of stool from ASF-colonized animals used for gavage in this study. Proportions of the different ASF lineages and other organisms are indicated by the color key. (C) Time course of 16S rRNA gene copy numbers during oral ABX treatment (14 days, vancomycin and neomycin) and upon discontinuation ofABXonday15(*n*=3pergroup). *$P < 0.0001$, for days 0–2 compared with the average of days 5–15 in the ABX group; **$P < 0.05$, between the ABX and control groups. Paired-sample *t* test and 2-tailed Student's *t* test.

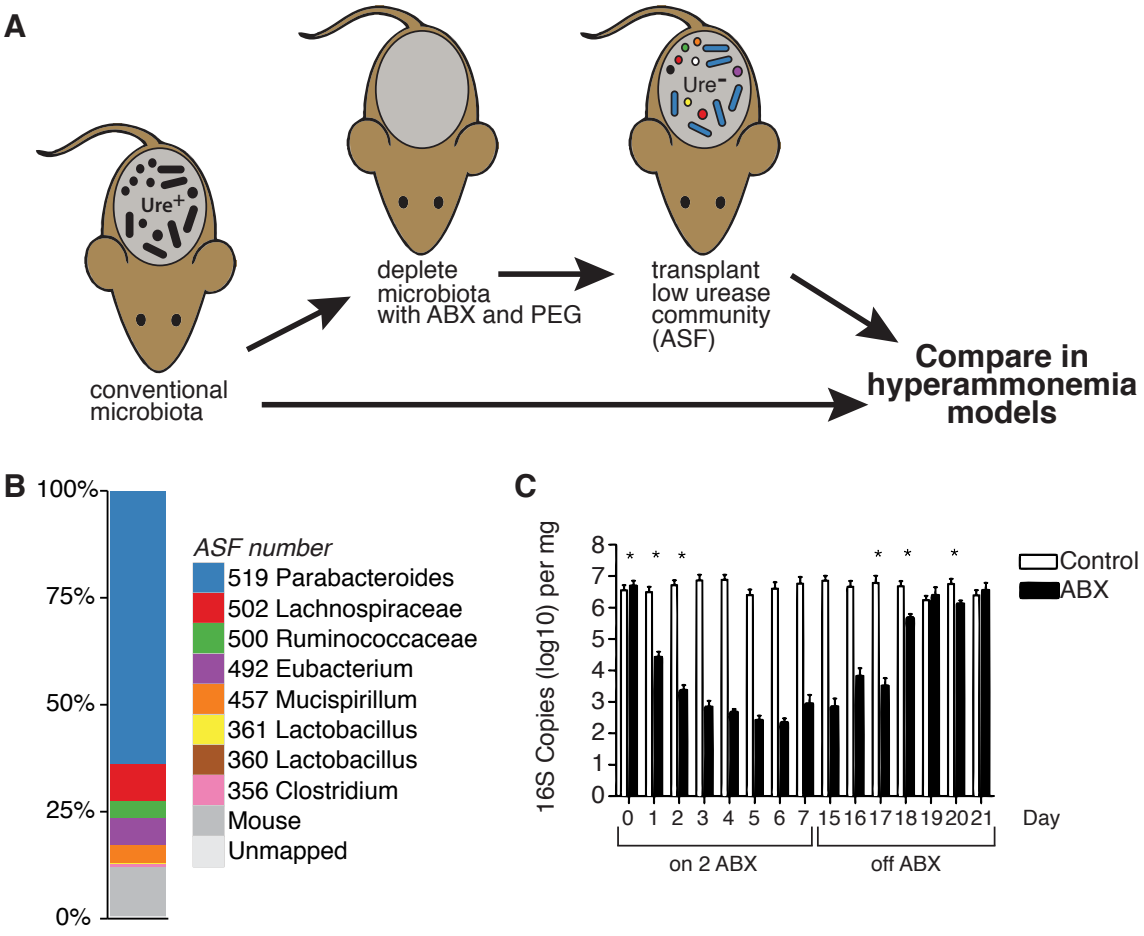


Figure 4.2. Heatmap showing the relative abundance of bacterial lineages over time in ASF-colonized mice and controls. Rows indicate bacterial lineages as annotated on the left. Relative abundance is indicated by the color key at the bottom of the figure. Columns summarize the sequencing results from individual fecal specimens. Elapsed time in days is shown along the bottom. The groups studied are indicated at the top of the heatmap and include (from the left) conventional mice that were gavaged with ASF stool without preparation (Conventional + ASF gavage); mice that were ASF colonized from birth, then transferred to a nonsterile SPF facility (ASF-colonized); mice that were germ-free, then gavaged with ASF (Germ-free + ASF gavage); and conven- tional mice that were prepared with ABX and PEG treatment, then gavaged with ASF (Prepared host + ASF gavage).

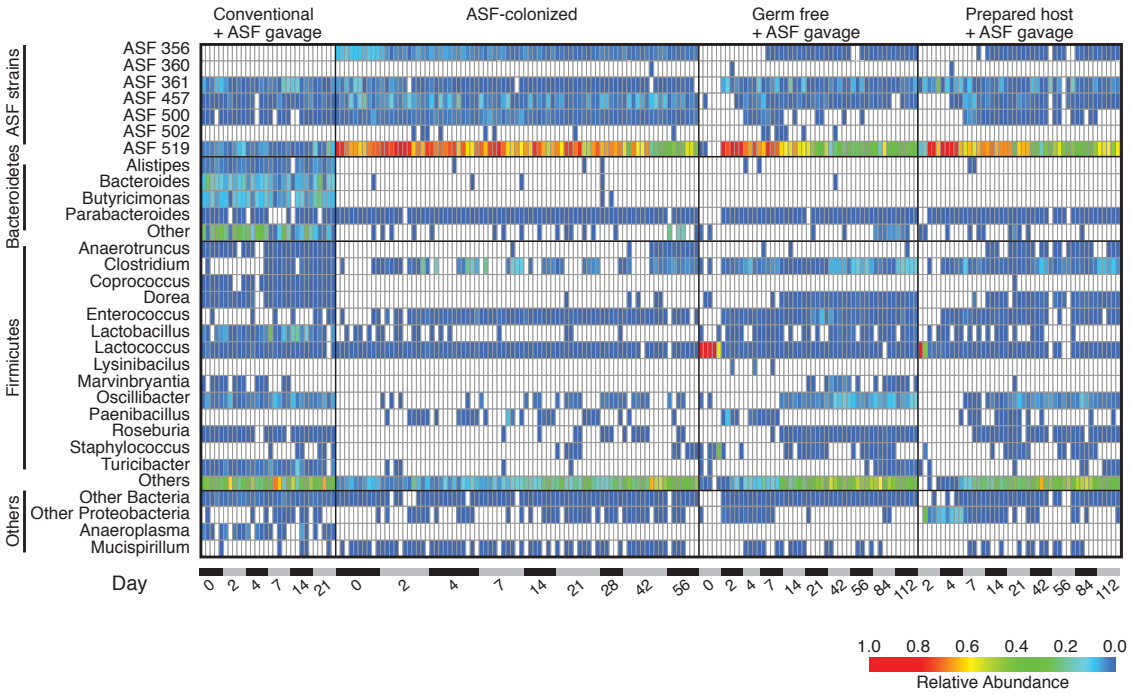


Figure 4.3. Development of a stable gut microbial community nucleated by inoculation with ASF. (A) Segmented regression analysis of communities in mice that were either germ-free or prepared conventional mice subjected to ASF gavage. The *y* axis shows the proportion of ASF lineages inferred from 16S rRNA gene tag pyrosequencing data. The *x* axis shows the number of days after transfer. Segmented regression analysis showed 2 phases, indicating a slow decline in the ASF proportion up to about day 30, followed by establish- ment of a new steady state consisting of approximately 40% ASF lineages. (B) Principal coordinates analysis (PCoA) ordination over time. Changes in community membership over time were analyzed using unweighted Unifrac (61). Progression of time is indicated by a gray arrow. (C) Shannon diversity of gut microbiota over time in the 4 hosts described in Figure 2. *n* = 5 per group for Conventional + ASF gavage, Germ-free + ASF gavage, and Prepared + ASF gavage. *n* = 10 for ASF-colonized.

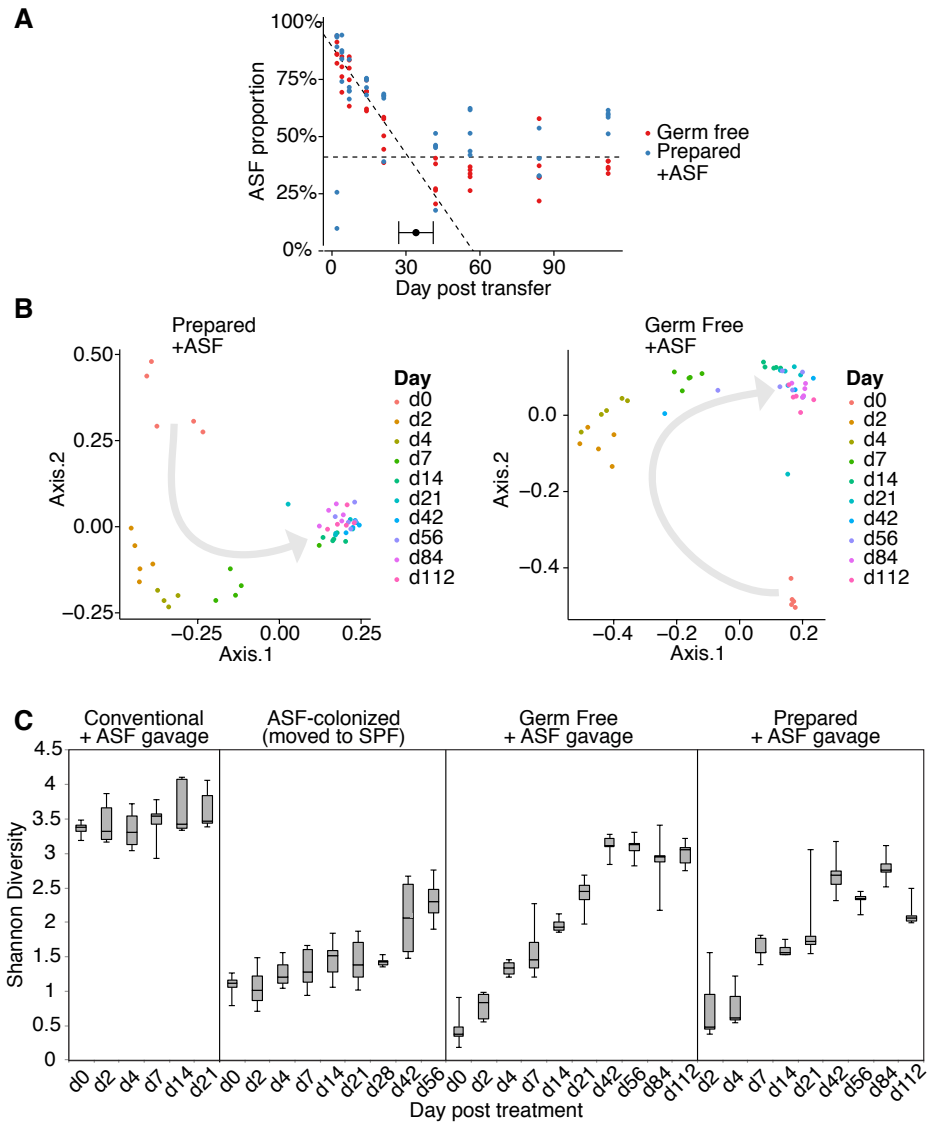


Figure 4.4. Comparison of non-ASF sequence reads from either Bacteroidetes or Firmicutes, illustrating selective repopulation with environmental Firmicutes. (A and B) Groups are color coded and represent change over time. *n* = 5 per group for Conventional + ASF gavage, Germ-free + ASF gavage, and Prepared + ASF gavage. *n* = 10 for ASF-colonized.

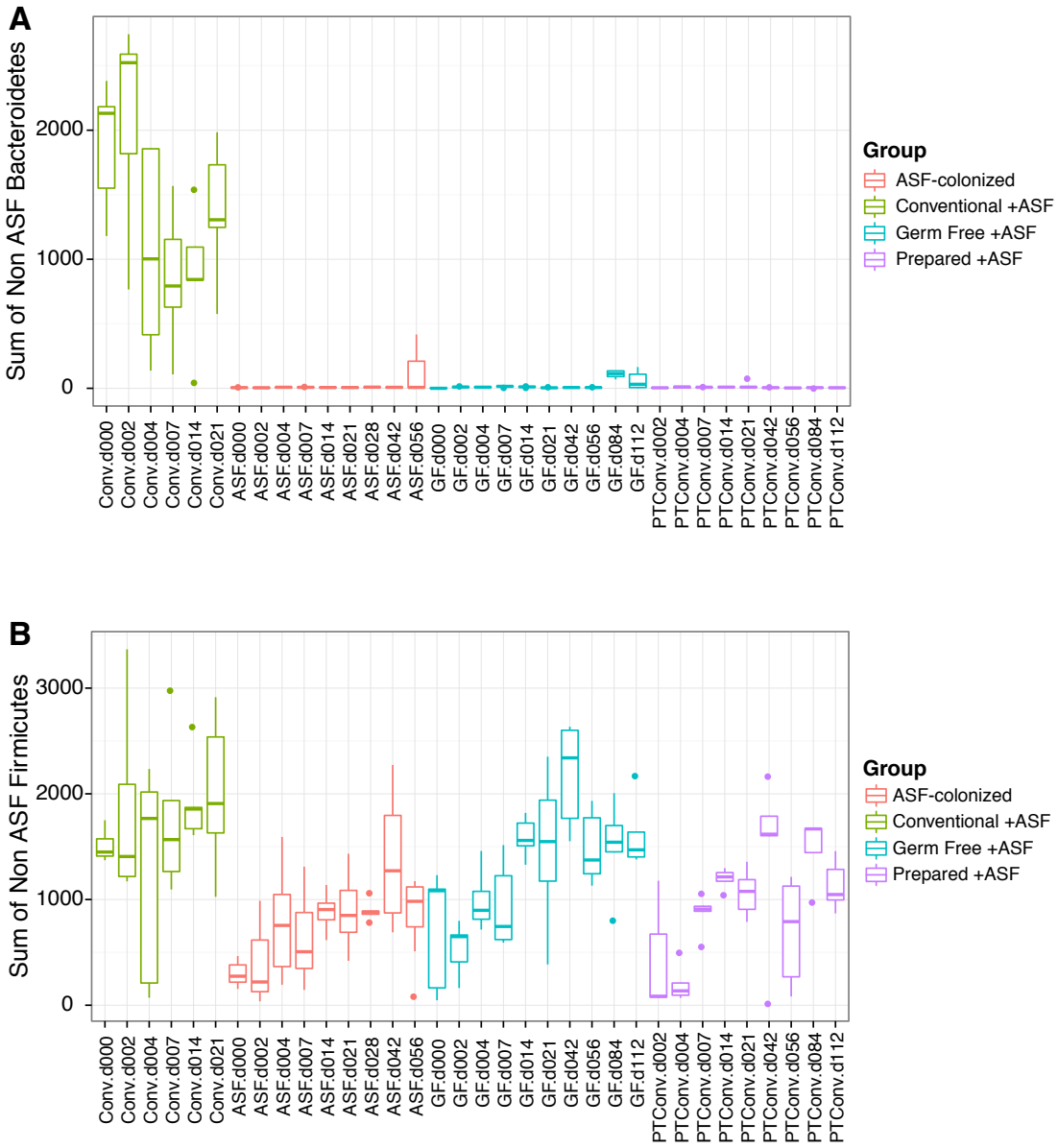


Figure 4.5. Transfer of ASF leads to a reduction in urease activity and fecal ammonia levels. (A) Urease activity in the feces of a conventionally housed mouse versus a mouse treated with ABX and a mouse colonized with ASF. (B) In vivo urease activity in conventionally housed (*n* = 5) and ASF-colonized mice (*n* = 5) quantified by the release of 13CO2 after i.v. injection of 13C-urea. (C) Fecal urease activity at the indicated time points after transplantation of ASF into prepared mice fed an irradiated diet (*n* = 3). (D) Fecal ammonia levels before and after transplantation of ASF into prepared mice fed a nonirradiated diet (*n* = 5). **P* < 0.01; ***P* < 0.001. Tukey's test for multiple comparisons.

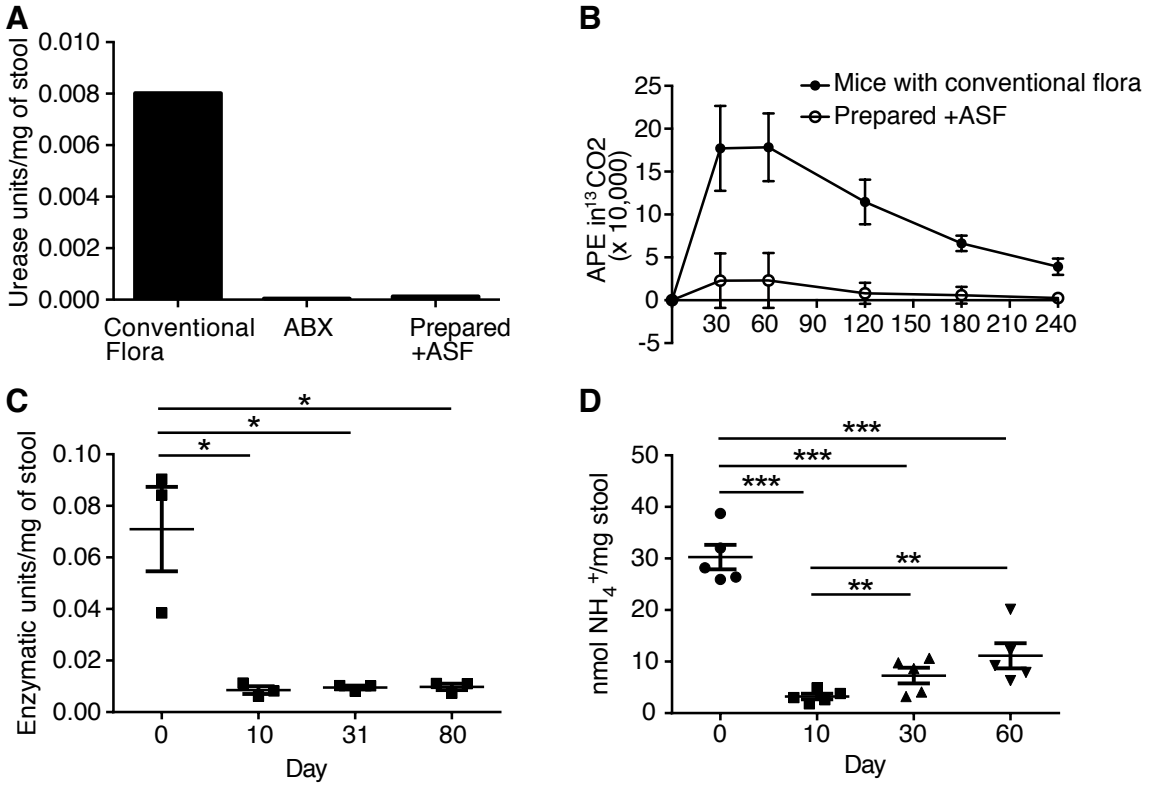


Figure 4.6. ASF transplantation into prepared mice reduces mortality after thioacetamide-induced hepatic injury and fibrosis. (A) Kaplan- Meier survival curves of high-dose TAA-induced acute hepatic injury in conventional versus prepared/ASF mice (*n* = 15 per group). (B) Kaplan- Meier survival curves of chronic, thrice weekly TAA administration at low, escalating doses, initiated 3 weeks after ASF transplantation (*n* = 15 per group). Survival curves were analyzed by the Kaplan-Meier method using the log-rank test.

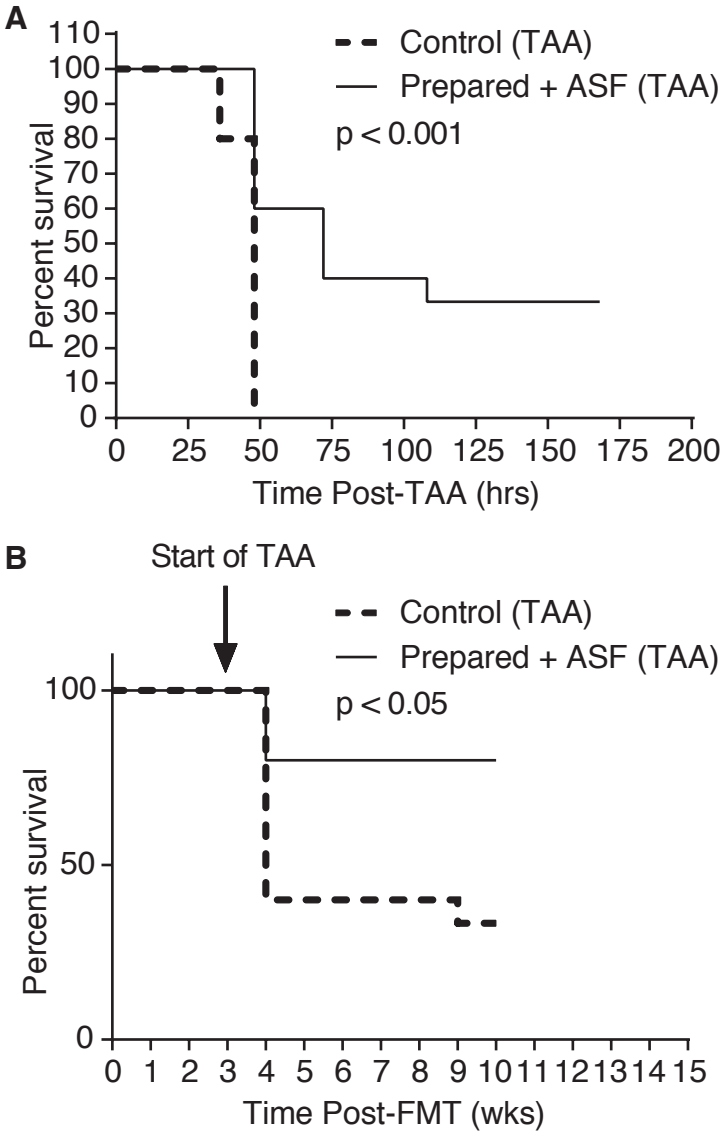
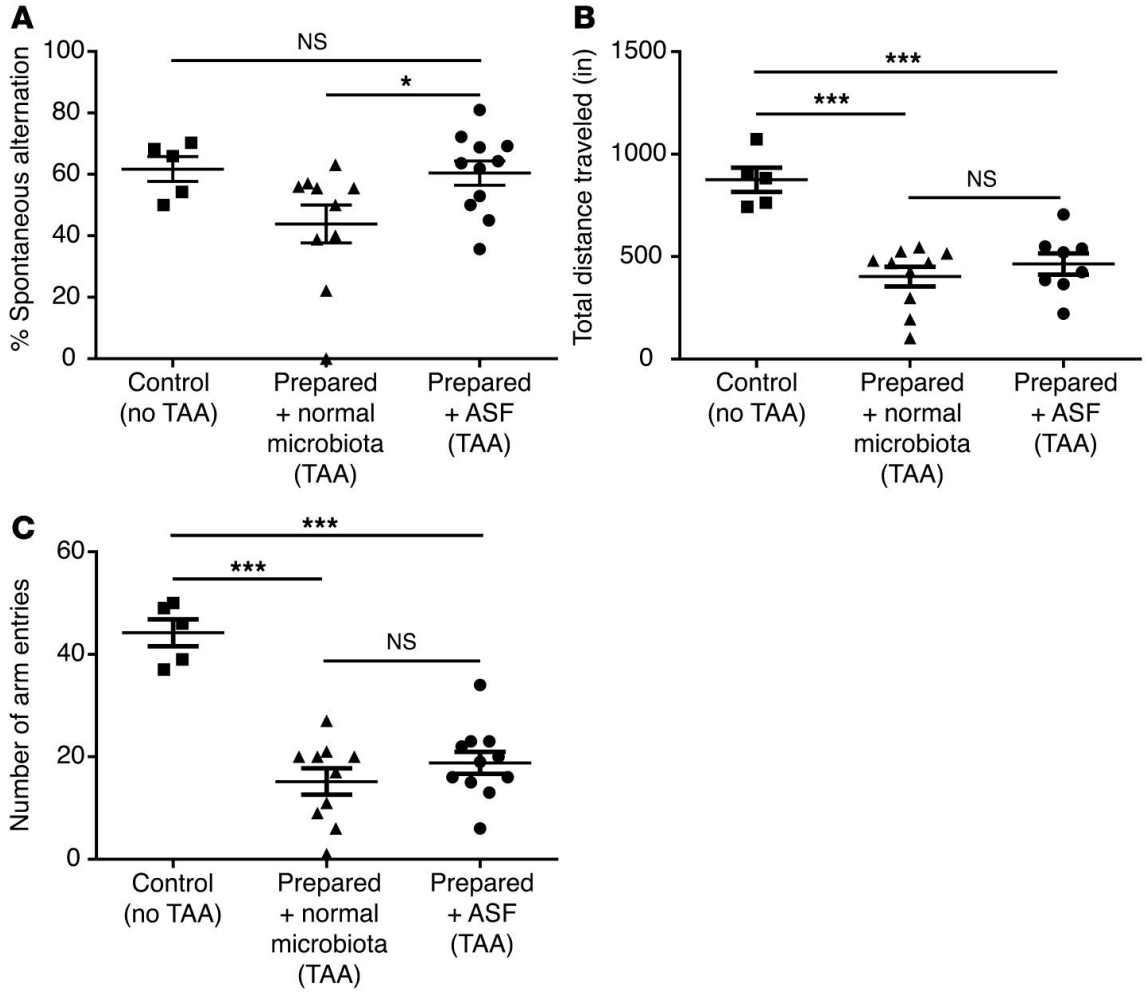


Figure 4.7. ASF transplantation into prepared mice restores cognitive, but not locomotor, deficits after thioacetamide-induced hepatic injury. (A) Spontaneous alternations after TAA treatment quantified by Y maze testing in prepared mice that were transplanted with either normal microbiota (n = 10) or ASF (n = 11) compared with untreated control mice (n = 5). *P < 0.05, by 2-tailed unpaired Student's t test; P = 0.04, by ANOVA. (B) Total distance traveled and (C) number of arm entries in the Y maze in prepared/ASF and prepared + normal microbiota mice after TAA treatment compared with untreated control mice. ***P < 0.001, by 2-tailed unpaired Student's t test; P < 0.001, by ANOVA.



References

1. Backhed F, et al. Defining a healthy human gut microbiome: current concepts, future directions, and clinical applications. *Cell Host Microbe*. 2012;12(5):611–622.
2. Lemon KP, Armitage GC, Relman DA, Fischbach MA. Microbiota-targeted therapies: an ecological perspective. *Sci Transl Med*. 2012;4(137):137rv135.
3. van Nood E, et al. Duodenal infusion of donor feces for recurrent Clostridium difficile. *N Engl J Med*. 2013;368(5):407–415.
4. Lawley TD, et al. Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing Clostridium difficile disease in mice. *PLoS Pathog*. 2012;8(10):e1002995.
5. Walser M, Bodenlos LJ. Urea metabolism in man. *J Clin Invest*. 1959;38:1617–1626.
6. Mobley HL, Hausinger RP. Microbial ureases: significance, regulation, and molecular characterization. *Microbiol Rev*. 1989;53(1):85–108.
7. Saudubray JM, Nassogne MC, de Lonlay P, Touati G. Clinical approach to inherited metabolic disorders in neonates: an overview. *Semin Neonatol*. 2002;7(1):3–15.
8. Riordan SM, Williams R. Treatment of hepatic encephalopathy. *N Engl J Med*. 1997;337(7):473–479.
9. Leitman P. Liver disease, aminoglycoside antibiotics and renal dysfunction. *Hepatology*. 1988;8(4):966–968.
10. Hampel H, Bynum G, Zamora E, El-Serag H. Risk factors for the development of renal dysfunction in hospitalized patients with cirrhosis. *Am J Gastroenterol*. 2001;96(7):2206–2210.
11. Bass N, et al. Rifaximin treatment in hepatic encephalopathy. *N Engl J Med*. 2010;362(12):1071–1081.
12. Mullen K, et al. Rifaximin is safe and well tolerated for long-term maintenance of remission from overt hepatic encephalopathy. *Clin Gastroenterol Hepatol*. 2014;12(8):1390–1397.
13. Bajaj JS, Sanyal AJ, Bell D, Gilles H, Heuman DM. Predictors of the recurrence of hepatic encephalopathy in lactulose-treated patients. *Aliment Pharmacol Ther*. 2010;31(9):1012–1017.
14. Nicaise C, et al. Control of acute, chronic, and constitutive hyperammonemia by wild-type and genetically engineered Lactobacillus plantarum
in rodents. *Hepatology*. 2008;48(4):1184–1192. 15. McGee RG, Bakens A, Wiley K, Riordan SM, Webster AC. Probiotics for patients with hepatic encephalopathy. *Cochrane Database Syst Rev*. 2011;(11):CD008716.
16. Lunia MK, Sharma BC, Sharma P, Sachdeva S, Srivastava S. Probiotics prevent hepatic encephalopathy in patients with cirrhosis: a randomized controlled trial. *Clin Gastroenterol Hepatol*. 2014;12(6):1003–1008.e1001.
17. Dewhirst FE, et al. Phylogeny of the defined murine microbiota: altered Schaedler flora. *Appl Environ Microbiol*. 1999;65(8):3287–3292.
18. Geuking MB, et al. Intestinal bacterial colonization induces mutualistic regulatory T cell responses. *Immunity*. 2011;34(5):794–806.
19. Willing BP, Vacharaksa A, Croxen M, Thanachayanont T, Finlay BB. Altering host resistance to infections through microbial transplantation. *PLoS One*. 2011;6(10):e26988.
20. Manichanh C, et al. Reshaping the gut microbiome with bacterial transplantation and antibiotic intake. *Genome Res*. 2010;20(10):1411–1419.
21. Wannemuehler MJ, Overstreet AM, Ward DV, Phillips GJ. Draft genome sequences of the altered schaedler flora, a defined bacterial community from gnotobiotic mice. *Genome Announc*. 2014;2(2):e00287-14.
22. Dollive S, et al. Fungi of the murine gut: episodic variation and proliferation during antibiotic treatment. *PLoS One*. 2013;8(8):e71806.
23. Hill DA, et al. Metagenomic analyses reveal antibiotic-induced temporal and spatial changes in intestinal microbiota with associated alterations in immune cell homeostasis. *Mucosal Immunol*. 2010;3(2):148–158.
24. Arumugam M, et al. Enterotypes of the human gut microbiome. *Nature*. 2011;473(7346):174–180.

25. Hildebrandt MA, et al. High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology*. 2009;137(5):1716–1724 e1711.
26. Singh R. Nutritional management of patients with urea cycle disorders. *J Inherit Metab Dis*. 2007;30(6):880–887.
27. Alexander T, Thomas K, Cherian A. Effect of three antibacterial drugs in lowering blood & stool ammonia production in hepatic encephalopathy. *Indian J Med Res*. 1992;96:292–296.
28. Zhao H, Wang H, Lu Z, Xu S. Intestinal microflora in patients with liver cirrhosis. *Chin J Dig Dis*. 2004;5(2):64–67.
29. James J, Ziparo V, Jeppsson B, Fischer J. Hyperammonaemia, plasma aminoacid imbalance, and blood-brain aminoacid transport: a unified theory of portal-systemic encephalopathy. *Lancet*. 1979;2(8146):772.
30. Fischer J, Baldessarini R. False neurotransmitters and hepatic failure. *Lancet*. 1971;2(7715):75–80.
31. Holecek M. Ammonia and amino acid profiles in liver cirrhosis: effects of variables leading to hepatic encephalopathy. *Nutrition*. 2015;31(1):14–20.
32. Miranda AS, et al. A thioacetamide-induced hepatic encephalopathy model in C57BL/6 mice: a behavioral and neurochemical study. *Arq Neuropsiquiatr*. 2010;68(4):597–602.
33. Ishikawa S, et al. CD1d-restricted natural killer T cells contribute to hepatic inflammation and fibrogenesis in mice. *J Hepatol*. 2011;54(6):1195–1204.
34. Elinav E, et al. Competitive inhibition of leptin signaling results in amelioration of liver fibrosis through modulation of stellate cell function. *Hepatology*. 2009;49(1):278–286.
35. Avraham Y, et al. Cannabidiol improves brain and liver function in a fulminant hepatic failureinduced model of hepatic encephalopathy in mice. *Br J Pharmacol*. 2011;162(7):1650–1658.
36. Hughes RN. The value of spontaneous alternation behavior (SAB) as a test of retention in pharmacological investigations of memory. *Neurosci Biobehav Rev*. 2004;28(5):497–505.
37. Hecht GA, et al. What is the value of a food and drug administration investigational new drug for fecal microbiota transplantation in Clostridium difficile infection? *Clin Gastroenterol Hepatol*. 2014;12(2):289–291.
38. Dominguez-Bello MG, Blaser MJ, Ley RE, Knight R. Development of the human gastrointestinal microbiota and insights from high-throughput sequencing. *Gastroenterology*. 2011;140(6):1713–1719.
39. Albenberg L, et al. Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota in humans and mice. *Gastroenterology*. 2014;147(5):1055-63.e8.
40. Human Microbiome Project Consortium. Structure, function diversity of the healthy human microbiome. *Nature*. 2012;486(7402):207–214.
41. Qin J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010;464(7285):59–65.
42. Momose Y, Park SH, Miyamoto Y, Itoh K. Design of species-specific oligonucleotide probes for the detection of Bacteroides and Parabacteroides by fluorescence in situ hybridization and their application to the analysis of mouse caecal Bacteroides-Parabacteroides microbiota. *J Appl Microbiol*. 2011;111(1):176–184.
43. Sakamoto M, Benno Y. Reclassification of Bacteroides distasonis, Bacteroides goldsteinii and Bacteroides merdae as Parabacteroides distasonis gen. nov., comb. nov., Parabacteroides goldsteinii comb. nov. and Parabacteroides merdae comb. nov. *Int J Syst Evol Microbiol*. 2006;56(Pt 7):1599–1605.
44. Smith MI, et al. Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. *Science*. 2013;339(6119):548–554.
45. Faust K, et al. Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol*. 2012;8(7):e1002606.
46. Wu GD, et al. Linking long-term dietary patterns with gut microbial enterotypes. *Science*. 2011;334(6052):105–108.
47. Freter R, Stauffer E, Cleven D, Holdeman LV, Moore WE. Continuous-flow cultures as in vitro models of the ecology of large intestinal flora. *Infect Immun*. 1983;39(2):666–675.

48. Lee SM, Donaldson GP, Mikulski Z, Boyajian S, Ley K, Mazmanian SK. Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature*. 2013;501(7467):426–429.
49. Eckburg PB, et al. Diversity of the human intestinal microbial flora. *Science*. 2005;308(5728):1635–1638.
50. Fischbach MA, Sonnenburg JL. Eating for two: how metabolism establishes interspecies interactions in the gut. *Cell Host Microbe*. 2011;10(4):336–347.
51. Torrallardona D, Harris C, Coates M, Fuller M. Microbial amino acid synthesis and utilization in rats: incorporation of 15N from 15NH4Cl into lysine in the tissues of germ-free conventional rats. *Br J Nutr*. 1996;75(5):689–700.
52. Belenguer A, Balcells J, Guada J, Decoux M, Milne E. Protein recycling in growing rabbits: contribution of microbial lysine to amino acid metabolism. *Br J Nutr*. 2005;94(5):763–770.
53. Metges C, et al. Incorporation of urea and ammonia nitrogen into ileal and fecal microbial proteins and plasma free amino acids in normal men and ileostomates. *Am J Clin Nutr*. 1999;70(6):1046–1058.
54. Bustamante J, et al. Prognostic significance of hepatic encephalopathy in patients with cirrhosis. *J Hepatol*. 1999;30(5):890–895.
55. Wu GD, et al. Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. *BMC Microbiol*. 2010;10:206.
56. Hoffmann C, et al. DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. *Nucleic acids research*. 2007;35(13):e91.
57. Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nature methods*. 2008;5(3):235–237.
58. Caporaso JG, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7(5):335–336.
59. Tuchman M, et al. N-carbamylglutamate markedly enhances ureagenesis in N-acetylglutamate deficiency and propionic acidemia as measured by isotopic incorporation and blood biomarkers. *Pediatr Res*. 2008;64(2):213–217.
60. Lalonde R. The neurobiological basis of spontaneous alternation. *Neurosci Biobehav Rev*. 2002;26(1):91–104.
61. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol*. 2005;71(12):8228–8235.

Acknowledgements

This work was supported by grants from the NIH (RO1-DK089472, to G.D. Wu; UH2/3-DK083981, to G.D. Wu, F.D. Bushman, and J.D. Lewis; and HD26979, to M. Yudkoff); the Molecular Biology Core and the Molecular Pathology Imaging Core of the Penn Cen- ter for Molecular Studies in Digestive and Liver Diseases (P30 DK050306); and the Joint Penn-CHOP Center for Digestive, Liver, and Pancreatic Medicine and the Metabolomic Core at the Children's Hospital of Philadelphia. Neurobehavior testing was conducted in collaboration with the Neurobehavior Testing Core at the University of Pennsylvania.

Contribution

KB and I performed the sequence analysis. I built figures 4.2, 4.3B, 4.3C and 4.4. I provided input regarding analysis strategy and helped author the manuscript.

Supplemental Information

Supplemental Figures and Tables

Supplemental figures and tables are available online: *Shen*, *Albenberg,* et al, JCI 2016.

CHAPTER 5: Dietary Regulation of the Gut Microbiota Engineered by a Minimal Defined Bacterial Consortium

The contents of this chapter will soon be published as:

> Shen T, **Chehoud C**, Ni J, Hsu E, Chen Y, Bailey A, Laughlin A, Bittinger K, Bushman F, Wu G. 2016. *Dietary Regulation of the Gut Microbiota Engineered by a Minimal Defined Bacterial Consortium*.

Abstract

We have recently reported that Altered Schaedler Flora (ASF) can be used to durably engineer the gut microbiota to reduce ammonia production as an effective modality to reduce morbidity and mortality in the setting of liver injury. Here we investigated the effects of a low protein diet on ASF colonization and its ability to engineer the microbiota. Initially, ASF inoculation was similar between mice fed a normal protein diet or low protein diet, but the outgrowth of gut microbiota differed over the ensuing month. Notable was the inability of the dominant *Parabacteroides* ASF taxon to exclude other taxa belonging to the Bacteroidetes phylum in the setting of a low protein diet. Instead, a poorly classified yet highly represented Bacteroidetes family, S24-7, returned within 4 weeks of inoculation in mice fed a low protein diet, demonstrating a reduction in ASF resilience in response to dietary stress. Nevertheless, fecal ammonia levels remained significantly lower than those observed in mice on the same low protein diet that received a transplant of normal feces. No deleterious effects were observed in host physiology due to ASF inoculation into mice on a low protein diet. In total, these results demonstrate that low protein diet can have a pronounced effect on engineering the gut microbiota but modulation of ammonia is preserved.

Introduction

The gut microbiota responds to multiple environmental stressors such as diet [1-3], antibiotic use [4], inflammation of the intestinal tract [5], and infection of the host with enteric

pathogens [6]. By studying the gut microbiota in pediatric patients with Crohn's disease, we have recently shown that the effects of these factors may be independent even if present simultaneously [7]. Amongst these, the impact of diet has received considerable attention as a potential modifiable factor that shapes the composition and/or function of the gut microbiota to prevent and/or treat disease [8]. The high-level efficacy of fecal microbiota transplantation (FMT) in the treatment of *Clostridium difficile* infections (CDI) is proof of concept that inoculating a host with a consortium of microbes has a meaningful effect on the composition of the gut microbiota [9]. The use of feces could be considered an untargeted approach with potential risks [10], but growing evidence suggests that the use of defined microbial consortia could be developed to treat disease [11, 12]. We have recently shown that the gut microbiota can be durably reconfigured to reduce fecal urease activity and ammonia production through oral inoculation of Altered Schaedler Flora (ASF), a defined microbial consortium that contains minimal urease gene content [12]. ASF comprises 8 murine gut commensal bacterial strains assembled in the 1970s and standardized by the National Cancer Institute in 1978 [13]. It is now commonly used to create gnotobiotic mice and/or to enhance the health of immunodeficient mouse strains.

Examples of co-metabolism between the gut microbiota and its mammalian host requiring host-derived substances include bile acids, mucous, and urea. The latter is particularly important for nitrogen flux between the host and the gut microbiota [14, 15]. As the primary source of nitrogen, dietary protein is essential to the synthesis of nucleic acids, amino acids, and other nitrogenous compounds. The catabolism of dietary protein by the host leads to hepatic formation of urea, a nitrogenous waste product that is excreted through the urine or delivered into the colon, where hydrolysis by bacterial urease results in the production of carbon dioxide and ammonia. Ammonia is a shared substrate for the synthesis of proteins, amino acids, and other small molecules by both the host and its microbiota. Although generally thought to be nutritionally beneficial to the host by enhancing nitrogen recycling, the production of ammonia by the gut microbiota can have deleterious effects in the setting of altered hepatic function, resulting in the

development of neurotoxicity [16-18]. Under such conditions, a low protein diet (LPD) can be used to reduce systemic ammonia levels [19, 20].

By inoculating mice with ASF after the endogenous microbiota has been reduced through the use of antibiotics and polyethylene glycol (PEG), the composition of the gut microbiota can be durably modified in composition as well as function. Functionally, there was a long-lasting reduction in fecal ammonia that was effective in reducing morbidity and mortality in the thioacetamide model of liver injury [12]. Since (1) the absorption of fecal ammonia produced by the gut microbiota may be an important source of nitrogen for the host especially in the setting of dietary protein restriction [14], and (2) low protein diets are used clinically in patients with hyperammonemic inborn errors of metabolism [21], there are a number of questions about the impact of diet on the engineering of the gut microbiota to reduce ammonia production. What is the effect of a LPD on the ability of a defined bacterial consortium to engraft into the gut? Does a LPD have an effect on the composition of the engineered microbiota? Will the ammonia reduction by microbiota engineering be sustained and exhibit lower levels than those achievable by a LPD alone? And lastly, will a significant reduction in gut microbiota ammonia production be deleterious to the host on a LPD?

Here, we address these questions by inoculating mice on a LPD with either feces from conventionally-reared mice (Normal Feces, or NF) or with ASF, monitoring the resultant composition of the gut microbiota over time by 16S tagged sequencing, assessing functionality by quantifying fecal ammonia levels, and investigating the impact on the host by metabolic profiling. Although a LPD has no effect on the ability of ASF to engraft into the gut of the host upon inoculation, the resultant engineered state of the microbiota is altered primarily due to the re-emergence of S24-7, a specific bacterial taxonomic family within the Bacteroidetes phylum. Despite this alteration, fecal ammonia levels remain diminished and without consequence to the metabolic physiology of the host on a LPD.

Results

LPD impacts host physiology and nitrogen metabolism but modestly alters the composition of the gut microbiota

We first set out to investigate the effects of a LPD on both the murine host and the gut microbiota. Fifteen adult female C57BL6J mice were placed on an open source irradiated purified rodent diet containing normal amount of dietary protein at 21% by kilocalories (AIN-76A), henceforth referred to as normal protein diet (NPD, Sup. Table 5.1) for one week upon arrival into the University of Pennsylvania SPF vivarium. Subsequently, ten of the 15 mice were switched to an irradiated low protein diet (LPD, Sup. Table 5.1) formulated using AIN-76A as the base that contains 3% protein by kilocalories. The LPD was made isocaloric by proportionally increasing carbohydrate content while keeping fat content unchanged. The remaining 5 mice continued to be fed the NPD. We monitored physiological changes in these mice using body weight and food intake measurements as well as body composition determination via nuclear magnetic resonance (NMR) imaging. We found that compared to NPD-fed mice, LPD-fed mice exhibited poor weight gain despite equivalent caloric consumption (Figure 5.1A). LPD-fed mice also demonstrated increased fat mass and decreased lean mass (Figure 5.1B and 5.1C). Corresponding to a reduction in serum urea level compared to NPD-fed mice (Figure 5.1D), LPD-fed mice exhibited significant reductions in fecal urea and fecal ammonia levels after ten weeks on the LPD (Figure 5.1E and 5.1F). These results are consistent with the fundamental role that dietary protein plays in host nitrogen balance. Reduction in dietary protein may have an effect on the gut microbiota by reducing the delivery of urea to the colonic environment leading to the reduction in fecal ammonia levels.

16S tagged sequencing (1045 to 5305 reads per sample, median = 3448 reads) revealed modest yet distinct differences in the composition of the fecal microbiota in mice after placement on the LPD. These difference can be visualized in principal coordinates analysis of weighted (Figure 5.2A) and unweighted (Figure 5.2B) UniFrac distance. The LPD led to a significant

increase in the diversity of the gut microbial community as assessed by the Shannon diversity index (Figure 5.2C and Sup. Figure 5.1). The LPD also led to significant increases in the relative abundance of Mollicutes and Coriobacteria and a decrease in the Firmicutes classes Erysipelotrichi and Clostridia (Figure 5.2D and 5.2E).

LPD has no effect on the initial engraftment of ASF into the host microbiota

We have previously shown that there is a reduction in gut bacterial biomass upon treatment of mice with oral antibiotics (vancomycin and neomycin) and PEG, thus permitting the engraftment of ASF upon inoculation by oral gavage [12]. However, the effect of a LPD on the engraftment of ASF into the gut of a prepared host remains unknown. After preparation with antibiotics and PEG, we orally inoculated five of the LPD-fed mice with ASF (herein referred to as "ASF-transplanted"). As a control group, we transplanted the other five LPD-fed mice using feces from conventionally-reared donor mice (herein referred to "NF-transplanted," for Normal Feces). Using 16S rRNA tagged sequencing, we tracked taxonomic alterations in the gut microbiota over time. We found that NF-transplanted mice exhibited minimal change in the composition of their gut microbiota (Sup. Figure 5.2). However, the gut microbiota of ASF-transplanted mice underwent a shift in composition in a similar fashion to that previously observed in NPD-fed mice transplanted with ASF [12], as shown in Figure 5.3A and 5.3B. In particular, the shifts along PC1 in both cohorts of mice represent changes due to the initial ASF inoculum (compare days 0 to 14 in both groups), whereas differences between the two cohorts of mice along PC2 may represent the effect of diet. These findings suggest that a LPD does not affect the initial engraftment of ASF into the host microbiota.

Diet affects the resilience of the gut microbiota engineered by inoculation with ASF

By tracking the composition of mice inoculated with ASF, we determined the effect of a LPD on the ability of ASF to engineer a different microbiota composition. In the setting of a NPD, we previously observed that ASF transplantation led to the development of a new steady state

community after one month composed of both ASF and the return of selected taxa of the Firmicutes phylum, but no non-ASF Bacteroidetes [12]. We proposed that *Parabacteroides* ASF519, the dominant taxon in ASF in feces, may have prevented the return of other Bacteroidetes taxa by competitive niche exclusion. Tracking compositional changes in the gut microbiota over time, we found that in the setting of the LPD, the gut microbiota engineered by ASF inoculation developed into an alternative rich community with diversity similar to that on the NPD (Figure 5.3C). However, ASF519 did not suppress the return of other Bacteroidetes. Instead, Bacteroidetes S24-7, a poorly classified yet common bacterial taxon in the commensal murine gut microbiota [22, 23], returned after ASF transplantation and reached an equilibrium state with ASF519 (Figure 5.4 and Sup. Figure 5.3). We plotted the progression of the transplanted ASF community over time. We found that ASF reached a new steady state in the setting of the LPD at around 4 weeks after transplantation, similar to what we previously observed in the setting of the NPD [12] (Figure 5.3A and 5.3B). However, this steady state more closely resembled the endogenous microbiota, likely as a result of the return of S24-7 on the LPD (best observed in Figure 5.3A along PC1 – compare the solid to dotted grey line). Overall, these findings suggest that ability of ASF lineages to compete is reduced in the presence of a LPD.

The ASF-engineered gut microbiota lowers fecal ammonia more effectively than LPD alone

We have previously shown that ASF transplantation durably reduces fecal ammonia by decreasing fecal urease activity [12]. Since a LPD itself mainly reduces fecal ammonia levels by decreasing the delivery of urea to the colon (Figure 5.1E and 5.1F), we sought to determine whether the ASF-engineered microbiota would be able to reduce fecal ammonia levels below those achieved by a LPD alone. We measured fecal urea and fecal ammonia levels in mice at baseline on the NPD, after ten weeks on the LPD, and compared ASF and NF transplantation on the LPD. As shown in Figure 5.5A, after the initial reduction in fecal ammonia levels induced by the LPD, ASF transplantation reduced fecal ammonia further than did NF transplantation. The ability of the ASF-engineered microbiota to lower fecal ammonia levels below those achieved by

the LPD alone is likely due to the reduction in fecal urease activity since there was no difference in fecal urea levels after NF and ASF transplantation (Figure 5.5B). These results indicate that the functionality of the ASF-engineered gut microbiota is not significantly altered in the setting of a LPD despite alterations in its composition.

The low fecal urease and fecal ammonia-producing microbiota engineered by ASF inoculation does not exacerbate host metabolic dysfunction induced by LPD

Urea is a nitrogenous waste product, but it is thought to contribute to host nutrition via urea nitrogen recycling by intestinal bacterial urease in both ruminants and non-ruminants, leading to microbial and/or host synthesis of peptides, amino acids, and other small molecules [14]. We asked whether this role of urea may become important for host physiology in the setting of a LPD, where systemic nitrogen is reduced. After transplanting the above cohort of mice with ASF or NF, we continued to monitor their body weight, food intake, and survival. Remarkably, despite the absence of colonic urea nitrogen recycling, ASF-transplanted mice did not differ significantly from NF-transplanted mice (Figure 5.5C, 5.5D, and Sup. Figure 5.4). Thus, in the setting of a LPD, ASF transplantation does not lead to significant detrimental changes to host physiology and metabolism.

Discussion

The success of FMT in the treatment of recurrent *Clostridium difficile* infections provides proof of concept that the gut microbiota can be a target for the treatment of disease in humans. The use of fecal transfer will likely be replaced by the use of defined microbial consortia with specific biological properties. As proof of concept, we have shown in murine models that a defined consortium of eight bacteria, known as Altered Schaedler Flora (ASF), can be used to engineer the gut microbiota with altered functionality, namely a reduction in fecal urease activity and ammonia production [12]. Critical to the success of this strategy is the substantial reduction in the biomass of the baseline microbiota to provide a niche into which the bacterial inoculum can engraft.

An important consideration for engineering the gut microbiota is resilience to environmental stress. An optimally engineered microbiota would be a rich community that stays intact in the presence of environmental stress. Diet is an important environmental stressor on the gut microbiota that should be considered when engineering gut microbial communities. As one example, the low ammonia-producing microbiota engineered by ASF, which has functional durability for several months in mice fed an irradiated diet, shows reduced resilience when the mice are fed a non-irradiated diet [12]. We chose to study the impact of dietary protein on the resilience of the ASF-engineered microbiota for several reasons: 1) Dietary protein has been shown to influence the composition of the gut microbiota in gnotobiotic mice [24]; 2) Protein consumption regulates the production of hepatic urea that may affect colonic urea delivery to the gut microbiota [14, 25]; 3) Protein-restricted diets are an important therapeutic modality for patients with hyperammonemic inborn errors of metabolism [20, 26].

Unlike the modulation of fat and fiber in mice, which have been shown to have a strong effect on the composition of the murine gut microbiota [1, 24], we show that severe restriction of dietary protein had a modest effect. Within the Firmicutes phylum the Clostridia and Erysipelotrichi classes decreased significantly on a LPD, consistent with the preference of taxa within Firmicutes, particularly Clostridia species, to metabolize amino acids and peptides [27, 28]. Alternatively, since we show that a LPD reduces serum urea levels with reduced delivery to the colon resulting in lower fecal ammonia levels, an alteration in nitrogen flux via ammonia into the gut microbiota [29, 30] may also have an effect on the composition of the bacterial microbiota.

Since we balanced protein with carbohydrate in the composition of the purified rodent diet, it is difficult to ascertain if the differences in the composition of the gut microbiota are due primarily to alterations in protein or in carbohydrate. We found that two bacterial phyla present at low abundances increased significantly on a LPD. Specifically, the classes Mollicutes (Tenericutes phylum) and Coriobacteria (Actinobacteria phylum) increased on a LPD. Previous work has shown that Mollicutes proliferated on a typical Western diet characterized by high-

fat/high-sugar content, likely because of their ability to import and process simple sugars [31]. Thus, an increase in the abundance of Mollicutes that we observed on a LPD could be due to the increase in carbohydrate content rather than the reduction in protein content. Another study also showed that gut colonization by Actinobacteria and Tenericutes was strongly correlated with decreased hepatic levels of glycogen and glucose [32], further suggesting the interplay between the host and these two phyla may be closely related to carbohydrate metabolism.

Despite the effect of a LPD on the composition of the gut microbiota at baseline, this did not have an effect on the initial engraftment of ASF at 2 weeks, demonstrating that the use of antibiotics and PEG effectively prepared the environment of the gut for inoculation by a minimal defined bacterial consortium. Subsequently, development of the resultant engineered microbiota, determined by emergence of various bacterial taxa in addition to ASF, was distinctly different in mice fed a NPD versus a LPD. On a NPD, we previously showed that the dominant taxon *Parabacteroides* (ASF519) was able to exclude the entire Bacteroidetes phylum yet permit the reappearance of specific taxa belonging to the Firmicutes phylum. The observation that bacterial lineages with similar phylogeny exhibit competitive niche exclusion has been demonstrated in the *Bacteroides* genus where successful competition for carbohydrate substrates plays an important role [33]. By contrast, on a LPD, the resultant engineered microbiota appears to be more similar to baseline primarily due to the reemergence of a single bacterial taxon belonging to the Bacteroides family, S24-7.

S24-7 has been previously recognized as a dominant taxonomic group in the murine microbiota. It was first characterized by Salzman *et al.*, who referred to the taxon as “mouse intestinal bacteria” [22]. The S24-7 taxon is phylogenetically distinct from other named genera in the order Bacteroidales. The taxon has been reported as altered in several recent mouse studies: it was increased in proportion following partial hepatectomy [34], associated with co-infection by *Hymenolepis* spp. [35], and decreased in proportion following antibiotic treatment for parenteral nutrition-associated liver injury [36]. However, to our knowledge, no study has previously

characterized competition between S24-7 and other Bacteroidetes species in mice. The S24-7 taxon is typically encountered at very low (<1%) abundance in fecal samples from human populations. However, one study of a previously uncontacted Amerindian population reported the taxon to be enriched in isolated Yanomami Amerindians relative to Guahibo Amerindian, Malawian, and U.S. subjects [37]. The average abundance of the taxon in Yanomami Amerindians was reported to be nearly 5% of total bacteria, suggesting a potential role for S24-7 in the human gut.

Upon reduction of bacterial biomass through a combination of antibiotics and polyethylene glycol (PEG), S24-7 is no longer detectable and shows no return over time after mice have been inoculated with ASF. Since both *Paracteroides* (ASF519) and S24-7 are closely related within the Bacteroidetes phylum, we speculate that S24-7 may be co-excluded from the luminal gut environment by ASF519 through competitive niche exclusion, a mechanism that has been hypothesized as the basis for the inversely-related proportions of Bacteroides and Prevotella in the human gut microbiota [2, 38], a predominant feature of “enterotypes” [39]. From a mechanistic standpoint, the basis of competitive niche exclusion may involve the competition of metabolic substrates as has been demonstrated for Bacteroides species in a reductionist model system [33]. Since S24-7 reappears and co-exists at approximately equal levels with ASF519 in LPD-fed mice, the alteration of substrate availability via diet may have altered the luminal environment of the gut that reduces the need for competition between these two taxa. For example, a LPD may have altered the balance of nitrogen flux into the gut microbiota via the uptake of ammonia. Indeed, despite the return of S24-7 and the similarities between the composition of a gut microbiota of a conventionally-housed mouse and the ASF-engineered community established in LPD-fed mice, fecal ammonia levels remained much lower in ASF-transplanted mice that those transplanted with normal feces. This suggests that S24-7 may be urease negative. Further elucidation of such mechanism(s) will require genomic characterization of S24-7 along with an evaluation of its biological properties.

The quantification of fecal ammonia was used to determine the impact of microbiota composition on the function of the community. Despite the modest alterations in the gut microbiota induced by the consumption of a LPD, there was a significant reduction in fecal ammonia levels reflecting the reduced abundance of urea substrate available for hydrolysis by the gut microbiota. This observation emphasizes the notion that diet may have an indirect impact on the gut microbiota by alteration of the host similar to the outgrowth of a pathobiont due to the enhanced production of sulfated bile acids in mice fed milk fat [40]. Importantly, engineering of the gut microbiota using ASF led to a reduction in fecal ammonia levels significantly greater than that observed on a LPD.

Since ammonia, produced by the gut microbiota via urease activity, is absorbed by the host where it can be used for amino acid synthesis, it has been hypothesized that this form of nitrogen recycling may be important for host health especially under conditions of limited protein intake [14, 41]. This might be a significant limitation of a strategy focused on reducing gut microbiota ammonia production for the treatment of hyperammonemia and hepatic encephalopathy [12]. Although LPD-fed mice did not exhibit growth, as would be expected, ASF transplantation with subsequent robust reduction of fecal ammonia levels did not lead to any effects on food intake, growth, or mortality relative to LPD-fed mice transplanted with normal feces who had much higher levels of fecal ammonia. Since patients with hyperammonemic inborn errors of metabolism are placed on a LPD to prevent metabolic crises, our observations provide preliminary evidence that the engineering of gut microbiota to reduce fecal ammonia production may be well tolerated in this patient population. However, additional safety studies are needed.

In summary, we show that diet has a significant effect on the ability of a defined microbial consortium to engineer the composition of the gut microbiota. Specifically, LPD alters the co-exclusion of two dominant taxa within the Bacteroidetes phylum. Given the alterations in the syntropic host-microbiota interactions in nitrogen flux that occur in the levels of urea delivery from the host to the gut microbiota, the reduced production of ammonia via bacterial urease, and the

uptake of ammonia by both the host and the gut microbiota, a LPD may be a particularly important environmental stressor that will impact upon the composition of an engineered microbiota. Nevertheless, the functionality of the engineered gut microbiota, as quantified by a reduction in fecal ammonia levels, remained intact. Together with the absence of detrimental effects on host physiology in the setting of a LPD, the reduction in fecal ammonia levels via engineering of the gut microbiota may be an effective therapeutic strategy for patients with hyperammonemic inborn errors of metabolism.

Methods

Animals

C57B6J female mice 8 to 12 weeks old (The Jackson Laboratory) were used in this study. Fecal pellets collected from ASF-colonized CB17 SCID mice (Taconic) served as the source of the ASF innoculum whereas conventionally-colonized C57B6J mice (The Jackson Laboratory) served as the source of the normal flora (NF) innoculum used in the FMT procedures as previously described [12]. Fecal homogenates were prepared by diluting 0.1 g feces 10-fold in PBS. Mice were prepared for FMT by oral delivery of antibiotics in drinking water (1.125 g aspartame, 0.15 g vancomycin, and 0.3 g neomycin in 300 mL sterile water) for 72 hours. During the final 12 hours, the water supply was exchanged with a 10% PEG solution (Merck), and the mice were fasted. Mice were then inoculated daily with fecal homogenates by oral gavage for 7 days. All mice were housed five per cage in a conventional specific-pathogen free (SPF) facility (and transferred from one conventional facility to another conventional facility within the University of Pennsylvania 10 weeks after the start of experiment for NMR imaging) and fed irradiated AIN-76A (Research Diets D10001, 21% protein by kilocalories – NPD, see Sup. Table 5.1). After one week, ten mice were switched to irradiated AIN-76A with lower protein content (Research Diets D08092201, 3% protein by kilocalories – LPD, see Sup. Table 5.1). Fecal pellets were collected at baseline on NPD, 10 weeks after placement on LPD, 2 weeks after FMT (15 weeks on LPD), 4 weeks after FMT (17 weeks on LPD), and 10 weeks after FMT (23 weeks on LPD) for bacterial

taxonomic and biochemical analyses. Fecal pellets were collected in 1.5 mL microcentrifuge tubes (Sigma-Aldrich) and immediately placed on dry ice then stored in -80°C freezer until time of analysis. Body composition was determined using NMR imaging via the Mouse Phenotyping, Physiology and Metabolism Core at the University of Pennsylvania. All animal studies were performed with the approval of the Institutional Animal Care and Use Committee of the University of Pennsylvania (Protocol Number: 803408).

16S V1-V2 sequencing

DNA was isolated from stool as previously described [2, 42]. 100 ng of DNA was amplified with barcoded primers annealing to the V1-V2 region of the 16S rRNA gene (forward primer, 5′-AGAGTTTGATCCTGGCTCAG-3′; reverse primer, 5′-CTGCTGCCTYCCGTA-3′; [43, 44] using AccuPrime Taq DNA Polymerase System with Buffer 2 (Life Technologies). PCR reactions were performed on a thermocycler using the following conditions: initiation at 95°C for 5 min followed by 20 cycles of 95°C × 30 s, 56°C × 30 s, and 72°C × 1 min 30 s, then a final extension step at 72°C for 8 min. The amplicons from each DNA sample, which was amplified in quadruplicate, were pooled and purified with Agencourt AMPure XP beads (Beckman Coulter) following the manufacturer's instructions. Purified DNA samples were then sequenced using the 454/Roche GS FLX Titanium chemistry (454 Life Sciences).

16S rRNA gene sequence analysis

16S rRNA gene sequence data was processed with QIIME v 1.8.0 [45] using default parameters. Sequences were clustered into operational taxonomic units (OTUs) at 97% similarity and then assigned Greengenes taxonomy [46] using the uclust consensus taxonomy classifier. Sequences were aligned using PyNAST [47] and a phylogenetic tree was constructed using FastTree [48]. Weighted and unweighted UniFrac [49] distances were calculated for each pair of samples for assessment of community similarity and generation of principal coordinate analysis (PCoA) plots. Statistical analyses for bacterial abundance difference was performed using non-

parametric Wilcoxon test, and p-values were corrected for multiple comparisons using the Benjamini and Hochberg procedure.

Measurement of fecal ammonia

Fecal ammonia concentrations were determined using an Ammonia Assay Kit (ab83360, Abcam, Cambridge, MA). Fecal pellets were suspended in the assay buffer provided at a concentration of 1 mg/10 uL, homogenized, and centrifuged at 13,000 x g for 10 minutes at room temperature to remove insoluble material. Ammonia concentration was then determined according to the kit protocol.

Measurement of serum and fecal urea

Urea concentrations were determined using the QuantiChrom™ Urea Assay Kit (DIUR-500, Bioassay Systems, Hayward, CA). Serum samples were assayed directly. Fecal pellets were suspended in ddH_2O at a concentration of 1 mg/10uL, homogenized, and centrifuged at 2,500 x g for 10 minutes at room temperature to remove insoluble material. Urea concentration was then determined according to the kit protocol.

Figures

Figure 5.1. Changes in murine physiology and nitrogen metabolism on a LPD. Differences in (A) body weight (n=5 in NPD group, n=10 in LPD group), (B) % fat mass, (C) % lean mass, and (D) serum urea levels between NPD-fed and LPD-fed mice. (E) Fecal urea and (F) fecal ammonia levels at baseline on the NPD and after placement on the LPD. Values represent mean ± SEM. Statistical significance in body weight determined by two-way ANOVA with repeated measures; statistical significance in other parameters determined by paired and unpaired two-tailed Student's t test. *p<0.05, **p<0.01, ***p<0.001.

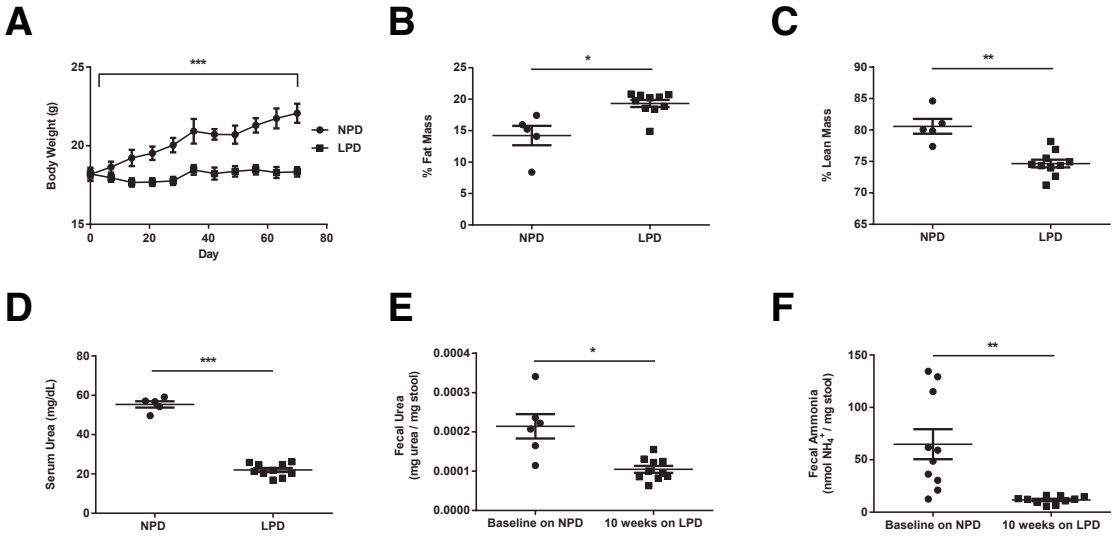


Figure 5.2. Effect of a LPD on the composition of the gut microbiota. Principal coordinates analysis (PCoA) ordination of mice before and after placement on the LPD for 10 weeks. Changes in community membership were analyzed using (A) weighted and (B) unweighted Unifrac. (C) The interquartile range of Shannon diversity values is shown for mice on the NPD who were later put on the LPD (Wilcoxon rank sum test p-value = 0.0001299). (D) Heatmap showing the relative abundance of bacterial lineages over time in mice who were on the NPD at baseline and then after ten weeks on the LPD. Rows indicate bacterial lineages annotated at the class taxonomic level on the left. The color key on the right of the figure indicates relative abundance. Columns summarize the sequencing results from individual fecal specimens. Each column represents a different mouse. The columns are grouped by diet. (E) Bacterial lineages that change on the LPD. Four bacterial classes significantly differed between the NPD and LPD (FDR-corrected Wilcoxon test p-value < 0.05). Relative abundance of each class in both diet groups is shown. Box and whiskers show the interquartile range; black circles mark the outlier samples.

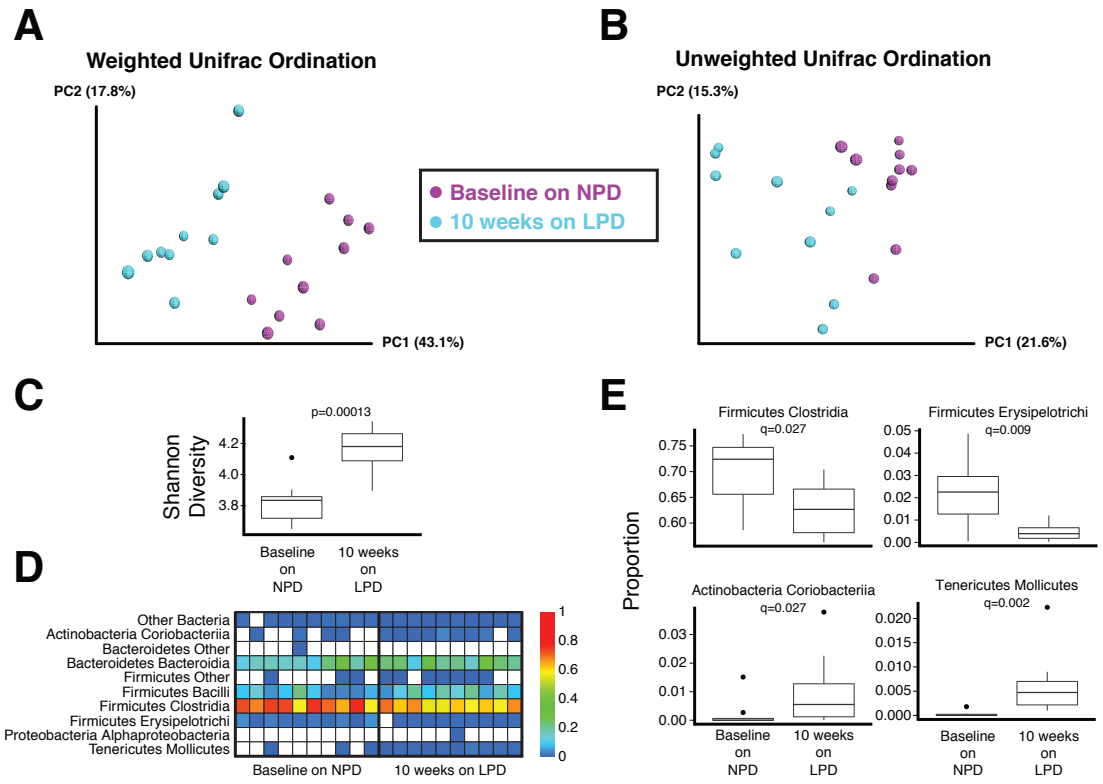


Figure 5.3. Effect of a LPD on the initial engraftment of ASF and subsequent resilience over time. Principal coordinates analysis (PCoA) ordination of mice after transplantation with ASF. Changes in community membership were analyzed using (A) weighted and (B) unweighted Unifrac. Dietary groups are color coded as indicated with the shades of colors indicating progression in time. Day 0 samples have gray circles around them (solid for NPD, dashed for LPD). The arrows were added to help visualize the progression of time after ASF transplantation. (C) The interquartile range of Shannon diversity values are shown for mice on the NPD and LPD inoculated with either ASF or Normal Feces (NF). Black circles mark the outlier samples.

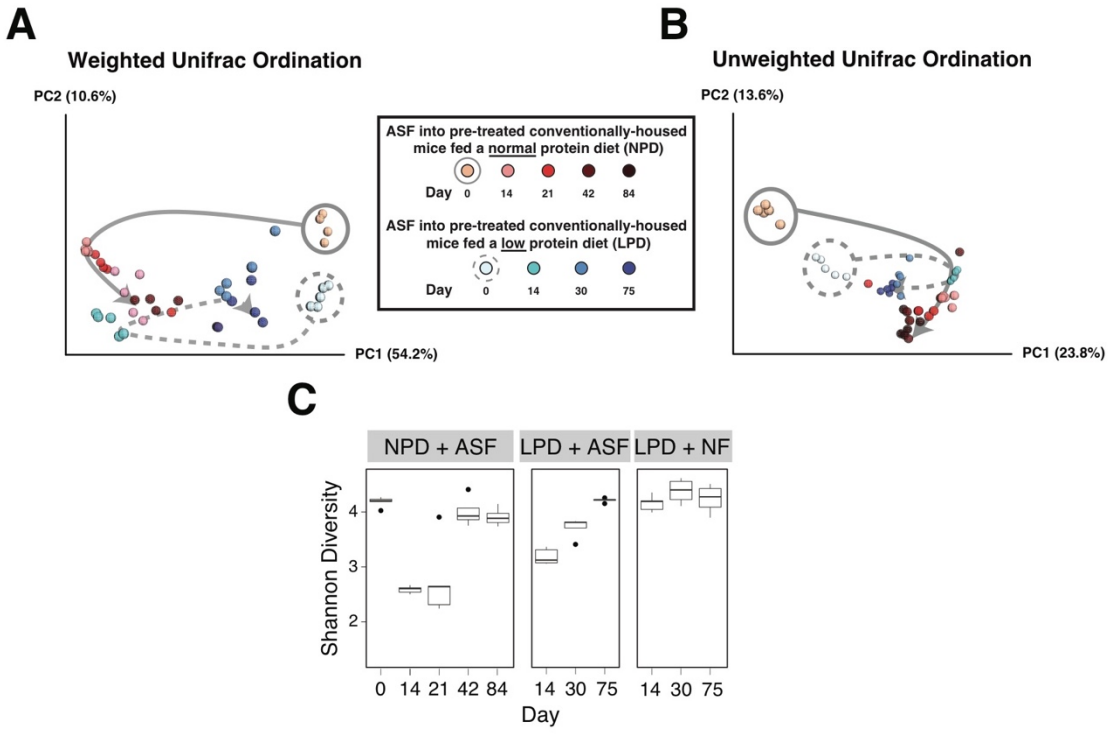


Figure 5.4. S24-7 returns after ASF transplantation into mice on a LPD but not on a NPD. (A) Relative abundance of bacterial taxa are shown. Each column represents a single sample of a pre-treated, ASF-inoculated mice on the LPD (LPD + ASF). Progression of inoculation is shown across multiple days post inoculation with ASF. Relative abundance of (B) Parabacteroides (including ASF 519) and (C) S24-7. Box and whiskers show the interquartile range; black circles mark the outlier samples.

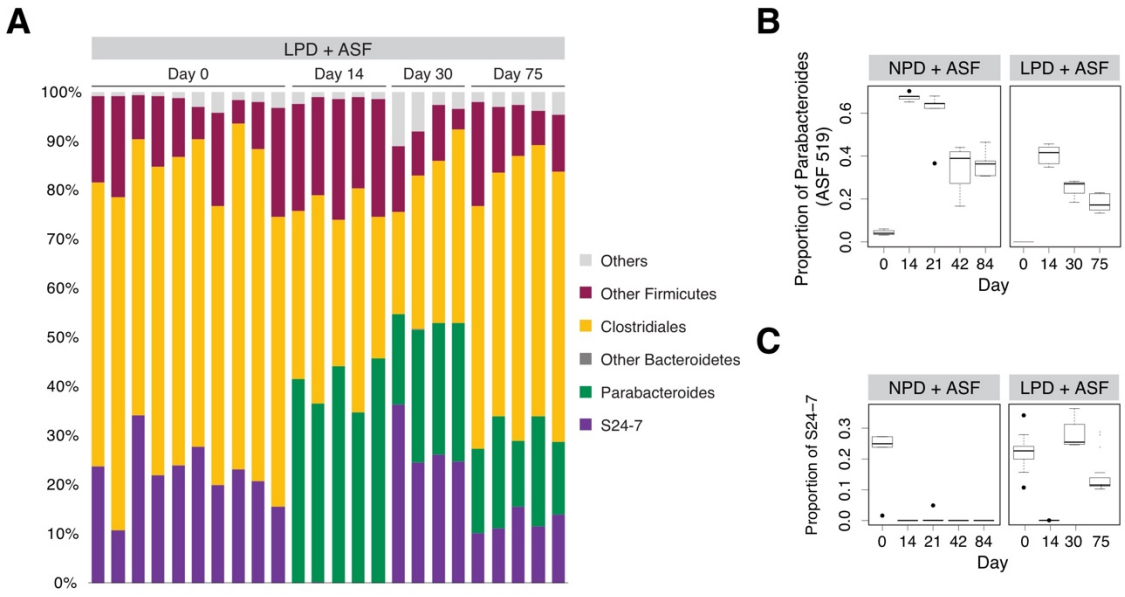
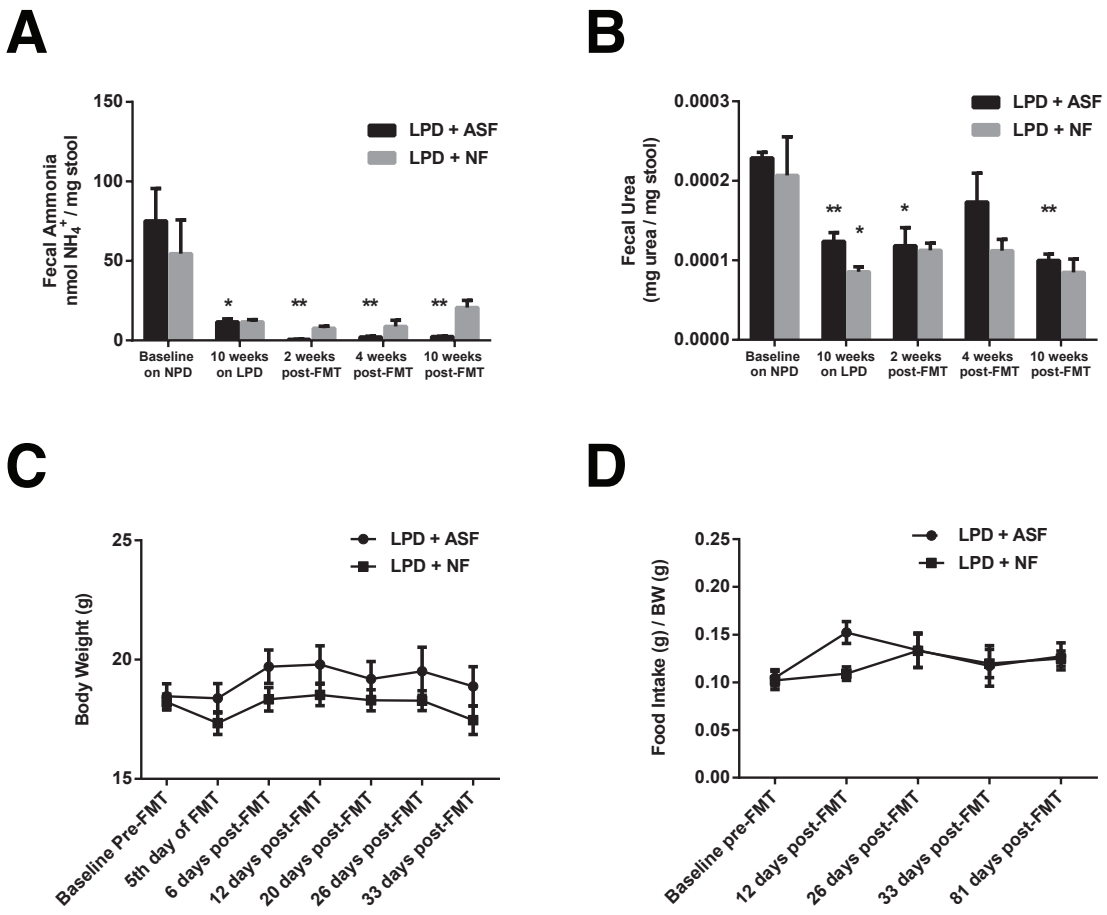


Figure 5.5. ASF transplantation alters colonic urea nitrogen recycling without significantly affecting host physiology. (A) ASF transplantation reduces fecal ammonia below the level achieved by the LPD alone (n=4-5 per group, *p<0.05 compared to baseline, **p<0.01 compared to 10 weeks on the LPD). (B) No difference in fecal urea level between ASF- and NF-transplanted mice (n=2-5 per group, *p<0.05 compared to baseline, **p<0.01 compared to baseline). No difference in (C) body weight or (D) food intake between ASF- and NF-transplanted mice (n=5 per group). Values represent mean ± SEM. Significance determined by two-tailed Student's t-test.



References

1. Hildebrandt MA, Hoffmann C, Sherrill-Mix SA, Keilbaugh SA, Hamady M, Chen YY, et al. High-fat diet determines the composition of the murine gut microbiome independently of obesity. Gastroenterology. 2009;137(5):1716-24 e1-2. Epub 2009/08/27. doi: S0016-5085(09)01457-7 [pii] 10.1053/j.gastro.2009.08.042. PubMed PMID: 19706296; PubMed Central PMCID: PMC2770164.
2. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, et al. Linking long-term dietary patterns with gut microbial enterotypes. Science. 2011;334(6052):105-8. Epub 2011/09/03. doi: science.1208344 [pii]
10.1126/science.1208344. PubMed PMID: 21885731.
3. David L, Maurice C, Carmody R, Gootenberg D, Button J, Wolfe B, et al. Diet rapidly and reproducibly alters the human gut microbiome. Nature. 2014;505(7484):559-63.
4. Dethlefsen L, Huse S, Sogin ML, Relman DA. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. PLoS biology. 2008;6(11):e280. doi: 10.1371/journal.pbio.0060280. PubMed PMID: 19018661; PubMed Central PMCID: PMC2586385.
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. doi: 0706625104 [pii]
10.1073/pnas.0706625104. PubMed PMID: 17699621; PubMed Central PMCID: PMC1959459.
6. Yurist-Doutsch S, Arrieta MC, Vogt SL, Finlay BB. Gastrointestinal microbiota-mediated control of enteric pathogens. Annu Rev Genet. 2014;48:361-82. doi: 10.1146/annurev-genet-120213-092421. PubMed PMID: 25251855.
7. Lewis JD, Chen EZ, Baldassano RN, Otley AR, Griffiths AM, Lee D, et al. Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease. Cell Host Microbe. 2015;18(4):489-500. doi: 10.1016/j.chom.2015.09.008. PubMed PMID: 26468751.
8. Albenberg LG, Wu GD. Diet and the intestinal microbiome: associations, functions, and implications for health and disease. Gastroenterology. 2014;146(6):1564-72. doi: 10.1053/j.gastro.2014.01.058. PubMed PMID: 24503132.
9. van Nood E, Vrieze A, Nieuwdorp M, Fuentes S, Zoetendal EG, de Vos WM, et al. Duodenal infusion of donor feces for recurrent Clostridium difficile. N Engl J Med. 2013;368(5):407-15. Epub 2013/01/18. doi: 10.1056/NEJMoa1205037. PubMed PMID: 23323867.
10. Hecht GA, Blaser MJ, Gordon J, Kaplan LM, Knight R, Laine L, et al. What Is the Value of a Food and Drug Administration Investigational New Drug for Fecal Microbiota Transplantation in Clostridium difficile Infection? Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association. 2013. Epub 2013/10/24. doi: S1542-3565(13)01641-8 [pii]
10.1016/j.cgh.2013.10.009. PubMed PMID: 24148361.
11. Lawley TD, Clare S, Walker AW, Stares MD, Connor TR, Raisen C, et al. Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing Clostridium difficile disease in mice. PLoS Pathog. 2012;8(10):e1002995. Epub 2012/11/08. doi: 10.1371/journal.ppat.1002995
PPATHOGENS-D-12-01455 [pii]. PubMed PMID: 23133377; PubMed Central PMCID: PMC3486913.
12. Shen TD, Albenberg L, Bittinger K, Chehoud C, Chen YY, Judge CA, et al. Engineering the gut microbiota to treat hyperammonemia. J Clin Invest. 2015;125(7):2841-50. doi: 10.1172/JCI79214. PubMed PMID: 26098218.
13. Dewhirst FE, Chien CC, Paster BJ, Ericson RL, Orcutt RP, Schauer DB, et al. Phylogeny of the defined murine microbiota: altered Schaedler flora. Appl Environ Microbiol. 1999;65(8):3287-92. Epub 1999/07/31. PubMed PMID: 10427008; PubMed Central PMCID: PMC91493.

14. Stewart G, Smith C. Urea nitrogen salvage mechanisms and their relevance to ruminants, non-ruminants and man. Nutr Res Rev 2005;18(1):49-62.
15. Fuller MF RP. Nitrogen cycling in the gut. Annu Rev Nutr 1998;18:385-411.
16. Vilstrup H, Amodio P, Bajaj J, Cordoba J, Ferenci P, Mullen K, et al. Hepatic Encephalopathy in Chronic Liver Disease: 2014 Practice Guideline by the American Association for the Study of Liver Diseases and the European Association for the Study of the Liver. Hepatology. 2014;60(2):715-35.
17. Riordan SM, Williams R. Treatment of hepatic encephalopathy. N Engl J Med. 1997;337(7):473-9. Epub 1997/08/14. doi: 10.1056/NEJM199708143370707. PubMed PMID: 9250851.
18. Saudubray JM, Nassogne MC, de Lonlay P, Touati G. Clinical approach to inherited metabolic disorders in neonates: an overview. Semin Neonatol. 2002;7(1):3-15. Epub 2002/06/19. doi: 10.1053/siny.2001.0083
S1084275601900831 [pii]. PubMed PMID: 12069534.
19. Nguyen DL MT. Protein restriction in hepatic encephalopathy is appropriate for selected patients: a point of view. Hepatol Int 2014;8(2):447-51.
20. Singh R. Nutritional management of patients with urea cycle disorders. J Inherit Metab Dis. 2007;30(6):880-7.
21. Brusilow S, Maestri N. Urea cycle disorders: diagnosis, pathophysiology, and therapy. Adv Pediatr. 1996;43:127-70.
22. Salzman NH, de Jong H, Paterson Y, Harmsen HJ, Welling GW, Bos NA. Analysis of 16S libraries of mouse gastrointestinal microflora reveals a large new group of mouse intestinal bacteria. Microbiology. 2002;148(Pt 11):3651-60. Epub 2002/11/13. doi: 10.1099/00221287-148-11-3651. PubMed PMID: 12427955.
23. Serino M, Luche E, Gres S, Baylac A, Berge M, Cenac C, et al. Metabolic adaptation to a high-fat diet is associated with a change in the gut microbiota. Gut. 2012;61(4):543-53. Epub 2011/11/24. doi: 10.1136/gutjnl-2011-301012. PubMed PMID: 22110050; PubMed Central PMCID: PMC3292714.
24. Faith JJ, McNulty NP, Rey FE, Gordon JI. Predicting a human gut microbiota's response to diet in gnotobiotic mice. Science. 2011;333(6038):101-4. Epub 2011/05/21. doi: science.1206025 [pii]
10.1126/science.1206025. PubMed PMID: 21596954; PubMed Central PMCID: PMC3303606.
25. Hamberg O. Regulation of urea synthesis by diet protein and carbohydrate in normal man and in patients with cirrhosis. Relationship to glucagon and insulin. Dan Med Bull. 1997;44(3):225-41.
26. Adam S, Almeida MF, Assoun M, Baruteau J, Bernabei SM, Bigot S, et al. Dietary management of urea cycle disorders: European practice. Mol Genet Metab. 2013;110(4):439-45.
27. Neis E, Dejong C, Rensen S. The role of microbial amino acid metabolism in host metabolism. Nutrients. 2015;7(4):2930-46.
28. Barker H. Amino acid degradation by anaerobic bacteria. Annu Rev Biochem. 1981;(50):23-40.
29. Metges C, El-Khoury A, Henneman L, Petzke K, Grant I, Bedri S, et al. Availability of intestinal microbial lysine for whole body lysine homeostasis in human subjects. Am J Physiol 1999;277(4):E597-607.
30. Metges C, Petzke K, El-Khoury A, Henneman L, Grant I, Bedri S, et al. Incorporation of urea and ammonia nitrogen into ileal and fecal microbial proteins and plasma free amino acids in normal men and ileostomates. Am J Clin Nutr 1999;70(6):1046-58.
31. Turnbaugh P, Bäckhed F, Fulton L, Gordon J. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell Host Microbe 2008;3(4):213-23.
32. Claus S, Ellero S, Berger B, Krause L, Bruttin A, Molina J, et al. Colonization-induced host-gut microbial metabolic interaction. MBio. 2011;2(2):e00271-10.
33. Lee SM, Donaldson GP, Mikulski Z, Boyajian S, Ley K, Mazmanian SK. Bacterial colonization factors control specificity and stability of the gut microbiota. Nature. 2013;501(7467):426-9. Epub 2013/08/21. doi: nature12447 [pii]
10.1038/nature12447. PubMed PMID: 23955152.

34. Liu HX, Rocha CS, Dandekar S, Wan YY. Functional analysis of the relationship between intestinal microbiota and the expression of hepatic genes and pathways during the course of liver regeneration. Journal of hepatology. 2015. Epub 2015/10/11. doi: 10.1016/j.jhep.2015.09.022. PubMed PMID: 26453969.
35. Kreisinger J, Bastien G, Hauffe HC, Marchesi J, Perkins SE. Interactions between multiple helminths and the gut microbiota in wild rodents. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2015;370(1675). Epub 2015/07/08. doi: 10.1098/rstb.2014.0295. PubMed PMID: 26150661; PubMed Central PMCID: PMC4528493.
36. Harris JK, El Kasmi KC, Anderson AL, Devereaux MW, Fillon SA, Robertson CE, et al. Specific microbiome changes in a mouse model of parenteral nutrition associated liver injury and intestinal inflammation. PLoS One. 2014;9(10):e110396. Epub 2014/10/21. doi: 10.1371/journal.pone.0110396. PubMed PMID: 25329595; PubMed Central PMCID: PMC4203793.
37. Clemente JC, Pehrsson EC, Blaser MJ, Sandhu K, Gao Z, Wang B, et al. The microbiome of uncontacted Amerindians. Science advances. 2015;1(3). Epub 2015/08/01. doi: 10.1126/sciadv.1500183. PubMed PMID: 26229982; PubMed Central PMCID: PMC4517851.
38. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, et al. Microbial co-occurrence relationships in the human microbiome. PLoS Comput Biol. 2012;8(7):e1002606. Epub 2012/07/19. doi: 10.1371/journal.pcbi.1002606
PCOMPBIOL-D-12-00158 [pii]. PubMed PMID: 22807668; PubMed Central PMCID: PMC3395616.
39. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. Nature. 2011;473(7346):174-80. Epub 2011/04/22. doi: nature09944 [pii]
10.1038/nature09944. PubMed PMID: 21508958.
40. Devkota S, Wang Y, Musch MW, Leone V, Fehlner-Peach H, Nadimpalli A, et al. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice. Nature. 2012;487(7405):104-8. Epub 2012/06/23. doi: nature11225 [pii]
10.1038/nature11225. PubMed PMID: 22722865; PubMed Central PMCID: PMC3393783.
41. Picou D, Phillips M. Urea metabolism in malnourished and recovered children receiving a high or low protein diet. Am J Clin Nutr. 1972;25:1261-6.
42. Wu GD, Lewis JD, Hoffmann C, Chen YY, Knight R, Bittinger K, et al. Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. BMC Microbiol. 2010;10:206. Epub 2010/08/03. doi: 1471-2180-10-206 [pii]
10.1186/1471-2180-10-206. PubMed PMID: 20673359; PubMed Central PMCID: PMC2921404.
43. Hoffmann C, Minkah N, Leipzig J, Wang G, Arens MQ, Tebas P, et al. DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. Nucleic acids research. 2007;35(13):e91. doi: 10.1093/nar/gkm435. PubMed PMID: 17576693; PubMed Central PMCID: PMCPMC1934997.
44. Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Methods. 2008;5(3):235-7. doi: 10.1038/nmeth.1184. PubMed PMID: 18264105; PubMed Central PMCID: PMCPMC3439997.
45. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335-6. Epub 2010/04/13. doi: nmeth.f.303 [pii]
10.1038/nmeth.f.303. PubMed PMID: 20383131; PubMed Central PMCID: PMC3156573.
46. McDonald D PM, 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.
47. Caporaso JG BK, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics. 2010;26(2):266-7.
48. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5(3):e9490. Epub 2010/03/13. doi: 10.1371/journal.pone.0009490. PubMed PMID: 20224823; PubMed Central PMCID: PMC2835736.

49. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol. 2005;71(12):8228-35. Epub 2005/12/08. doi: 10.1128/AEM.71.12.8228-8235.2005. PubMed PMID: 16332807; PubMed Central PMCID: PMC1317376.

Acknowledgements

This work was supported by NIH grants P30 DK050306, T32 DK007066, and R01 GM103591 as well as the PennCHOP Microbiome Program. Metabolic phenotyping was performed by Mouse Phenotyping, Physiology and Metabolism Core at the University of Pennsylvania.

Contribution

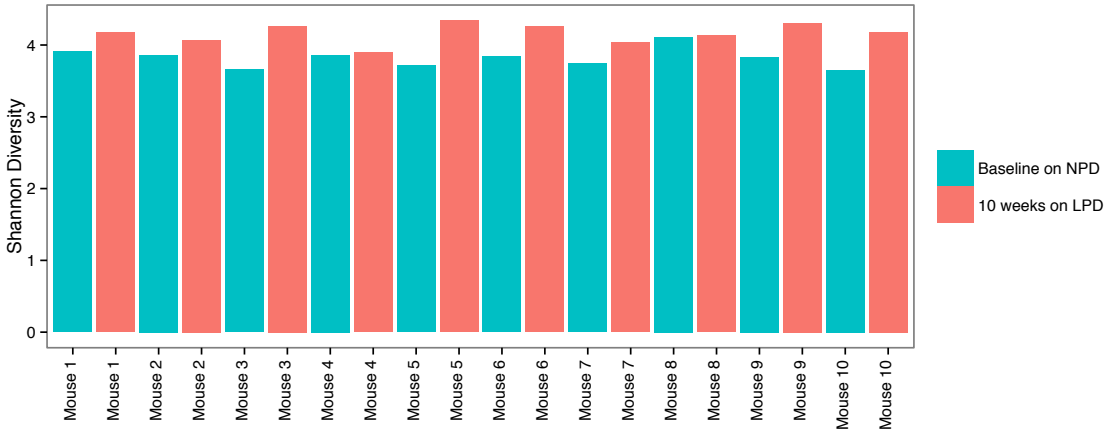
I performed all of the sequence analysis for this publication. I generated figures 5.2, 5.3 and 5.4, as well as supplemental figures 5.1, 5.2, and 5.3. I worked closely with DS and GW to generate the analysis strategy and author the manuscript.

Supplemental Information

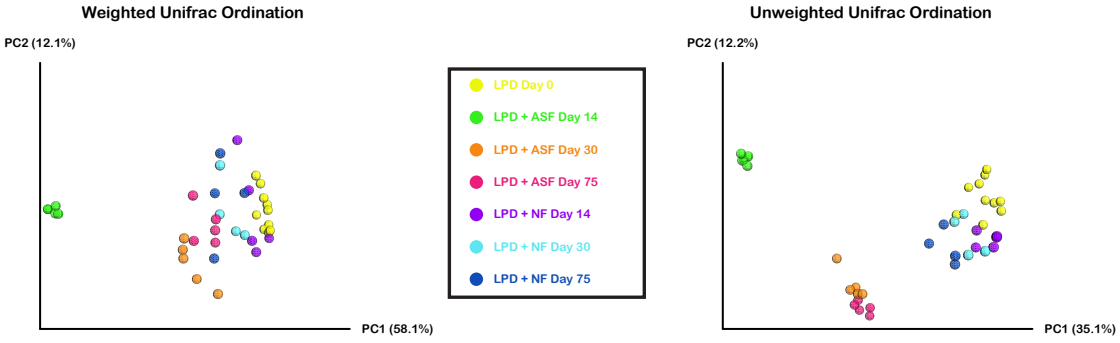
Sup. Table 5.1. Components of normal protein and low protein diets.

	NPD (kcal%)	LPD (kcal%)
Protein	20	3
Carbohydrate	68	85
Fat	12	12
	Ingredient (kcal)	
Casein	800	120
DL-Methionine	12	1.8
Corn starch	600	760
Sucrose	2000	2530
Corn oil	450	450
Vitamin mix	40	40

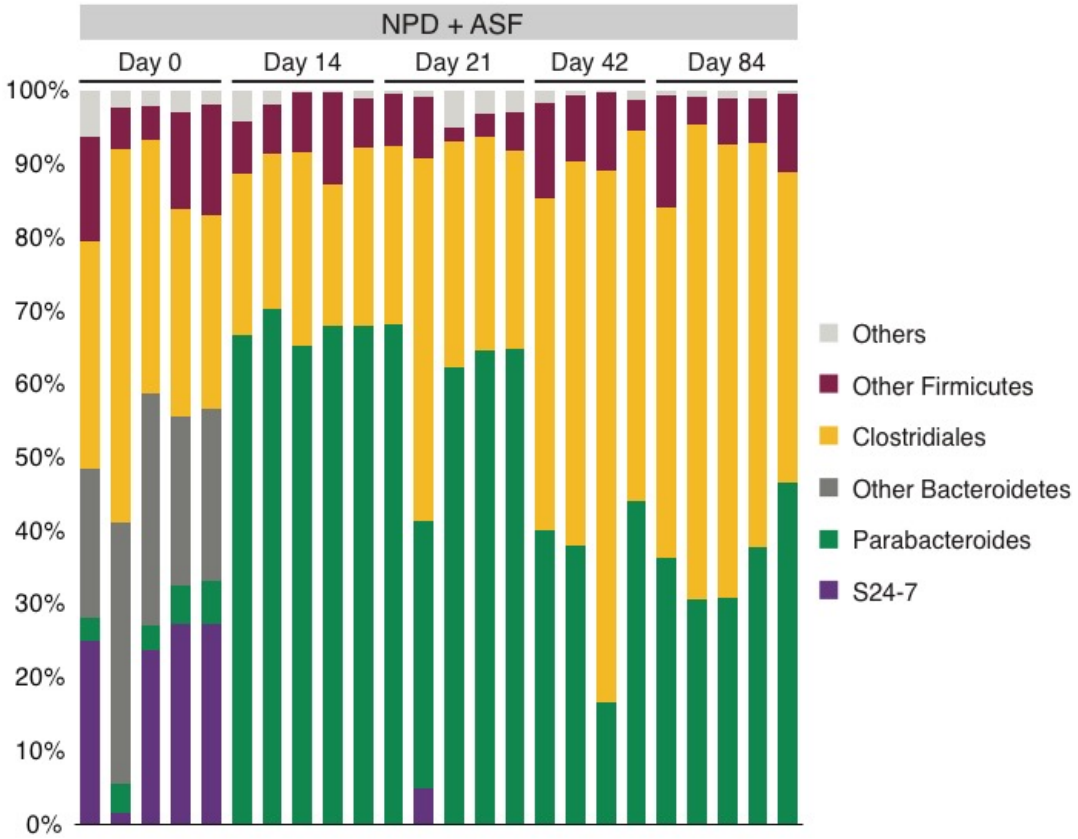
Sup. Figure 5.1. Diversity in each mouse after a LPD. Shannon diversity is shown for all ten mice while on the NPD (blue) and after ten weeks on the LPD (salmon).



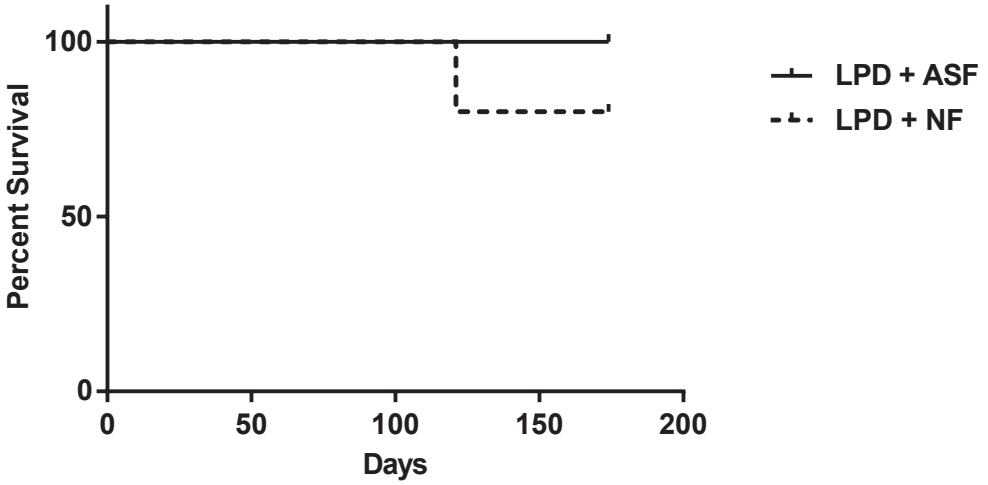
Sup. Figure 5.2. Principal coordinates analysis ordination of mice on the LPD after transplantation with ASF or NF. Changes in community membership were analyzed using (A) weighted and (B) unweighted Unifrac.



Sup. Figure 5.3. Relative abundance of bacterial taxa after ASF transplantation. Each bar represents a single sample. Samples represent pre-treated ASF-inoculated mice on the NPD (NPD + ASF). Progression is shown across multiple days post-inoculation with ASF.



Sup. Figure 5.4. Murine mortality on a LPD. Kaplan Meier curve showing no significant difference in survival between ASF- and NF-transplanted mice on the LPD (n=5 per group at start of experiment).



CHAPTER 6: Transfer of Viral Communities between Human Individuals during Fecal Microbial Transplantation.

The contents of this chapter have been published as:

Chehoud C, Dryga A, Hwang Y, Nagy-Szakal D, Hollister E, Luna R, Versalovic J, Kellermayer R, Bushman F. 2016. *Transfer of Viral Communities between Human Individuals during Fecal Microbial Transplantation.* mBio (PMID: 27025251).

Abstract

Fecal microbiota transplantation (FMT) is a highly effective treatment for refractory *Clostridium difficile* infections. However, concerns persist about unwanted co-transfer of pathogenic microbes such as viruses, which has not previously been investigated. Here we study FMT from a single, healthy human donor to three pediatric ulcerative colitis patients, each of whom received a course of 22 to 30 FMT treatments. Viral particles were purified from donor and recipient stool samples and sequenced; the reads were then assembled into contigs corresponding to viral genomes or partial genomes. Transfer of selected viruses was confirmed by quantitative PCR. Viral contigs present in the donor could be readily detected in recipients, with up to 32 different donor viral contigs appearing in a recipient sample. Reassuringly, none of these were viruses are known to replicate on human cells. Instead, viral contigs either scored as bacteriophage or could not be attributed taxonomically, suggestive of unstudied phage. The two most frequently transferred gene types were associated with temperate phage replication. Also, Siphoviridae, the group of typically temperate phages that includes phage lambda, were found to be transferred with significantly greater efficiency than other groups. Based on these findings, we propose that the temperate phage replication style may promote efficient phage transfer between human individuals. In summary, we documented transfer of multiple viral lineages between human

individuals through FMT, but in this case series none were from viral groups known to infect human cells.

Introduction

Fecal microbiota transplantation (FMT) has been used to treat relapsing *Clostridium difficile* infection with great success and shows promise for other indications (1, 2). However, many pathogenic human viruses are known to be transmitted via the fecal-oral route, and viral transfer accompanying FMT has not been studied previously. Here we investigate viral transfer during FMT using model-independent metagenomic sequencing.

Introduction of single viruses into humans has been closely studied, but the transfer of populations of viruses is much less well-understood. Multiple viruses are commonly found as co-infections in humans, though it is usually unknown whether acquisition of these viruses was sequential or simultaneous. Careful tracking of transmitted founder viruses during HIV and HCV transmission has revealed examples of probable simultaneous transfer of multiple lineages (3, 4). HIV and HCV commonly co-occur and may be acquired simultaneously during intravenous drug use (5). For the case of satellite viruses (e.g. Hepatitis B and its satellite virus Hepatitis D) the two are inferred to commonly infect together because the satellite depends on its helper for replication (6). Multiple viruses that replicate in human cells may be transferred together during organ transplantation (7). For bacterial viruses, simultaneous acquisition of multiple lineages has rarely been investigated.

The human fecal virome is immense in the numbers of viruses present—comparable to the number of bacteria, which is 10^{11} per gram of feces—and the number of different types. Gut bacteriophages are so numerous and diverse that sequence databases contain only a small fraction of the global population (8-14). Thus, simply sequencing total DNA from stool and aligning to database genomes generally misses most of the bacteriophage sequences present, necessitating specialized methods.

In this study, we investigated the transfer of viral communities between humans through FMT and characterized features associated with particularly efficient transmission. We took advantage of a case series where feces from a single donor were used to treat three children with ulcerative colitis (UC) (15). These subjects received a particularly intensive procedure, which consisted of 22 to 30 FMT treatments per subject delivered by colonoscopy or enema during a 6 to12 week period. We performed metagenomic analysis of viral particle fractions purified from fecal samples from the donor, the donor product, and recipients before, during, and after transplantation. To track viral transfer, we assembled contigs representing viral genomes and monitored their abundances through the transplantation process. All recipients were detectably colonized by multiple donor viruses, documenting transfer of whole viral populations. None of the transferred viruses were from groups known to replicate on human cells, supporting the safety of the FMT procedure in this application.

Results

Subjects studied

Three pediatric UC patients receiving immunotherapy regimens were switched to receiving a course of 22 to 30 FMT treatments via colonoscopy or enema during a 6 to12 week period (Figure 6.1A) (15). Donor stool specimens were from a single, healthy, 37-year-old male. FMT in all three patients was associated with a symptom-free period of at least 4 weeks, supporting the temporary withdrawal of immunotherapy (no treatment other than mesalamine). Treatment was associated with endoscopic and histologic remission for at least 2 weeks after the last FMT. All three subjects eventually relapsed and resumed therapy with immunomodulators.

Analysis of transfer of bacterial lineages

Bacterial microbiota composition was investigated by purifying DNA from stool and sequencing a segment of the 16S rRNA gene as described (15). The donor samples contained high levels of *Bacteroides* and lower levels of Firmicutes such as *Dialister* and *Subdoligranulum*.

Prior to FMT, the three recipients harbored communities dominated by *Subdoligranulum*, *Prevotella*, and *Bacteroides* respectively. During FMT, *Bacteroides* dominated in all three recipients (Figure 6.1B). After completion of the FMT treatment, *Bacteriodes* persisted, but Firmicutes lineages also increased in abundance in recipients 2 and 3. Global community analysis showed that for all three recipients, microbial communities during FMT became detectably (though modestly) closer to that of the donor. After FMT, all three recipients' microbial communities increased in diversity (Sup. Figure 6.1), suggestive of acquisition of healthier microbiota.

Composition of the fecal virome in donors and recipients

To study viral transfer, virus-like particles (VLPs) were purified from 18 samples. We compared four donor samples of two types: two were crude stool samples and the other two were processed to generate the liquid product used for FMT. Recipient samples were taken prior to FMT, during the FMT intervention (after the 13th FMT treatment), and after the FMT was completed. The five control samples were 1) a stool sample from a healthy 25-year-old male, 2) two stool samples from pediatric Crohn's Disease patients on immunotherapy but not receiving FMT, and 3) two buffer only negative control samples. The first three control samples were from well-studied subjects and serve as positive controls. The two pediatric Crohn's Disease patients were selected because a previous metagenomic study had shown them to harbor Anelloviruses (Torque Teno virus) and Parvoviruses (Human Bocavirus) (16). The two buffer controls were included to document contamination from reagents and environmental admixture.

VLPs were purified from the 18 samples, then DNA isolated from the particles. VLP DNA was amplified using multiple displacement amplification (MDA) and the product sequenced using Illumina technology. Quantitative PCR for 16S was used to assay contaminating bacterial DNA, documenting only low level carry over (Sup. Table 1). Here we only investigate DNA viruses. RNA viruses are less abundant in stool than DNA viruses, and previous studies from us and

others suggest that many of the RNA viruses present in stool are transient plant RNA viruses derived from food ((17) and unpublished).

At least one million reads were acquired from each fecal sample (Sup. Table 2), allowing detailed investigation of the viral populations. Low quality reads were filtered out, as were reads mapping to the human genome (hg18) and phage phiX174 (the latter is added to sequencing mixtures for technical reasons). Reads were aligned to a curated viral database (18) using a permissive alignment threshold (BLAST e-value 10^{-3}). After filtering out low abundance negative control samples (<50,000 reads) plots of the proportions of the most abundant taxa are shown (Figure 6.1C). The majority of reads did not align to reference database sequences (an average of 17.9% of reads returned significant matches).

Of the reads that did align, the most common assignments were Microviridae, which are non-enveloped, single stranded DNA phages from the family including phiX174, and Caudovirales, which is a broad family including all of the tailed phages. Of the two controls with known animal cell viruses, Anelloviruses were recovered efficiently, and Bocavirus less efficiently. The MDA step is known to amplify single stranded circular DNA particularly efficiently, likely explaining the abundance of Anelloviruses and Microviridae, both of which have small single stranded circular DNA genomes.

To examine the VLP genomes in our samples more fully, reads were assembled into contigs using deBruijn graph assembly (19) and overlap consensus methods (20). A total of 3634 contigs were built from the four donor samples (Sup. Table 3). Of these, 261 were greater than 3000 base pairs in length (mean length 8161 bp and maximum length 65335 bp). For the recipients, a total of 486, 1177, and 1936 contigs were built from each pre-FMT sample; of these contigs, 38, 138, and 127 contigs were greater than 3000 base pairs (Sup. Table 4). These contigs were assessed for length, circularity, and number of open reading frames (ORFs). Nine of the 261 contigs were circular, suggesting that we had sequenced the entire viral genome (Sup. Figure 6.2). With one exception, these clustered at specific lengths suggestive of Anelloviruses

(~3000 bp), Microviridae (~5000-6000 bp), and Siphoviridae (typically >30000 bp). One circular contig (8662 bp) was annotated as Podoviridae (typically >40000 bp), suggesting it may have assembled as a circle due to the presence of a repeated sequence. Linear contigs are either complete linear genomes or genome fragments.

Proteins encoded in the gut virome

To catalog the genes present, open reading frames (ORFs) were predicted, conceptually translated and aligned to several protein databases (19, 21, 22), including one composed of conserved protein domains (from the Conserved Domains Database (CDD)) known to be associated with viruses (12, 14). A total of 3,321 ORFs were predicted. Protein types found in the donor and in each recipient pre-FMT are summarized in Figure 6.1D. Only an average of 7.3% of proteins were annotated, which emphasizes the size of the pool of uncharacterized bacteriophage genes. Prominent CDD types were associated with phage capsids, tails, portals, terminases and other structural components. Additional annotations were associated with effector proteins including bacteriocin, beta-lactamase, reverse transcriptase, and restriction modification.

Contigs of length >3000 bp had on average 6 ORFs, allowing us to use gene family membership to annotate the probable viral family of origin. We found that on average each contig had two ORFs matching a viral genome (BLAST e-value 10^{-5}). The majority of classified VLP genomes belonged to tailed phages (Caudovirales), which include Siphoviridae, Myoviridae, and Podoviridae. Microviridae were also prominent. In several cases single ORFs annotated as belonging to animal cell virus lineages (Poxviridae, Adenoviridae, and Herpesviridae), but multiple other ORFs on the same contig annotated as Caudovirales, so in these cases the Caudovirales attribution was retained, or multiple other ORFs were unclassified weakening the validity of an animal cell lineage. The two Crohn's disease samples pre-selected for containing animal cell viruses returned contigs representing Anellovirus and Bocavirus as expected.

Transfer of viral lineages with FMT

We then used the contigs of >3000 bp to investigate the transfer of phages during FMT (Figure 6.2, Sup. Figure 6.3). To score detection of a donor contig in a recipient, we required >50% coverage of the contig and alignment of at least 5 paired reads, where both reads in each pair showed high quality alignment to the target contig. We note that these criteria are conservative, so the numbers of contigs transferred represent minimal estimates. Comparing to recipient samples at the first time point after initiation of FMT, while FMT was ongoing, we detected 21, 42, and 16 donor contigs respectively. At the second time point, 5 to 13 weeks later and 2 to 4 weeks after the completion of the FMT treatment, we detected 5, 7 and 34 donor viral contigs, respectively. Sup. Figure 6.4 shows examples of sequence alignments of reads from recipients at time points following transfer onto contigs built from donor reads, illustrating high coverage.

To confirm the transfer from the donor to recipients, we used quantitative PCR (qPCR) to analyze VLP DNA samples (Figure 6.3). As template we analyzed fecal VLP DNA specimens that were not subjected to MDA amplification, which is known to distort relative abundances. To confirm that some lineages transferred efficiently, we analyzed four VLP contigs from the donor that were detected in at least two different recipients. QPCR assays were designed for each of the four contigs (Sup. Table 5), and positive control DNAs were synthesized for use as copy number standards. We compared results for the four donor samples to the nine recipient samples (three patients, three time points each). We also tested the same positive controls used in the viral sequencing analysis, specifically stool VLP DNA from a healthy 25-year-old male and two pediatric Crohn's Disease patients. As a negative control, we used a commercial phage lambda DNA sample.

All four qPCR assays detected the expected contigs in the donor samples and in at least two recipients. Levels were low in all controls. For contig 100-94 (annotating as Microviridae) and 100-182 (Unattributed) donor contigs were detected after FMT in all three recipients (Figure 6.3A and 6.3B). For contigs 100-21 (Podoviridae) and 19 (Myoviridae), transfer was robust to subjects

2 and 3 (Figure 6.3C and 6.3D). The maximum copy numbers detected were in the range of 8-15X10^3 copies per nanogram of VLP DNA.

Viral features associated with efficient viral transfer

We next investigated features of viral contigs associated with efficient transmission during stool transplantation. We first asked whether any inferred phage-encoded proteins could be identified that correlated with efficient transmission. Figure 6.4A shows the top two conserved domain families that correlated with frequency of transmission (p-values<0.05, odds ratio >4.2 times more likely to be transferred). These transferred domains were found to be associated with the temperate replication style, in which phages form quiescent long-term associations with their bacterial hosts. The highest scoring family (pfam01051) is annotated as “initiator replication protein”. The best-studied member of this group is RepA protein of the temperate phage P1. RepA initiates DNA replication on the P1 genome, and related proteins also mediate replication of Herpes viruses. The second highest scoring family (pfam14659) is a domain from the lambda integrase family of integration proteins. Lambda integrase is well known to mediate integration during prophage formation. Thus both the analysis of preferential transfer of viral family types and transfer of viral genes indicated preferential transmission of temperate phage.

We next asked whether any viral families were preferentially transferred by comparing frequencies in the donor versus recipient VLP populations. We focused the analysis on our most accurate detections, specifically the donor viral pool consisting of contigs >3000 bp. Comparison showed that Siphoviridae were significantly more likely to be transferred (p=0.0072 by Fisher’s exact test, Figure 6.4B). We repeated these tests by looking at Siphoviridae compared to all contigs and again found Siphoviridae to be significantly more likely to be transferred (p<0.0001 by Fisher’s exact test, Figure 6.4C). Siphoviridae include phage lambda, 434, and P22 and are known to include viruses with the capacity to integrate into host genomes (23).

To confirm transfer, we selected four contigs annotating as Siphoviridae and called as transferred, and designed qPCR assays to verify transfer. VLP DNA samples (not subjected to MDA) were tested, and contigs were detected in the donor and recipient as expected from the sequencing data (Figure 6.4 D-G). In all four cases, the transferred phage were detected in at least two out of the three recipients, indicating transfer to multiple individuals.

Discussion

Here we report transfer of gut viral communities associated with FMT. No viruses infecting human cells were detected in the donor, so reassuringly no transfer could be documented in recipients. Numerous recipient VLP sequence reads showed high quality alignments to donor VLP contigs, suggesting transfer with FMT. Many annotated as containing genes known to be phage encoded, supporting the inference that the VLP contigs corresponded to phage. From 5 to 34 transferred contigs could be detected in recipient samples. Simultaneous transfer of multiple viruses has rarely been documented in humans but was readily detected here, potentially associated with the large number of FMT treatments (22 to 30 per subject). Based on these observations, we propose that transfer of bacteriophage populations is a general characteristic of FMT.

An alternative interpretation for our data could be that viruses were not transferred from donor to recipients, but instead were present in recipients initially below the level of detection. According to this view, viruses pre-existing in the recipient would have grown out after FMT, resulting in spurious appearance of transfer. We cannot completely rule this out, but several aspects of phage biology in the human gut support our interpretation. Bacteriophage populations in the intestines of healthy humans are known to show extreme differences among individuals (8-14). However within an individual, populations show notable longitudinal stability, at least in healthy subjects (8, 10, 11). As a result, most VLP genomes are likely recognizable by their individual of origin. Thus, while it is possible that some new detections in recipients represent the

outgrowth of pre-existing viruses, or even *de novo* colonization, we expect that most recipient sequences resembling donor VLPs are genuinely derived from the donor stool product.

Our most unexpected finding was that lysogeny is associated with efficient transfer by FMT. This was indicated first by the preferential transfer of the Siphoviridae, the group including the well-studied temperate phage lambda. Analysis of genes associated with transfer also made this point (Figure 6.4A). The most strongly associated gene encoded the RepA protein, important for replication of the temperate phage P1, and the second, encoded a domain of the lambda integrase proteins important for prophage formation.

Based on our findings, we propose that the temperate phage replication style may have evolved in part to optimize phage transmission between environments. If bacteriophage must pass through hostile environments to reach locations favorable for replication, transportation as a prophage may promote dispersal. Transmission of gut viruses between human hosts potentially involves persistence outside the gut, likely representing a transmission barrier. Viruses transferred efficiently by the fecal-oral route are rarely enveloped, suggesting that tough physical properties may be needed to survive fecal-oral transmission. In feces, an enveloping membrane may be damaged by the detergent effects of bile salts or by drying after shedding in feces. Respiratory viruses, in contrast, are commonly spread on respiratory droplets and are often enveloped. Integration of prophage DNA into the bacterial genome may provide a protective vehicle supporting phage transmission despite exposure to feces or other hostile environmental pressures. Thus, we propose that the temperate phage replication style may in part have evolved to maximize survival during transmission.

In summary, we presented a metagenomic survey of DNA viruses transferred during FMT. Numerous temperate phage were found to be transferred, but no viruses detected corresponded to pathogenic viruses that infect human cells. Methods for preparing and introducing feces into patients vary and may influence the likelihood of viral transmission, so additional characterization of viral transfer can help guide development of safest practices.

Methods

Human subjects

Human subjects and methods used for FMT have been described (15). As stated in the supplemental methods, fresh stool specimens from the healthy adult donor were aliquoted to ~50 g aliquots. Cold sterile normal saline solution was added prior to homogenization in a strainer bag with 500-μm pore size (Seward Laboratory Systems Inc., Port Saint Lucie, FL) using the Smasher Laboratory Blender/Homogenizer (AES CHEMUNEX Inc., Cranbury, NJ). Both fresh stool samples and the homogenized stool preparations which were used for the transplantation were analyzed in this study. Experiments were carried out under a protocol approved by the Institutional Review Board of Baylor College of Medicine (H-30591).

DNA purification and sequencing

Detailed methods are presented in the Supplemental Material and summarized here. Phage particles were purified from stool samples by suspension in SM buffer and filtration through a 0.22 μm filter (EMD, Millipore). The filtrate was concentrated using the 100 kDa molecular weight cutoff filter Centricon Plus 70 (Millipore), resuspended in 40 mL Buffer SM, and concentrated. The concentrate was treated with DNase I and RNase (Roche) at 37°C for 30 min to eliminate non-encapsulated nucleic acids, then the enzymes were deactivated at 65°C for 15 min. Total phage DNA was extracted from particles using the QIAamp DNA Stool Kit (Qiagen). Extracted phage DNA was amplified using Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare). Libraries were made using the Illumina Nextera XT Samples Prep Kit and quantified using the SYBR Fast Illumina Library Quantification Kit (Kapa Biosystems).

Purified VLP DNAs were sequenced on three MiSeq runs (250 base per read, paired-end sequencing) according to manufacturer's instructions (Illumina). Two negative control samples yielded 1242 and 36156 reads respectively. Contigs were built from the four donor samples and were annotated with open reading frames (ORFs), integrase genes, and viral family names based

on similarity to reference viral genomes. ORFs were also aligned to a database of conserved domain integrase genes (to assess potential temperate replication style), and also the Virulence Factor Database, Aclame database of mobile genetic elements, NCBI's reference viral database, NCBI's reference bacterial database, and NCBI's nucleotide database to assess additional functional features (summarized in Supplemental Tables 3 and 4). Reads from all of the samples were mapped to contigs built from the four donor samples using Bowtie2 (24). Coverage metrics were calculated, including concordance rate between paired reads, average depth of coverage, and the percentage each contig was covered by reads.

QPCR validation

The qPCR for quantitating contigs was performed using FastSYBR Green Master Mix (Applied Biosystems). Primers are described in Supplemental Table 5. The ZeroBlunt TOPO Vector (ThermoFisher Scientific), containing the corresponding gBLOCK synthetic copy number control sequence, was linearized and used as a standard control.

Quantification of 16S rRNA copy number was performed using the real-time TaqMan method with TaqMan® Environmental Master Mix 2.0 and primers as indicated in Supplemental Table 5. The ZeroBlunt TOPO Vector containing a near full-length clone of *Escherichia coli* 16S was linearized and used as a standard control.

Statistical analysis

Fisher's exact test and odds ratios were used to assess significance of conserved domains or viral family transfer.

Assessing contamination of phage DNA preparations with bacterial and human DNA

Purities of the phage DNA preparations were assessed using qPCR to quantify bacterial 16S rRNA gene copies (Sup. Table 6.1). Values ranged from 7 to 5021 16S rRNA gene copies/ng of purified phage DNA (average of 257 copies/ng). The contribution of bacterial DNA per

nanogram of purifed phage DNA could be estimated by assuming the average bacterial genome has 2 to 5 rRNA gene copies and is 2.5 to 5.5 Mb (25) total (MW 1.6 x 10^9 to 3.3x10^9), implying that implying that on average 1.0 nanogram of phage DNA was contaminated with 1.1 picograms of bacterial DNA.

Human DNA contamination was also assayed by aligning sequence reads from the VLP DNA samples to the human genome. Most samples contained <1% human DNA; the highest value recorded was 4% (Sup. Table 2). We thus conclude that bacterial and human DNA were minor contributors to the VLP genomic DNA samples studied.

Figures

Figure 6.1. Fecal microbiota transplantation (FMT) to treat ulcerative colitis. A) Diagram of the treatment regimens. Patient 1 received 30 rounds of FMT, Patient 2 received 25, and Patient 3 received 22. All patients were in remission (gray shaded box) while receiving the FMT course and remained in remission for more than 11 weeks following their last FMT. However, all three eventually experienced a relapse requiring immunotherapy. Mesalamine, given orally or through a rectal enema, was allowed during the trial, depending on clinical disease activity. Three samples (indicated as triangles) were taken from each patient: one before the beginning of the FMT course (pre), the second after the 13th FMT (during), the third two to three weeks after finishing the FMT course (post). B) Bacterial lineages in the samples studied. Bacterial taxonomic representation and abundance was characterized by sequencing of 16S rRNA gene tags. Details are described in (15). C) Viral families detected in the samples studied, assessed by alignment of reads to a curated database of viral sequences from (18). D) Gene types inferred from analysis of the VLP contigs. VLP contigs were analyzed for open reading frames (ORFs), then ORFs annotated using CDD. Output was interpreted using custom annotation relating pfam domains to viral gene functions. Gene types identified are summarized to the right of the figure. Percentage of classified phage ORFs was 8.9%, 7.1%, 3.8% and 9.5% respectively.

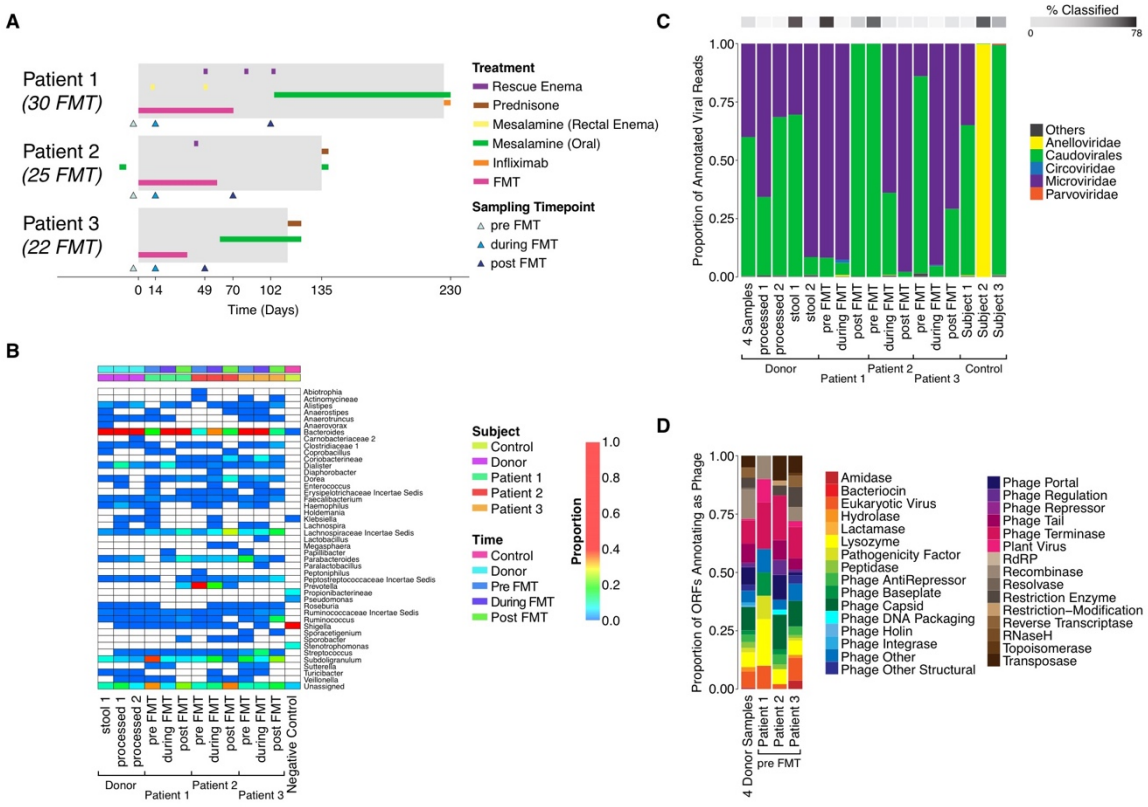


Figure 6.2. Transfer of phage between human individuals. Transmission is shown using Sankey diagrams. The patients 1-3 are shown from top to bottom. For each subject, contigs >3000 bp are shown as the box at the left, with recipient contigs on top and donor contigs below. In the middle, contigs detected at the during FMT time point are shown, to the right are contigs detected after cessation of FMT (post). Curves with darker colors show persistence of donor contigs, curves with lighter colors persistence of recipient contigs. The color of each strand represents a single viral family. Numbers indicate the number of donor contigs transferred from each viral family.

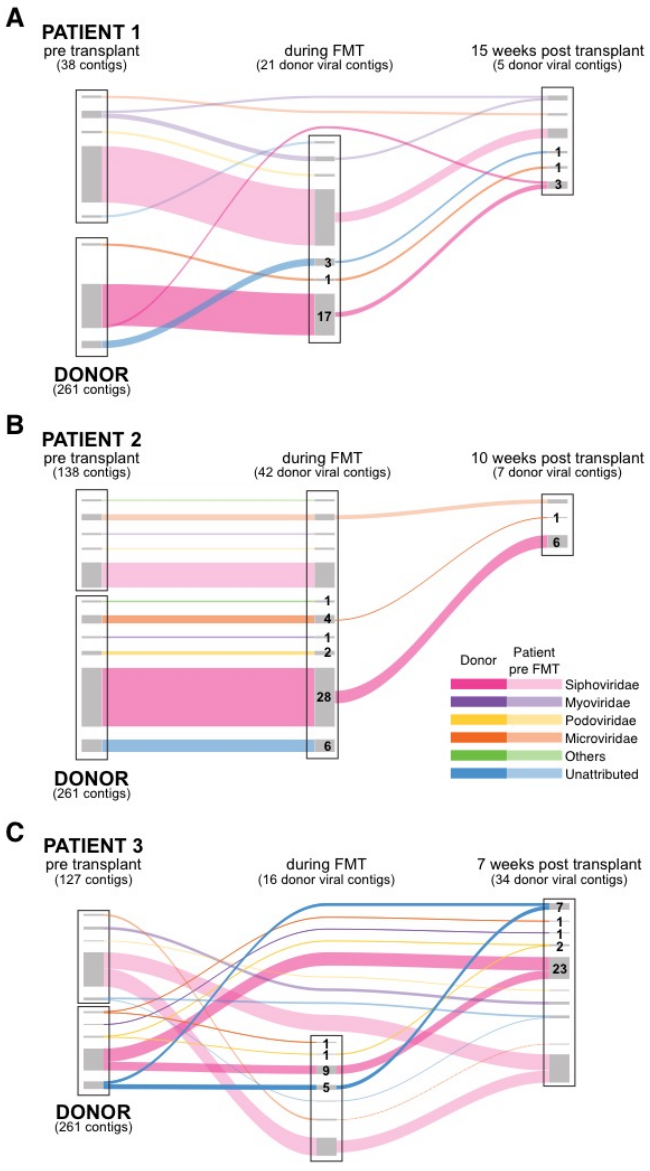


Figure 6.3. Analysis of repeatedly transferred VLP contigs by qPCR. VLP contigs built from donor samples that are indicated by sequence data to be present in patients after FMT treatment, but not in recipient pre-FMT samples, were quantified by qPCR. VLP DNA from stool was tested without use of Genomifi Amplification (to reduce distortions of abundance). Abundance was quantified from 4 samples from the donor (two whole stool and two purified product prepared for FMT), the 3 recipient patients at three time points each, and 3 control individuals. As a negative control, Lambda DNA was also tested (far right). The dashed line along the baseline indicates the inferred background level in the lambda DNA negative control. Contig length, circularity, and viral lineage as attributed by ORF annotation is indicated on each panel. A) Contig 100-182 (unattributed); B) Circular Contig 100-94 (Microviridae); C) Contig 100-21 (Podoviridae); D) Contig 19 (Myoviridae).

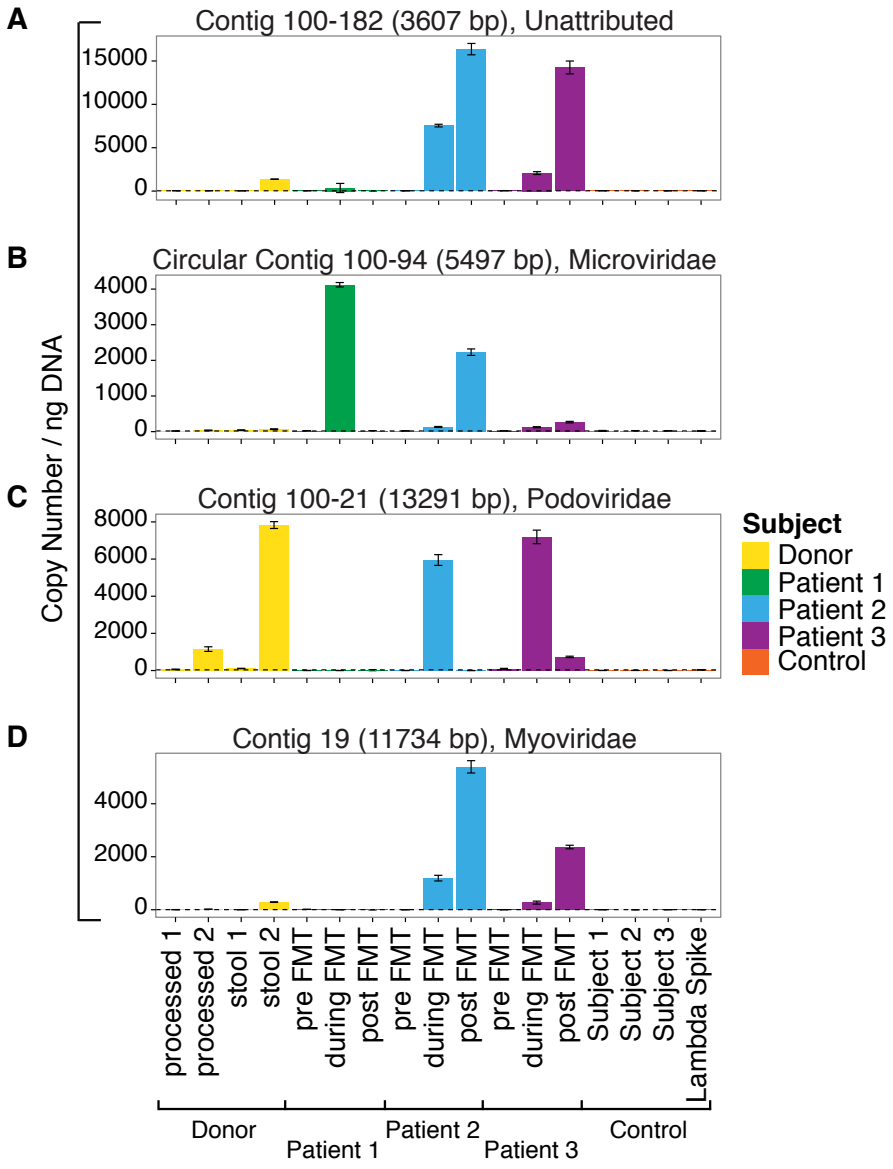
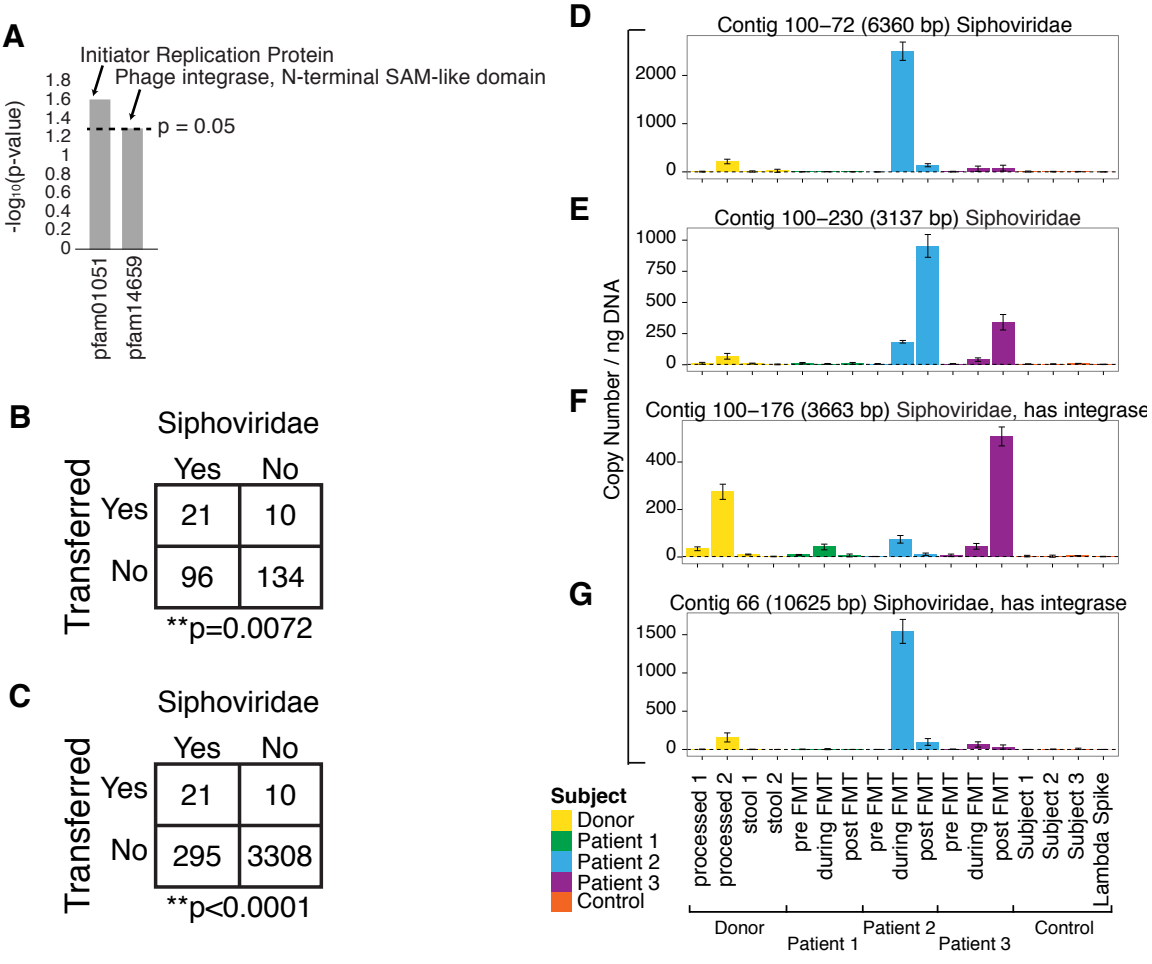


Figure 6.4. Preferential transfer of Siphoviridae between human individuals. A) Top differentially transferred genes based on CDD annotation in the pfam database. A total of 302

CDD hits were found in the contigs. The top two domains (p-value < 0.05, odds ratio > 4) based on CDD annotation which were most frequently transferred are shown. B) Comparison of the frequency of transfer to least 2 of the 3 patients for Siphoviridae contigs versus all others contigs >3000 bp. Here transfer is defined as achieving >50% coverage and >5 paired reads, where both reads in each pair detect the same contig. If transfer is defined as detection of a donor contig during FMT, post-FMT, or both in at least one patient, the favoring of Siphoviridae achieves p<0.0001. C) Comparison of the frequency of transfer to least 2 of the 3 patients for Siphoviridae contigs versus all other contigs built. D-G) Contigs annotating as Siphoviridae found in the donors and suggested by sequence data to be transferred to multiple recipients validated using qPCR. Samples tested and annotation were the same as in Figure 3. D) Contig 100-72; E) Contig 100-230; F) Contig 100-176; G) Contig 66.



References

1. van Nood E*, et al.* (2013) Duodenal infusion of donor feces for recurrent Clostridium difficile. *The New England journal of medicine* 368(5):407-415.
2. Kelly CR*, et al.* (2015) Update on Fecal Microbiota Transplantation 2015: Indications, Methodologies, Mechanisms, and Outlook. *Gastroenterology* 149(1):223-237.
3. Parrish NF*, et al.* (2013) Phenotypic properties of transmitted founder HIV-1. *Proceedings of the National Academy of Sciences of the United States of America* 110(17):6626-6633.
4. Salazar-Gonzalez JF*, et al.* (2009) Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *The Journal of experimental medicine* 206(6):1273-1289.
5. Hagan H & Des Jarlais DC (2000) HIV and HCV infection among injecting drug users. *The Mount Sinai journal of medicine, New York* 67(5-6):423-428.
6. Rizzetto M*, et al.* (1977) Immunofluorescence detection of new antigen-antibody system (delta/anti-delta) associated to hepatitis B virus in liver and in serum of HBsAg carriers. *Gut* 18(12):997-1003.
7. Eastlund T (1995) Infectious disease transmission through cell, tissue, and organ transplantation: reducing the risk through donor selection. *Cell transplantation* 4(5):455-477.
8. Reyes A*, et al.* (2010) Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 466(7304):334-338.
9. Reyes A, Semenkovich NP, Whiteson K, Rohwer F, & Gordon JI (2012) Going viral: next-generation sequencing applied to phage populations in the human gut. *Nat Rev Microbiol* 10(9):607-617.
10. Reyes A, Wu M, McNulty NP, Rohwer FL, & Gordon JI (2013) Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. *Proceedings of the National Academy of Sciences of the United States of America* 110(50):20236-20241.
11. Minot S*, et al.* (2013) Rapid evolution of the human gut virome. *Proceedings of the National Academy of Sciences of the United States of America* 110(30):12450-12455.
12. Minot S*, et al.* (2011) The human gut virome: inter-individual variation and dynamic response to diet. *Genome research* 21(10):1616-1625.
13. Minot S, Wu GD, Lewis JD, & Bushman FD (2012) Conservation of gene cassettes among diverse viruses of the human gut. *PloS one* 7(8):e42342.
14. Minot S, Grunberg S, Wu GD, Lewis JD, & Bushman FD (2012) Hypervariable loci in the human gut virome. *Proceedings of the National Academy of Sciences of the United States of America* 109(10):3962-3966.
15. Kellermayer R*, et al.* (2015) Serial fecal microbiota transplantation alters mucosal gene expression in pediatric ulcerative colitis. *The American journal of gastroenterology* 110(4):604-606.
16. Lewis JD*, et al.* (2015) Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease. *Cell host & microbe* 18(4):489-500.
17. Zhang T*, et al.* (2006) RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS biology* 4(1):e3.
18. Norman JM*, et al.* (2015) Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* 160(3):447-460.
19. Peng Y, Leung HC, Yiu SM, & Chin FY (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28(11):1420-1428.
20. Treangen TJ, Sommer DD, Angly FE, Koren S, & Pop M (2011) Next generation sequence assembly with AMOS. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* Chapter 11:Unit 11 18.
21. Leplae R, Lima-Mendez G, & Toussaint A (2010) ACLAME: a CLAssification of Mobile genetic Elements, update 2010. *Nucleic acids research* 38(Database issue):D57-61.
22. Marchler-Bauer A*, et al.* (2015) CDD: NCBI's conserved domain database. *Nucleic acids research* 43(Database issue):D222-226.

23. Ptashne M (1986) *A genetic switch: Gene control and phage. lambda* p Medium: X; Size: Pages: (138 p).
24. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods* 9(4):357-359.
25. Nayfach S & Pollard KS (2015) Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. Genome Biology 16:51.

Acknowledgements

We are grateful to members of the Bushman laboratory for help and suggestions. We thank Laurie Zimmerman for assistance with the figures. This work was supported by P30AI 045008 (FDB), AI 082020-05A1 (FDB), T32AI007632 (CC), and the PennCHOP Microbiome Program. RK was supported by the Gutsy Kids Fund including philanthropic donations from the Karen and Brock Wagner family. JV was supported by the National Institutes of Health (R01 AT004326, UH3 DK083990, and U01 CA170930) and Texas Medical Center Digestive Diseases Center (P30 DK56338).

Contribution

Working with AD, I was the major contributor to the viral annotation analysis pipeline. I performed all of the sequence analysis, figure generating, and statistics testing. I also worked with FB to generate the analysis strategy and author the manuscript.

Supplemental Information

Supplemental Methods

Classifying viral read

All MiSeq reads were BLAST aligned to a custom database (Norman et al., 2015) using an e-value 10^{-3} and max_target_seqs 100. All default settings were used unless otherwise stated. The BLAST outputs (compressed BLAST XML format) were analyzed using MEGAN (Huson et al., 2011) with the following parameters: minsupport=1, minscore=30.0, toppercent=10.0, and mincomplexity=0.44. All classifications that did not include a viral family (e.g. “root;Viruses;dsDNA

viruses, no RNA stage; unclassified dsDNA phages", "root;Viruses;unclassified phages;Enterococcus phage EF62phi") were considered Unclassified. Proportions of the classified viral families were calculated. Samples with less than 50,000 reads (i.e. two negative control SM buffer samples) were excluded.

Bacterial data

Previously published bacterial 16S data (Kellermayer et al., 2015) was reanalyzed using R packages qiimer and vegan. Shannon diversity was calculated for each sample. A principal coordinate ordination plot, based on the Weighted Unifrac distances, was built. The Weighted Unifrac distance between each patient sample and the centroid of the donor samples was calculated.

Assembling, annotating, and visualizing viral contigs

Sequence reads were first assembled into contigs using IDBA-UD (Peng et al., 2012) with pre-correction and minimum and maximum k-mer lengths set to default 20 and 100 bp respectively. Minimo (Treangen et al., 2011), a program which utilizes a conservative overlap-layout-consensus algorithm, was used to further combine the contigs. Contigs were built from 1) all 4 donor samples, including two crude stool samples and two processed samples which were blended and filtered to generate the liquid preparation used for the FMTs, 2) patient 1 pre-FMT's samples, 3), patient 2 pre-FMT samples, and 4) patient 3 pre-FMT samples. Each of the three patients' pre-FMT sample (after Genomiphi amplification) was sequenced thrice. The sequencing data from each patient's pre-FMT samples was combined.

After the contigs were built, reads from each donor and patient sample were aligned using Bowtie2 (Langmead and Salzberg, 2012) to each of the four sets of contigs. Bowtie's very-sensitive parameters (i.e. -D 20 -R 3 -N 0 -L 20 -i S,1,0.50) were used. Samtools and bedtools (Li et al., 2009; Quinlan and Hall, 2010) were used to parse the alignment results and assess genome coverage.

In addition to the number of reads matching each of the four sets of viral contigs, length of the viral contig was determined. The circularity of each contig was also assessed by looking for at least 10 and at most 1000 bp of overlap between the end and beginning of the contig sequences that were greater than 3500 bp in length.

Entire contig sequences were then compared using BLAST with an e-value threshold 10^{-5} to NCBI reference viral database, nucleotide (NT) database, and bacterial genomes database (downloaded January 7, 2015 from NCBI Refseq). Top matches were recorded. Additional "host" information was given to each contig using the following rules: 1) if a match to a reference viral database was found, the top match was used, 2) if no match to the viral database was found but a match to a bacterial genome was found, the match to the bacterial genome was used, and 3) if no match to the viral or bacterial reference genomes was found but a match was found to the NT database, that match was used.

Contigs were then annotated with open reading frames (ORFs) using Glimmer (Salzberg et al., 1998), a program that utilizes the Interpolated Context Model to predict genes. Minimum gene length was set to 100 bp. The number of predicted ORFs was recorded. Predicted ORFs, in protein form, were BLAST aligned to ORFs of known viral families. The number of ORFs matching predicted ORFs from reference viral family sequences was recorded. A putative viral family was given to each contig based on the viral family with the most ORF matches. For example, if a contig had 4 predicted ORFs, 3 of which matched ORFs in known Siphoviridae viruses and 1 of which matched an ORF in a known Podoviridae virus, the contig would be putatively annotated as belonging to the Siphoviridae family. If only 1 of the 4 predicted ORFs matched an ORF in a known Siphoviridae virus and the remaining 3 ORFs did not have any matches, the contig would be assigned to the Siphoviridae family. In the rare instance were an equal number of ORFs matched two families (e.g. 2 ORFs matched one viral family and another 2 ORFs matched another family), a viral family was selected at random.

Seven contigs contained ORFs that showed a top blast hit to an animal cell virus ORF. In all cases, the matched ORFs were a minority of all ORFs on the contig. In 5 out of 7 cases, the contig as a whole found a better alignment to a phage or bacterial sequence. In the remaining 2 cases, no better matches were found but because the animal cell viruses matched 1 out of 3 or 1 out of 6 contig ORFs, we concluded that these contigs were likely derived from phage or bacteria as well.

In addition to using ORFs for providing putative viral family names, ORFs were aligned to conserved domains of integrase proteins with Pfam (Sonnhammer et al., 1998), the Virulence Factor Database (Chen et al., 2012), and Aclame database of mobile genetic elements (Leplae et al., 2010). The number of matches to each of these databases was recorded.

Predicted ORFs, in protein form, were also compared to the Pfam (Sonnhammer et al., 1998) database of conserved domains with Reversed Position Specific BLAST (RPS-BLAST). Low compositional complexity regions were not masked and default e-value of 0.01 was used for the search. Pfam families were grouped using custom database with Pfam domain identifiers linked to phage regulation function (Minot et al., 2011). Contigs from 1) 4 donor samples, 2) patient 1 pre-FMT, 3) patient 2 pre-FMT, and 4) patient 3 pre-FMT were annotated with phage gene types.

Last, the number of reads from each sample mapping to each contigs were plotted. The size of the circle was set to represent percent of the contig covered by reads from that sample. The colors of each sample were used to distinguish the donor, each patient, and the control samples. Contig names were annotated with length and putative viral family (if any).

Donor contigs in patients during and post-FMT

Contig transfer was assessed separately for each patient and and then for all patients combined. Transfer meant 1) an increase in the number of reads matching a contig in a patient's during FMT and/or post-FMT sample(s) from the patient's pre-FMT sample, 2) at least 5 paired

reads matching the contig during and/or post-FMT and 3) at least 50% coverage of the contig during and/or post-FMT. The number of viral contigs that transferred (as well as the contigs' putative viral family name) was recorded for each patient. Viral contigs, built from the donor samples, could have been seen in a patient's during FMT sample, patient's post-FMT sample or both. The number of transfers as well as putative viral family classification was represented using Sankey plots (using R riverplot package). Contigs that were seen to be transferred in at least 2 of the 3 patients were also recorded. 2x2 contingency tables were built for each viral family and transfer, as well as pfams and transfer. Fisher's exact test, and odds ratio, was used to calculate significance.

Verification of donor contigs

To verify our sequence results, four contigs were selected for verification with qPCR. Prior to qPCR, the sequence data from all of the samples was re-aligned to each contig and visualized using IGV (Robinson et al., 2011). Manual visual inspection confirmed the contig transfer plots, and qPCR assays were developed. After seeing an overrepresentation of Siphoviridae being transferred, 4 contigs specifically annotating as Siphoviridae were assessed for qPCR. Alignment of all sample reads to these contigs was again visualized in IGV and manually confirmed to match the contig transfer plots.

Supplemental Methods References

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I.*, et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. Nature methods *7*, 335-336.
Chen, L., Xiong, Z., Sun, L., Yang, J., and Jin, Q. (2012). VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. Nucleic acids research *40*, D641-645.
Huson, D.H., Mitra, S., Ruscheweyh, H.J., Weber, N., and Schuster, S.C. (2011). Integrative analysis of environmental sequences using MEGAN4. Genome research *21*, 1552-1560.
Kellermayer, R., Nagy-Szakal, D., Harris, R.A., Luna, R.A., Pitashny, M., Schady, D., Mir, S.A., Lopez, M.E., Gilger, M.A., Belmont, J.*, et al.* (2015). Serial fecal microbiota transplantation alters mucosal gene expression in pediatric ulcerative colitis. The American journal of gastroenterology *110*, 604-606.
Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods *9*, 357-359.

Leplae, R., Lima-Mendez, G., and Toussaint, A. (2010). ACLAME: a CLAssification of Mobile genetic Elements, update 2010. Nucleic acids research *38*, D57-61.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.
Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2011). The human gut virome: inter-individual variation and dynamic response to diet. Genome research *21*, 1616-1625.
Norman, J.M., Handley, S.A., Baldridge, M.T., Droit, L., Liu, C.Y., Keller, B.C., Kambal, A., Monaco, C.L., Zhao, G., Fleshner, P*., et al.* (2015). Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell *160*, 447-460.
Peng, Y., Leung, H.C., Yiu, S.M., and Chin, F.Y. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics *28*, 1420-1428.
Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.
Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nature biotechnology *29*, 24-26.
Salzberg, S.L., Delcher, A.L., Kasif, S., and White, O. (1998). Microbial gene identification using interpolated Markov models. Nucleic acids research *26*, 544-548.
Sonnhammer, E.L., Eddy, S.R., Birney, E., Bateman, A., and Durbin, R. (1998). Pfam: multiple sequence alignments and HMM-profiles of protein domains. Nucleic acids research *26*, 320-322.
Treangen, T.J., Sommer, D.D., Angly, F.E., Koren, S., and Pop, M. (2011). Next generation sequence assembly with AMOS. Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al] *Chapter 11*, Unit 11 18.

Supplemental Figures and Tables

Supplemental figures and tables are available online: *Chehoud* et al, MBio 2016.

CHAPTER 7: Conclusion

This dissertation describes changes in the microbial communities associated with disease states, how the microbial communities can be modulated to ameliorate certain disease states, and the possible obstacles of such modulation. Like the microbiome field, in this thesis, we shift from characterizing and associating microorganisms with disease states to investigating the possibilities of engineering communities of microorganisms to alleviate diseases. Microbiota transplantation has been shown to be highly efficacious in treating *C. difficile* infections, but much more work is in progress investigating its uses for other disease. We answer fundamental questions about the long-term durability of transplants in mice and how they are affected by diet. We investigate the use of microbiota transplants in the treatment of hyperammonemia. We also investigate the behavior of viruses during transplantations to assess a potential obstacle. This work serves as an important contribution to the field, particularly as we are just beginning to harness the power of the microbiome to cure disease.

I begin my dissertation by characterizing the microbial community in two contexts. In Chapter 2, targeted sequencing was used to characterize the bacterial, archaeal, and fungal communities in pediatric Inflammatory Bowel Disease (IBD) patients. Similar to previous studies (1-5), decreased bacterial diversity was observed in pediatric IBD patients. Fungal diversity was reported to have decreased in IBD patients compared to healthy controls, and one Candida member, associated with at least five names, including *Pichia jadinii*, *Candida utilis*, *Cyberlindnera jadinii*, *Torulopsis utilis*, and *Torula utilis* (6), was found to be enriched in IBD patients. Only three pediatric IBD patients had detectable archaea in their stool, and all archaeal sequences were members of the *Methanobrevibacter* genus. A random forest classifier was used to partition pediatric samples by disease status. The greatest accuracy of the classifier, 93 percent, could be achieved when the abundance of both bacterial and fungal community members was taken into account. This chapter illustrates the importance of investigating fungal

community members in conjunction with bacterial members, and presents clear evidence that fungal dysbiosis is present in pediatric IBD.

An important question arising from this study is whether this fungal signature, or a combination of fungal and bacterial signatures, is useful as a molecular diagnostic tool. To explore this further, one must first verify that the signatures found in this paper can be reproduced. Recent studies have found the same *Candida* organism, also known as *Pichia jadinii*, to be in more abundant in IBD patients both via tagged-sequencing and metagenomics sequencing (7, 8). It is also important to assess whether the fungal signatures observed in Chapter 2 correlate with parameters such as disease severity, medication success, IBD subtype etc. This would allow us to determine whether fungal blooms are a response to certain IBD treatments or if the fungal community membership can help us distinguish between IBD subtypes. The few Ulcerative Colitis (UC) patients in our study did not allow us to discriminate between the UC and Crohn's Disease (CD) phenotype. The number of subjects in the UC subtype would need to be expanded significantly to better assess differences. If studies continue to support fungal signatures clearly distinguishing between IBD and healthy patients, correlations between the fungal signatures and variables such as age, antibiotic usage, medication, immune cells, need to be performed. This would set the groundwork for follow-up, mechanistic studies that can link fungi and inflammation within our gastrointestinal tract and help answer questions concerning the role of specific yeast lineages in the etiology of IBD.

Chapter 3 discusses the characterization of DNA virus in lung-transplant recipients. A highly abundant and diverse set of anelloviruses was found in the lungs of lung-transplant recipients. The anelloviruses were verified with Q-PCR and found to be 56-fold more abundant in lung transplant recipients compared with healthy controls or HIV+ subjects. Novel anelloviruses were built from the metagenomics data and phylogenetically compared to previously known anelloviruses. This revealed a vast diversity of anelloviruses, even within single individuals.

The primary follow-up to this work is to perform long-term longitudinal studies to determine whether levels of anelloviruses in the lung allograft are associated with, and possibly play a role in, transplant outcomes. Can we use abundance levels of anelloviruses or anellovirus diversity to track and predict lung transplant failures? Is this possible with other organ transplants? A study preceding ours found anelloviruses in serum following organ transplantation (9). They reported that anellovirus genome levels correlated with the extent of immunosuppression, suggesting that anellovirus DNA in blood might serve as an assay for the overall level of immunosuppression. This leads us to ask about the correlation between anellovirus levels in serum and lung allograft. Are there differences in anelloviruses species in different body sites within the same person? In addition to looking at anelloviruses, we must investigate other viruses that may be useful as indicators of immunosuppression. Our current tools for identifying viruses are biased towards DNA viruses and viruses that amplify well with Genomiphi. We need to identify others viruses known to be ubiquitous in humans and assess via targeted Q-PCR how their abundance levels track with immunosuppression. Viral markers can provide a new wave of clinical immune-monitoring strategies for organ transplant recipients.

In Chapters 4 and 5, we replaced the entire gut microbial community of mice with a defined consortium of eight bacteria containing minimal urease gene content, known as Altered Schaedler Flora (ASF). This engineered gut microbiota was able to engraft long-term and reduce fecal urease activity and ammonia production, leading to decreased morbidity and mortality in a murine model of hepatic injury. Since the gut microbiota responds differently to environmental stressors such as diet, we further investigated the effects of dietary protein restriction on the structure and function the microbiota community engineered by inoculation with ASF. We found that community membership after the initial ASF inoculation was similar between mice on a normal protein diet (NPD) and a low protein diet (LPD), but the outgrowth of gut microbiota differed over the ensuing month. Nevertheless, fecal ammonia levels remained significantly lower in ASF-transplanted mice on a LPD compared to mice transplanted with conventional microbiota on a LPD, showing that the functionality of the engineered gut microbiota remained intact. These

findings provide insight into engineering a gut microbiota with a defined bacterial consortium, and, how this is affected by diet.

It is widely known that diet is one of many factors that affect our gut microbiota. Although, we specifically investigated the effect of a LPD versus NPD on the engraftment of an engineered community, further studies investigating other perturbations are warranted. For example, how resilient is the engineered community if a new dose of antibiotics is given without an additional ASF gavage or if a conventional mouse is added to the cage of mice with established engineered communities? How different are the effects of the perturbations as we get farther away from the time of the initial ASF inoculation? How does the engraftment change in mice of different genetic backgrounds? Investigating these questions in the controlled-setting of animal models will help us predict the dynamics of defined communities in humans.

In Chapter 6, we investigated whether viruses also transfer when microbial communities are transferred from one human to another. Since fecal microbiota transplantation (FMT) is a highly successful treatment for relapsing *Clostridium difficile* infection and, potentially, other gastrointestinal diseases, we set out to investigate the transfer of viral communities during FMT in pediatric UC patients. We documented transfer of multiple viral lineages between humans; none of which were viruses that replicated on animal cells or are known to be pathogenic. We found that temperate bacteriophage, which form stable associations with their hosts, were significantly more likely to be transferred during FMT. We conclude that our findings support a model in which the viral temperate replication style may have evolved in part to support efficient viral transmission between environments.

The most important follow-up to the work presented in chapter 6 is to develop better tools to investigate the virome. Most of the viruses we find are novel viruses that cannot be annotated with taxonomic information. Furthermore, we need to improve the methods by which we can match viruses with their bacterial hosts. Our current methods include correlating abundance profile or using genetic homology, CRISPRs, or oligonucleotide profiles (reviewed in (10)). A

recent paper challenged 74 genetically defined Propionibacterium strains with 15 fully-sequenced phages to assess phage infection capabilities (11). We should be able to use sequence data coupled with these types of wet-lab results to build models to predict infectivity of bacteriophage and bacterial host pairs. With enough data (particularly with experiments investigating how single polymorphisms can alter infectivity), we should be able to take viral sequences and assign each nucleotide an importance score. This would allow us to take the data presented in Chapter 6 and predict whether viruses in a donor's stool would survive in a recipient's, as well as use these tools to investigate viral-bacterial relationships in a wide range of applications.

This dissertation examines a broad range of questions relating to the human microbiome. I first discuss characterization of fungal, archaeal, and bacterial communities in pediatric IBD patients and viral communities in lung-transplant recipients. To better understand the dynamics between different bacteria, I describe studies investigating defined microbial communities in animal models and end with a study investigating viral communities during FMT in humans. This entire work took advantage of the capabilities of high-throughput sequencing, which enabled an unprecedented, large-scale investigation of microbial communities. The methods developed and discussed in this dissertation will allow other scientists to investigate the bacterial, archaeal, fungal, and viral communities under a different set of circumstances. The research presented here is meant to serve as a foundation for subsequent studies investigating the dynamic microorganisms that live within us.

References

1. Frank DN, St Amand AL, Feldman RA, et al. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci U S A. 2007;104:13780-13785
2. Gophna U, Sommerfeld K, Gophna S, et al. Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis. J Clin Microbiol. 2006;44:4136-4141
3. Manichanh C, Rigottier-Gois L, Bonnaud E, et al. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. Gut. 2006;55:205-211
4. Martinez-Medina M, Aldeguer X, Gonzalez-Huix F, et al. 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. 2006;12:1136-1145

5. Walker AW, Sanderson JD, Churcher C, et al. High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. BMC Microbiol. 2011;11:7
6. Available at: http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=4903
7. Lewis JD, Chen EZ, Baldassano RN, et al. Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease. Cell Host Microbe. 2015;18:489-500
8. Sokol H, Leducq V, Aschard H, et al. Fungal microbiota dysbiosis in IBD. Gut. 2016
9. De Vlaminck I, Khush KK, Strehl C, et al. Temporal response of the human virome to immunosuppression and antiviral therapy. Cell. 2013;155:1178-1187
10. Edwards RA, McNair K, Faust K, et al. Computational approaches to predict bacteriophage-host relationships. FEMS Microbiol Rev. 2016;40:258-272
11. Liu J, Yan R, Zhong Q, et al. The diversity and host interactions of Propionibacterium acnes bacteriophages on human skin. ISME J. 2015;9:2078-2093

APPENDIX 1: List of Publications

(17) Bryson A*, **Chehoud C*,** Dryga A, Young J, Zost S, Loy E, Chen E, Li H, Roberts R, Minot S, Clark T, Korlach J, Sherrill-Mix S, Bushman F. 2016. *Phage Predation in the Human Gut Microbiome* (in preparation).

(16) **Chehoud C**, Stieh D, Bailey A, Laughlin A, Hope T, Bushman F. 2016. *The Human Microbiome and Socioeconomic Status: Dynamics of the Vaginal Microbiota in Women from two Chicago Hospitals* (in preparation).

(15) Shen T, **Chehoud C**, Ni J, Hsu E, Chen Y, Bailey A, Laughlin A, Bittinger K, Bushman F, Wu G. 2016. *Dietary Regulation of the Gut Microbiota Engineered by a Minimal Defined Bacterial Consortium* (in submission).

(14) Connor K, **Chehoud C**, Altrichter A, DeSantis T, Chan L, Lye S. 2016. *Programming of Immune and Metabolic Status during Pregnancy is Mediated by Maternal Diet and the Gut Microbiome* (in submission).

(13) **Chehoud C**, Dryga A, Hwang Y, Nagy-Szakal D, Hollister E, Luna R, Versalovic J, Kellermayer R, Bushman F. 2016. *Transfer of Viral Communities between Human Individuals during Fecal Microbial Transplantation.* mBio (PMID: 27025251).

(12) Karabucak B, Bunes A, **Chehoud C**, Kohli M, Setzer F. 2016. *Prevalence of Apical Periodontitis in Endodontically Treated Premolars and Molars with Untreated Canal: A Cone-beam Computed Tomography Study*. Journal of Endodontics (PMID: 26873567).

(11) Lewis J*, Chen E*, Baldassano R, Otley A, Griffiths A, Lee D, Bittinger K, Bailey A, Friedman E, Hoffmann C, Albenberg L, Sinha R, Compher C, Nessel L, Grant A, **Chehoud C**, Li H, Wu G, Bushman F. 2015. *Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease.* Cell Host and Microbe (PMID: 26468751).

(10) Shen T*, Albenberg L*, Bittinger K, **Chehoud C**, Chen Y, Judge C, Wang L, Sheng M, Lin A, Wilkins B, Lewis J, Daikhin Y, Nissim I, Yudkoff M, Bushman F, Wu G. 2015. *Engineering the Gut Microbiota to Treat Hyperammonemia.* Journal of Clinical Investigation (PMID: 26098218).

(9) **Chehoud C***, Albenberg L*, Judge C, Hoffmann C, Grunberg S, Bittinger K, Baldassano R, Lewis J, Bushman F, Wu G. 2015. *Fungal Signature in the Gut Microbiota of Pediatric Patients with IBD.* Inflammatory Bowel Diseases Journal (PMID: 26083617).

(8) Lee D, Baldassano R, Otley A, Albenberg L, Griffiths A, Compher C, Chen E, Li H, Gilroy E, Nessel L, Grant A, **Chehoud C**, Bushman F, Wu G, Lewis J. 2015. *Comparative Effectiveness of Nutritional and Biologic Therapy in North American Children with Active Crohn's Disease.* Inflammatory Bowel Diseases Journal (PMID: 25970545).

(7) Cantarel B, Waubant E, **Chehoud C**, Kuczynski J, DeSantis T, Warrington J, Venkatesan A, Fraser C, Mowry E. 2014. *Gut Microbiota in MS: Possible Influence of Immunomodulators.* Journal of Investigative Medicine (PMID: 25775034).

(6) Wu G, Compher C, Chen E, Smith S, Shah R, Bittinger K, **Chehoud C**, Albenberg L, Nessel L, Gilroy E, Star J, Weljie A, Flint H, Metz D, Bennett M, Li H, Bushman F, Lewis J. 2014. *Comparative Metabolomics in Vegans and Omnivores Reveal Constraints on Diet-Dependent Gut Microbiota Metabolite Production.* Gut (PMID: 25431456).

(5) Young J, **Chehoud C**, Bittinger K, Bailey A, Diamond J, Cantu E, Haas A, Abbas A, Frye L, Christie J, Bushman F, Collman R. 2014. *Blooms of Anelloviruses in the Respiratory Tract of Lung Transplant Recipients.* American Journal of Transplantation (PMID: 25403800).

(4) Miezeiewski M, Schnaufer T, Muravsky M, Wang S, Caro-Aguilar I, Secore S, Thiriot D, Hsu C, Rogers I, DeSantis T, Kuczynski J, Probst A, **Chehoud C**, Steger R, Warrington J, Bodmer J, Heinrichs J. 2014. *An in vitro Culture Model to Study the Dynamics of Colonic Microbiota in Syrian Golden Hamsters and their Susceptibility to Infection with Clostridium difficile.* Multidisciplinary Journal of Microbial Ecology (PMID: 25036923).

(3) **Chehoud C**, Rafail S, Tyldsley AS, Seykora JT, Lambris JD, Grice EA. 2013. *Complement Modulates the Cutaneous Microbiome and Inflammatory Milieu.* Proceedings of the National Academy of Sciences (PMID: 23980152).

(2) Minot S, Bryson A, **Chehoud C**, Wu GD, Lewis JD, Bushman FD. 2013. *Rapid Evolution of the Human Gut Virome.* Proceedings of the National Academy of Sciences (PMID: 23836644).

(1) Noval Rivas M, Burton OT, Wise P, Zhang YQ, Hobson SA, Garcia Lloret M, **Chehoud C**, Kuczynski J, DeSantis T, Warrington J, Hyde ER, Petrosino JF, Gerber GK, Bry L, Oettgen HC, Mazmanian SK, Chatila TA. 2013. *A Microbiota Signature Associated with Experimental Food Allergy Promotes Allergic Sensitization and Anaphylaxis.* The Journal of Allergy and Clinical Immunology (PMID: 23201093).