UNIVERSITY OF CALIFORNIA,
IRVINE

Ecological and evolutionary drivers of microbial community assembly: The influence of host, diet, and richness on *Bifidobacterium*

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Cynthia Itzel Rodriguez

Dissertation Committee:
Professor Jennifer B.H. Martiny, Chair
Associate Professor Katrine Whiteson
Professor Brandon Gaut

2022

DEDICATION

To

my parents, sister, and mentors

in recognition of your support, encouragement, and love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First, I want to express my gratitude to my Principal Investigator, Dr. Jennifer Martiny, for serving as a guide and role model throughout my graduate career. Your intelligence, insightful advice, support, and expertise as a microbial ecologist have played a fundamental role in my academic journey. In addition, your caring and compassionate character make you an extraordinary mentor and I feel beyond privileged to have been in your lab for the past four years and a half.

Thank you to my committee members, Dr. Katrine Whiteson and Dr. Brandon Gaut, for always being enthusiastic while listening to my work and making the time to meet with me. Your help, feedback, and constant support are much valued and will never be forgotten.

To the members of the J. Martiny lab and Microbial Group, I am thankful for your presence and for making a lab environment that was enjoyable and productive. Our lunches, happy hours, and outings were key in my development. To Zain Alhashem, thank you for your constant scientific and emotional support, I appreciate all the times you lent me an ear to listen to my frustrations, your presence in my life has been invaluable.

To my sister, Tania Rodriguez, I want to thank you for your love, understanding, and help throughout my Ph.D. You have helped me emotionally by being a great listener and companion and with my lab work by labeling hundreds of vials and staying up with me multiple nights. I am thankful for our accountability hours over zoom and in person and for your encouragement. Having you as my younger sister is a true blessing.

To my parents, I am very grateful to you for always believing in me and being proud of me. I appreciate all the support you have given me since I was little to pursue an education and will always remember our study sessions and early morning drives to school. I am very thankful to you for everything you have sacrificed by coming to this country to give me and my sister a better life, none of this would be possible without your hard work and perseverance.

Lastly, I would like to thank all the people in the department and university that have supported my graduate career. Dr. Mike Mulligan and Dr. Kathleen Treseder thank you for helping in my transition into the J. Martiny lab.

This work was funded by the Faculty Mentor Program fellowship, the NIH T32 Pre-doctoral Fellowship for Training in Microbiology & Infectious Diseases (Award# AI141346; NIH T32AI141346), the UC President's Dissertation Year Fellowship, the Rose Hills Foundation Science & Engineering Fellowship, and the UCI Microbiomes Initiative Center.

The text of **Chapter 1** of this dissertation is a reprint of the material as it appears in BMC Genomics. This material is distributed under the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which "permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made." Jennifer B.H. Martiny, a co-author listed in this publication, directed and supervised the research, which forms the basis for this dissertation. No changes were made to the original text, except for the exclusion of Additional files 1 and 2, those can be found in the online publication only.

Rodriguez, C.I., Martiny, J.B.H. Evolutionary relationships among bifidobacteria and their hosts and environments. BMC Genomics 21, 26 (2020). https://doi.org/10.1186/s12864-019-6435-1

VITA

Cynthia Itzel Rodriguez

EDUCATION

2022	Ph.D., Biological Sciences, University of California, Irvine (UCI) *Advisor: Jennifer Martiny, Ph.D.*
2019	M.S., Biological Sciences, University of California, Irvine (UCI)
2016	B.S., Biological Sciences, University of California, Irvine (UCI)

PUBLICATIONS

J. Phan, J. Kapcia III, **C.I. Rodriguez**, V.L. Voguel, S.J.B. Dunham, K. Whiteson. 2022. Capturing actively produced microbial volatile organic compounds from human associated samples with vacuum assisted sorbent extraction. JoVE. doi: 10.3791/62547

C.I. Rodriguez and J.B.H Martiny. 2020. Evolutionary relationships among bifidobacteria and their hosts and environments. BMC Genomics. doi: 10.1186/s12864-019-6435-1

SELECTED FELLOWSHIPS AND AWARDS

2022	Summer Teaching Apprenticeship Program Fellowship (STAP)
2021	Rose Hills Foundation Science & Engineering Fellowship
2021	UCI Associated Graduate Students Conference Grant
2021	UC President's Dissertation Year Fellowship
2020	Rose Hills Foundation Science & Engineering Fellowship
2020	Registration Scholarship for 2020 SACNAS Conference
2020-2021	NIH T32 Pre-doctoral Fellowship for Training in Microbiology & Infectious Diseases
2019-2020	Faculty Mentor Program Fellowship (FMP)

SELECTED TEACHING EXPERIENCES

2022	Instructor of record for E134 Microbiomes, UCI - Ecology and Evolutionary Biology Department
2019	Instructor for Microbial Community: PCR Workshop, UCI-Microbiome Initiative
2019	Instructor for Characterizing Microbiome Diversity: DNA extraction Workshop, UCI-Microbiome Initiative
2017-2018	Teaching assistant for Bio94 From Organisms to Ecosystems, UCI-Ecology and Evolutionary Biology Department

SELECTED PROFESSIONAL PRESENTATIONS

2022	**Oral presenter**, International Society for Microbial Ecology Symposium (ISME)- Lausanne, Switzerland
2022	**Oral presenter**, American Society for Microbiology (ASM)- Washington, D.C.
2021	**Oral Presenter**, Ecological Society of America (ESA)- Virtual
2021	**Convener**, ASM World Microbiome Forum– Session: Ecological and Evolutionary Theory of Host and Microbiomes - Virtual
2020	**Poster Presenter**, Society for Advancement of Chicanos/Hispanics and Native Americans in Science (SACNAS)- Virtual
2020	**Poster Presenter**, Microbial Ecology and Evolution (MEE)- Virtual
2018	**Poster Presenter**, International Society for Microbial Ecology Symposium (ISME)- Leipzig, Germany

SELECTED OUTREACH EXPERIENCES

2022	Speaker for COSMOS program, Irvine, CA
2018-2022	Judge for the Irvine Unified School District Annual Science Fair, Irvine, CA
2022	Speaker for 21st Century Career Conference, Irvine Unified School District, CA
2021	Judge for the World Microbe Forum iPosters, ASM and FEMS collaboration, virtual
2021	Speaker for VASE program, virtual
2020	Sharing Opportunities and Insights for a Successful Academic Journey, Saddleback College for The Transfer Mentor Program, Saddleback, CA
2018-2020	Ask-A-Scientist-Night-Irvine Unified School District, Irvine, CA
2019	Science lab tours with Godinez Fundamental High School (SACNAS), Irvine, CA
2018-2019	Instructor for the Tech Trek Program- Encouraging middle school girls to join STEM, Irvine, CA
2018	Science demonstration to Boyle Heights children's club (SACNAS), Boyle Heights, CA
2018	Moderator for workshop Reclaiming STEM (SACNAS), Irvine, CA

PROFESSIONAL CERTIFICATIONS

2021	Transforming Your Research into Teaching Course Access
2021	Mentoring Excellence Program
2018	Improv for Teaching Program
2018	Research Mentoring Certificate

ABSTRACT OF THE DISSERTATION

Ecological and evolutionary drivers of microbial community assembly: The influence of host, diet, and richness on *Bifidobacterium*

by

Cynthia Itzel Rodriguez

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2022

Professor Jennifer B.H. Martiny, Chair

Humans harbor a diverse microbiome in their gut including a diversity of bifidobacteria species. The bacterial group bifidobacteria (genus *Bifidobacterium*) is fundamental to the health of humans and animals. They are inhabitants of the gastrointestinal tract, vagina, and mouth of mammals and some insects. Bifidobacteria have gained notable attention as their presence in the gut has been correlated with many health-promoting benefits. While human bifidobacteria have been well studied, most focus on one strain at a time, without considering the variety that coexist in the gut. My thesis focused on the ecological and evolutionary forces driving the assembly, coexistence, and functioning of bifidobacterial diversity. Using a multidisciplinary approach that encompasses ecological and evolutionary theory, genomics, and microbial metabolism, my objectives were to: 1) uncover the trait and evolutionary associations bifidobacteria have with their animal hosts (Chapter 1); 2) assess bacterial responses to dietary fiber consumption in the human gut (Chapter 2); and 3) identify the functional consequences of bifidobacteria diversity (Chapter 3).

To address objective 1, I used a comparative genomics approach to investigate the adaptation of bifidobacteria to their hosts. I analyzed all the bifidobacteria genomes available in The National Center for Biotechnology Information (NCBI) repository. I identified the hosts from which the bifidobacteria strains were isolated, performed a multilocus phylogenetic analysis, compared the genetic relatedness of the strains to different hosts, and tested the degree to which variation in traits can be attributed to their hosts. I found that different species of bifidobacteria colonized different animal hosts and that traits related to fiber degradation were associated with particular hosts.

To tackle objective 2, I conducted a meta-analysis that included dietary fiber interventions that have examined the human gut microbiome (n=21). By synthesizing, reanalyzing, and conducting an in-depth phylogenetic analysis, I found consistent bacterial responses to short-term increases in dietary fiber consumption in healthy humans. Specifically, I found that fiber interventions decreased bacterial diversity and explained an average of 1.5% of compositional variation. Moreover, I identified specific bacterial taxa that responded to dietary fiber in humans. One taxon that drastically increased in response to fiber consumption across interventions was bifidobacteria.

To address objective 3, I tested how the diversity and richness of bifidobacteria isolates influence their coexistence and functioning (i.e., fiber degradation) using laboratory experiments. To do this, I obtained bifidobacteria isolates from the Human Microbiome Project (HMP), and I selectively isolated new strains from fecal samples collected at the University of California, Irvine. I conducted microcosm experiments that vary the diversity of bifidobacteria and characterized the functioning of the bacterial communities using flow cytometry (to count bacterial cells), lactate assays (to measure metabolite concentration), and next-generation

sequencing (to decipher the bacterial community's composition). I found that the diversity of isolates persisted, such that up to 7 strains coexisted. Moreover, metabolite production (e.g., lactate) increased with increasing bifidobacteria diversity. However, this was not the case for biomass production.

Focusing on the ecology and evolution of an important taxon like bifidobacteria within the diverse gut community will provide a deeper understanding on the community assembly mechanisms that bacteria use to colonize specific hosts.

INTRODUCTION

The human gut microbiome has shown to mediate the regulation of many human processes including intestinal integrity, digestion, host immunity, protection against pathogens, harvesting and production of nutrients, among others (Thursby and Juge 2017). Even activities known to be beneficial to us, like exercise, have shown to provide benefits by modifying the composition of gut microbes (Clauss et al. 2021; Monda et al. 2017). In contrast, an altered gut microbiome (dysbiosis) is often correlated to health problems such as inflammation, obesity, diabetes, cancer, and reduced immunity (Rastelli, Knauf, and Cani 2018; Fan and Pedersen 2021; Shreiner, Kao, and Young 2015). However, despite the importance of the gut microbiome, the mechanisms that promote the assembly of a healthy microbial community and its functioning are not fully understood. Thus, the aim of my thesis is to investigate the ecology and evolution of microbial diversity in the gut using a model group of bacteria.

The four major bacterial phyla in the human gut microbiome are Proteobacteria, Firmicutes, Actinobacteria, and Bacteroidetes (Thursby and Juge 2017; Turnbaugh et al. 2007). Within the phylum Actinobacteria, one bacterial group that has proven to be of particular importance for health is bifidobacteria. These are gram-positive, anaerobic, saccharolytic, irregular-shaped bacteria, members of the genus *Bifidobacterium* (Turroni et al. 2018). Bifidobacteria usually reside in the gastrointestinal tract, vagina, and mouth of mammals, including humans and some insects (Milani et al. 2014; Turroni, van Sinderen, and Ventura 2011). However, they have also been found in human blood, breast milk, and sewage (Esaiassen et al. 2017; Lamendella et al. 2008; Martín et al. 2009).

Bifidobacteria were first isolated in 1899 by Henri Tissier from feces of breast-fed infants (Tissier, H. 1899). Since then, they are often associated with newborns as specific species can

degrade human milk oligosaccharides (HMOs) derived from breast milk (Ruiz-Moyano et al. 2013; Thomson, Medina, and Garrido 2018). However, adults also have bifidobacteria in their gut but at lower levels than breast-fed infants and with different species' composition (Arboleya et al. 2016). Bifidobacteria have been correlated with health-promoting benefits including the including the production of beneficial metabolites (vitamins, antioxidants, and short-chain fatty acids), immune system development, protection from certain gut diseases (enterocolitis and acute diarrhea), and degradation of fiber (O'Callaghan and van Sinderen 2016). Because of these positive associations, bifidobacteria are a popular probiotic added to yogurt and dietary supplements. I am centering my projects on *Bifidobacterium* because this genus is abundant and widespread in mammals and has important health benefits to humans.

My thesis focuses on the ecological and evolutionary forces driving the assembly, coexistence, and functioning of bifidobacteria. Using a multidisciplinary approach that encompasses ecological and evolutionary theory, genomics, and microbial metabolism, my objectives were to: 1) uncover the trait and evolutionary associations bifidobacteria have with their animal hosts (Chapter 1); 2) assess bacterial responses to dietary fiber consumption in the human gut (Chapter 2); and 3) identify the functional consequences of bifidobacteria diversity (Chapter 3). Studying the ecology and evolution of bifidobacteria as a diverse group provides an opportunity to uncover general mechanisms that determine bacterial diversity in the gut as well as their specific role in human health.

REFERENCES

Arboleya, Silvia, Claire Watkins, Catherine Stanton, and R. Paul Ross. 2016. "Gut Bifidobacteria Populations in Human Health and Aging." *Frontiers in Microbiology* 7 (August). https://doi.org/10.3389/fmicb.2016.01204.

Clauss, Matthieu, Philippe Gérard, Alexis Mosca, and Marion Leclerc. 2021. "Interplay Between Exercise and Gut Microbiome in the Context of Human Health and Performance." *Frontiers in Nutrition* 8. https://www.frontiersin.org/articles/10.3389/fnut.2021.637010.

Esaiassen, Eirin, Erik Hjerde, Jorunn Pauline Cavanagh, Gunnar Skov Simonsen, and Claus Klingenberg. 2017. "Bifidobacterium Bacteremia: Clinical Characteristics and a Genomic Approach To Assess Pathogenicity." *Journal of Clinical Microbiology* 55 (7): 2234–48. https://doi.org/10.1128/JCM.00150-17.

Fan, Yong, and Oluf Pedersen. 2021. "Gut Microbiota in Human Metabolic Health and Disease." *Nature Reviews Microbiology* 19 (1): 55–71. https://doi.org/10.1038/s41579-020-0433-9.

Lamendella, Regina, Jorge W. Santo Domingo, Catherine Kelty, and Daniel B. Oerther. 2008. "Bifidobacteria in Feces and Environmental Waters." *Applied and Environmental Microbiology* 74 (3): 575–84. https://doi.org/10.1128/AEM.01221-07.

Martín, Rocío, Esther Jiménez, Hans Heilig, Leonides Fernández, María L. Marín, Erwin G. Zoetendal, and Juan M. Rodríguez. 2009. "Isolation of Bifidobacteria from Breast Milk and Assessment of the Bifidobacterial Population by PCR-Denaturing Gradient Gel Electrophoresis and Quantitative Real-Time PCR." *Applied and Environmental Microbiology* 75 (4): 965–69. https://doi.org/10.1128/AEM.02063-08.

Milani, Christian, Gabriele Andrea Lugli, Sabrina Duranti, Francesca Turroni, Francesca Bottacini, Marta Mangifesta, Borja Sanchez, et al. 2014. "Genomic Encyclopedia of Type Strains of the Genus Bifidobacterium." *Applied and Environmental Microbiology* 80 (20): 6290–6302. https://doi.org/10.1128/AEM.02308-14.

Monda, Vincenzo, Ines Villano, Antonietta Messina, Anna Valenzano, Teresa Esposito, Fiorenzo Moscatelli, Andrea Viggiano, et al. 2017. "Exercise Modifies the Gut Microbiota with Positive Health Effects." *Oxidative Medicine and Cellular Longevity* 2017. https://doi.org/10.1155/2017/3831972.

O'Callaghan, Amy, and Douwe van Sinderen. 2016. "Bifidobacteria and Their Role as Members of the Human Gut Microbiota." *Frontiers in Microbiology* 7 (June). https://doi.org/10.3389/fmicb.2016.00925.

Rastelli, Marialetizia, Claude Knauf, and Patrice D. Cani. 2018. "Gut Microbes and Health: A Focus on the Mechanisms Linking Microbes, Obesity, and Related Disorders." *Obesity (Silver Spring, Md.)* 26 (5): 792–800. https://doi.org/10.1002/oby.22175.

Ruiz-Moyano, Santiago, Sarah M. Totten, Daniel A. Garrido, Jennifer T. Smilowitz, J. Bruce German, Carlito B. Lebrilla, and David A. Mills. 2013. "Variation in Consumption of Human Milk Oligosaccharides by Infant Gut-Associated Strains of Bifidobacterium Breve." *Appl. Environ. Microbiol.* 79 (19): 6040–49. https://doi.org/10.1128/AEM.01843-13.

Shreiner, Andrew B., John Y. Kao, and Vincent B. Young. 2015. "The Gut Microbiome in Health and in Disease." *Current Opinion in Gastroenterology* 31 (1): 69–75. https://doi.org/10.1097/MOG.0000000000000139.

Thomson, Pamela, Daniel A. Medina, and Daniel Garrido. 2018. "Human Milk Oligosaccharides and Infant Gut Bifidobacteria: Molecular Strategies for Their Utilization." *Food Microbiology*, New Tools to Detect and Prevent Foodborne Outbreaks from Farm to Fork, 75 (October): 37–46. https://doi.org/10.1016/j.fm.2017.09.001.

Thursby, Elizabeth, and Nathalie Juge. 2017. "Introduction to the Human Gut Microbiota." *Biochemical Journal* 474 (11): 1823–36. https://doi.org/10.1042/BCJ20160510.

Tissier, H. 1899. Le bacterium coli et la reaction chromophile d'Escherich. Crit. Rev. Soc. Biol. 51:943-945.

Turnbaugh, Peter J., Ruth E. Ley, Micah Hamady, Claire M. Fraser-Liggett, Rob Knight, and Jeffrey I. Gordon. 2007. "The Human Microbiome Project." *Nature* 449 (7164): 804–10. https://doi.org/10.1038/nature06244.

Turroni, Francesca, Christian Milani, Douwe Van Sinderen, and Marco Ventura. 2018. "Chapter 12 - Bifidobacteria: Ecology and Coevolution With the Host." In *The Bifidobacteria and Related Organisms*, edited by Paola Mattarelli, Bruno Biavati, Wilhelm H. Holzapfel, and Brian J. B. Wood, 213–20. Academic Press. https://doi.org/10.1016/B978-0-12-805060-6.00012-0.

Turroni, Francesca, Douwe van Sinderen, and Marco Ventura. 2011. "Genomics and Ecological Overview of the Genus Bifidobacterium." *International Journal of Food Microbiology*, 3rd International Symposium on Propionibacteria and Bifidobacteria: Dairy and Probiotic applications, Oviedo 1-4 June 2010, 149 (1): 37–44. https://doi.org/10.1016/j.ijfoodmicro.2010.12.010.

CHAPTER 1

Evolutionary relationships among bifidobacteria and their hosts and environments

ABSTRACT

The assembly of animal microbiomes is influenced by multiple environmental factors and host genetics, although the relative importance of these factors remains unclear. Bifidobacteria (genus *Bifidobacterium,* phylum Actinobacteria) are common first colonizers of gut microbiomes in humans and inhabit other mammals, social insects, food, and sewages. In humans, the presence of bifidobacteria in the gut has been correlated with health-promoting benefits. Here, we compared the genome sequences of a subset of the over 400 *Bifidobacterium* strains publicly available to investigate the adaptation of bifidobacteria diversity. We tested 1) whether bifidobacteria show a phylogenetic signal with their isolation sources (hosts and environments) and 2) whether key traits encoded by the bifidobacteria genomes depend on the host or environment from which they were isolated. We analyzed *Bifidobacterium* genomes available in the PATRIC and NCBI repositories and identified the hosts and/or environment from which they were isolated. A multilocus phylogenetic analysis was conducted to compare the genetic relatedness the strains harbored by different hosts and environments. Furthermore, we examined differences in genomic traits and genes related to amino acid biosynthesis and degradation of carbohydrates. We found that bifidobacteria diversity appears to have evolved with their hosts as strains isolated from the same host were non-randomly associated with their phylogenetic relatedness. Moreover, bifidobacteria isolated from different sources displayed differences in genomic traits such as genome size and accessory gene composition and on particular traits related to amino acid production and degradation of carbohydrates. In contrast, when analyzing

diversity within human-derived bifidobacteria, we observed no phylogenetic signal or differences on specific traits (amino acid biosynthesis genes and CAZymes).

Overall, our study shows that bifidobacteria diversity is strongly adapted to specific hosts and environments and that several genomic traits were associated with their isolation sources. However, this signal is not observed in human strains alone. Looking into the genomic signatures of bifidobacteria strains in different environments can give insights into how this bacterial group adapts to their environment and what types of traits are important for these adaptations.

INTRODUCTION

Bacteria are central to the evolution and ecology of animals influencing their genomes, development, and physiology (McFall-Ngai et al. 2013). The composition of bacterial communities in the animal gut are thought to be shaped by host physiology and diet on daily timescales, but also by host evolutionary history over much longer timescales (David, Materna, et al. 2014; David, Maurice, et al. 2014; Muegge et al. 2011). A major challenge in animal microbiome research is therefore to disentangle the ecological and evolutionary processes underlying the variation in gut communities. One approach to tackling these questions is to focus on a specific bacterial group within the larger gut community (Groussin et al. 2017; Moeller et al. 2016).

A widespread and abundant group of bacteria in mammalian guts is bifidobacteria. Bifidobacteria are gram-positive, anaerobic, saccharolytic bacteria, members of the genus *Bifidobacterium* of the phylum Actinobacteria (Klijn, Mercenier, and Arigoni 2005). Their presence in the gut has been correlated with health-promoting benefits in humans and mouse models including the production of metabolites like vitamins and antioxidants, immune system development, and protection from certain gut diseases such as enterocolitis and acute diarrhea

(O'Callaghan and van Sinderen 2016). In newborns, specific species of bifidobacteria are important for degrading human milk oligosaccharides (HMOs) derived from breast milk (Ruiz-Moyano et al. 2013; LoCascio et al. 2007). The fermentation of HMOs promotes the wellness of infants and prevents colonization from potential pathogenic bacteria (Bode 2012; Marcobal and Sonnenburg 2012). Bifidobacteria also excel at degrading and fermenting carbohydrates (Liu et al. 2015; Rivière et al. 2014). This process produces short-chain fatty acids (SCFAs) such as butyrate, acetate, and propionate, which have been linked to reducing the risk of inflammatory diseases, heart disease, type II diabetes, and other adverse conditions such as cancer (Slavin 2013).

Here, we take a comparative genomics approach to investigate the relationship between bifidobacteria diversity and their hosts and environments. Bifidobacteria are ubiquitous inhabitants of the gastrointestinal tract, vagina, and mouth of mammals, including humans and are also present in guts of insects such as bees (Milani et al. 2014; Turroni, van Sinderen, and Ventura 2011). They have also been found in human blood, breast milk, and sewage (Esaiassen et al. 2017; Lamendella et al. 2008; Martín et al. 2009). The genomic signatures of bifidobacteria strains in different environments can give insights into how this bacterial group adapts to their environment and what types of traits are important for these adaptations. The few studies that have considered the association between bifidobacteria diversity and their hosts and environments have found contradictory results. Some studies observe no relationship between hosts and the type of genes bifidobacteria carry (Milani et al. 2017; Freitas and Hill 2018), while others do (Sharma, Mobeen, and Prakash 2018; Sun et al. 2015; Turroni et al. 2018).

We analyzed a subset of the 400 bifidobacteria genomes publicly available to answer two questions: 1) Do bifidobacteria show a phylogenetic signal with their isolation sources (hosts and

environments)? and 2) Do key traits encoded by the bifidobacteria genomes depend on the host or environment from which they were isolated? The term "phylogenetic signal" generally refers to the tendency of related species to resemble one another more than they would resemble a species drawn randomly from the same phylogenetic tree (Münkemüller et al. 2012; Kamilar and Cooper 2013).

Since most bacterial traits are phylogenetically conserved (Martiny, Treseder, and Pusch 2013), our first hypothesis was that bifidobacteria are adapted to the hosts (and other environments) from which they are isolated. We predicted that this adaptation would be reflected in the phylogeny of bifidobacteria, despite horizontal gene transfer (HGT) and rapid evolution. Secondly, we hypothesized that bifidobacteria strains would further adapt to their environment through genomic signatures like genome size and overall composition of accessory genes, as well as the composition of particular traits. Genome size is broadly associated with different bacterial lifestyles (Cobo-Simón and Tamames 2017; McCutcheon and Moran 2012; Dini-Andreote et al. 2012), and accessory gene composition can capture horizontally transferred regions of the genome, which are thought to allow for rapid adaptation to a specific environment (Hall, Brockhurst, and Harrison 2017). We specifically focused on two particular classes of genes: amino acid biosynthesis genes and carbohydrate-active enzymes (CAZymes). The abundance and diversity of amino acid biosynthesis genes may vary as amino acids can be exchanged between different hosts and bacteria (Neis, Dejong, and Rensen 2015; Graf and Ruby 1998), allowing for the loss or gain of these genes. Bacterial CAZyme profiles are also known to vary by environment, suggesting a mechanism for bacteria to adapt to the local carbohydrate supply (Cantarel, Lombard, and Henrissat 2012; Berlemont and Martiny 2016). Moreover,

bifidobacteria are key degraders of carbohydrates in host guts, and we expected that strains might adapt to host diet.

RESULTS

Phylogenetic relationships between bifidobacteria strains and isolation sources

To investigate the phylogenetic relationships between bifidobacteria strains isolated from different environments and hosts, two phylogenetic trees were constructed based on 107 concatenated core genes. These trees included one with 60 human-derived strains (**Fig. 1.1A**) and one with 129 strains from different environments and hosts (**Fig. 1.1B**). In both trees, members of the same taxonomic species clustered closely, and the phylogenetic structure of the trees was similar to previous reports based on 16S rRNA sequences and based on various core genes (Milani et al. 2014; Sun et al. 2015; Ventura et al. 2006; Lugli et al. 2014; Turroni, Berry, and Ventura 2017). For instance, *B. breve* and *B. longum* strains were found to be closely related as well as *B. bifidum* and *B. scardovii*. One difference was that the *B. asteroides* phylogroup has been previously shown to be positioned in the deepest branches of the bifidobacteria lineage (Milani et al. 2014; Sun et al. 2015; Bottacini et al. 2012); however, in our human-derived strains phylogenetic tree the deepest branch corresponded to a member of the *B. thermophilum* species. In the larger tree, the deepest branches corresponded to strains from the *B. simiarum, B. primatium, B. vansinderenii*, and *B. tissieri* species followed by *B. asteroides* group.

The strains isolated from a variety of human stages and body locations showed no phylogenetic signal (ANOSIM: $R= 0.022$, $p>0.05$). For example, strains isolated from infants were not more genetically similar to one another than those isolated from adults (**Fig. 1.1A**). Similarly, strains isolated from the blood were not more genetically similar to one another than those found in milk or in the urogenital tract.

By contrast, when comparing across multiple host species and environments, the habitat from which the strains were isolated was strongly associated with the bacteria's phylogenetic distribution (**Fig. 1.1B;** ANOSIM: R= 0.420, p<0.001). For instance, bee, primate, and rodent derived strains are tightly clustered in the phylogenetic tree within their categories (**Fig. 1.1B**). These broader evolutionary patterns seem particularly robust for strains isolated from the orders Artiodactyla (pig and cattle-derived strains), Hymenoptera (bee-derived strains), and Primates (human and non-human primate-derived strains) as they clustered mostly within the same branches (**Fig. 1.1B**).

Genomic features and content among isolation sources

Genome size analysis

Within the human-derived strains, genome size did not differ by the particular human habitat (e.g., urogenital or gut) or between different human stages (e.g., infant or elderly) (**Fig. 1.2A**; Kruskal-Wallis H = 10.428, p>0.05, df = 7). Conversely, strains isolated from diverse animal hosts and environments (e.g. primates, bees, wastewater, etc.) differed significantly in genome size (**Fig. 1.2B**; H = 26.244, p<0.01, df = 9). Strains isolated from non-human primates had the highest genome size (2.9 Mb $\pm$ 0.19 SD), whereas strains isolated from bees had the lowest genome size (2.0 Mb $\pm$ 0.21 SD).

Pangenome analysis

The analysis on 129 bifidobacteria strains revealed that their pangenome is composed of 438 core genes, 115 soft core genes, 1,802 shell genes, and 24,550 cloud genes, for a total of 26,905 gene clusters (**Fig. 1.3)**. This resonates with previous studies with fewer genomes that found this genus to have between 400-500 core genes (Milani et al. 2014; Sun et al. 2015). The composition of accessory genes excluding the core genome and singletons (~6,400 genes), was associated

with both the bacteria's isolation source (ANOSIM: R=0.394, p<0.001), and the phylogeny of the bifidobacteria strains (based on 107 core genes; RELATE test, Spearman's ϱ =0.52, p<0.001).

Amino acid biosynthesis analysis

Beyond general genomic characteristics, we investigated how a variety of specific traits, such as amino acid biosynthesis genes varied among the strains. There was a significant difference in abundance of amino acid biosynthesis genes between different animal hosts and environments (**Fig. 1.4A**; H = 62.216, p<0.001, df = 11) (*post hoc* Dunn's test). For instance, bees showed the lowest abundance of amino acid biosynthesis genes (87 genes $\pm$ 13 SD) while non-human primates showed the highest number (100 genes $\pm$ 2.9 SD) (**Fig. 1.4A**).

Furthermore, the diversity of amino acid biosynthesis genes also differed among hosts and environments (**Fig. 1.4B**; H = 76.594, p<0.001, df = 11) (*post hoc* Dunn's test); the bee-derived strains showed the lowest diversity of amino acid biosynthesis genes (78 genes $\pm$ 12 SD). Strains isolated from the other host categories carried between 86 and 90 genes (**Fig. 1.4B**).

Carbohydrate-active enzymes (CAZymes)

Since bifidobacteria are known to be excellent degraders of complex carbohydrates, we also searched for CAZymes in their genomes. On the one hand, the abundance of CAZymes among the different human-derived strains did not differ significantly (**Fig. 1.5A**; H = 9.6557, p>0.5, df = 7). On the other hand, when comparing strains derived across different hosts and environments, we found a significant difference between categories (**Fig. 1.5B**; H = 60.9, p<0.001, df = 11). In the human environments, the oral-derived strains encoded the highest number of CAZymes (103 genes $\pm$ 2.8 SD), whereas strains derived from adults (gut-derived) encoded the lowest number (55.8 genes $\pm$ 12 SD). Across all hosts and environments, non-

human primates carried more CAZymes than any other host (84 genes ± 20 SD), while wastewater exhibited the fewest (42 genes ± 10 SD) (**Fig. 1.5**).

DISCUSSION

Studying the diversity of bifidobacteria and their trait associations provides insights into the mechanisms that underlie their assembly within a larger microbial community. Bifidobacteria strains isolated from the same host or environment were non-randomly associated with their phylogenetic relatedness. This pattern is consistent with the hypothesis that bifidobacteria specialize, or at least prefer, particular hosts, in agreement with several other studies (Sun et al. 2015; Lamendella et al. 2008; Milani et al. 2016). For example, Lamendella *et al*. (Lamendella et al. 2008) found that bifidobacteria strains from the same host, including those isolated from birds and pigs, tended to cluster by clade. We also observed that all *B. pseudolongum* subsp. *pseudolongum* strains were isolated from pigs as previously noted (Lugli et al. 2019). Similarly, bee-derived bifidobacteria clustered within two relatively deep branches (Sun et al. 2015). Notably, this clustering was not perfect; for instance, some primate-derived strains clustered with more ancient branches than the bee-derived strains, and rodent-isolated strains could be found within several clades. This pattern of imperfect clustering suggests that host-specialization of bifidobacteria has occurred several times within different branches of the genus. In addition, the clades of strains from mixed isolation sources may indicate that many bifidobacteria are not strict specialists but are capable of colonizing non-preferred host types (Milani et al. 2017).

The bifidobacteria genomes also reveal adaptation to their host environment through genomic signatures like accessory genes and specific gene sets, supporting our second hypothesis. Sun et al. (Sun et al. 2015) also observed that bifidobacteria isolated from bees, pigs, and humans shared unique sets of genes. However, the correlation we observed between

accessory genes and isolation sources was weaker than the association with the phylogeny based on core genes to the whole genus. Thus, it appears that specialization by bifidobacteria to a host species is primarily determined by vertically inherited traits, whereas horizontal gene transfer of traits captured through accessory gene composition plays a secondary role.

More specifically, bifidobacteria strains isolated from different hosts differed in the abundance and diversity of amino acid biosynthesis genes. Notably, bee-derived strains encoded the lowest abundance and diversity of amino acid biosynthesis genes, while non-human primates encoded the highest. Similarly, the bee strains also showed the smallest genome size. Given that species isolated from bees dominate the more ancient lineages, bifidobacteria may have coevolved longer with bees than with other hosts (Bottacini et al. 2012). One might speculate a longer coevolutionary history allowed bee-derived bifidobacteria to lose genes by evolving to use amino acids and other nutrients produced by the host or other gut bacteria, similar to the selection for smaller genome sizes observed in obligate bacterial symbionts (McCutcheon and Moran 2012; Graf and Ruby 1998).

Bifidobacteria are also known to degrade a range of carbohydrates ranging from simple to complex molecules, and there was genomic evidence of carbohydrate specialization by bifidobacteria isolated from different hosts. In particular, strains isolated from primates (including humans) carry relatively high abundances of CAZyme encoding genes. This difference could be due to more varied, plant diets of primates as well as the complexity and diversity of their milk oligosaccharides (Tao et al. 2011).

While bifidobacteria strains appear to be adapted to different hosts, there was little evidence that they are adapted to particular habitats and life stages within humans. In particular, we expected that different strains might be adapted to adults or infants, as bifidobacteria

composition varies over age (Arboleya et al. 2016; Kato et al. 2017). Indeed, some subspecies such as *B. longum subps. infantis* are specialized to breakdown human milk oligosaccharides (LoCascio et al. 2007). Perhaps we could not see the pattern at this finer scale due to the limited diversity within each bifidobacteria species in our analysis. However, a recent study also found that strains within just two species, *B. breve* and *B. longum,* isolated from the vagina and gut of humans were indistinguishable based on phylogenetic and genomic trait analyses (Freitas and Hill 2018). Thus, at least for these two habitats, that may be connected by dispersal, there are not specialized strains even when focusing on a finer genetic scale.

The lack of differences in CAZyme abundance among human categories was also surprising. This is contrary to previous studies that have found the highest abundance of CAZymes in gut bacterial communities (Berlemont and Martiny 2016; O'Callaghan and van Sinderen 2016; Cantarel, Lombard, and Henrissat 2012). In particular, we expected high numbers of CAZymes from infant strains as some bifidobacteria can degrade HMOs in the babies' gut allowing the modulation of the immune system and succession of the microbiome in the infants (Cantarel, Lombard, and Henrissat 2012; Thomson, Medina, and Garrido 2018; LoCascio et al. 2007). A point worth noting is the blood-derived strains, which we suspect are not specialized in their isolation source but instead are transient. Indeed, the strain classified as *B. scardovii* JCM 12489^T = DSM 13734^T (accession number AP012331) has been reported to have one of the largest genomes consisting of 3,158,347 bp with no plasmids and with the largest number of glycosyl hydrolase genes (Toh et al. 2015).

Our conclusions are limited by data issues inherent to the reanalysis of publicly available genomes that could be addressed in future research. First, the sampling among host animals is quite uneven, and larger sample sizes among a broader range of hosts would strengthen the

results. Second, signals of host or habitat adaptation will be stronger at a higher genetic resolution (i.e. within bifidobacteria species), and thus there is a need for deeper sampling of strains to resolve finer-scale adaptation. Related to this, we had to exclude many human-derived genomes that were not accompanied by information about the specific isolation site and age stage of the host. Lastly, it is unclear whether some of the observed patterns might have been influenced by different isolation methods, which likely varied across different studies.

CONCLUSION

This comparative genomic analysis reveals that bifidobacteria are adapted to their hosts. This adaptation is reflected in the evolutionary history of the shared core genome as well as their accessory gene composition and specific gene sets. At the same time, there is little evidence within the genus for specialization on particular human habitats or stages, which may be due to sampling limitations or a higher degree of bacterial dispersal within humans than appreciated. In sum, the assembly of bifidobacteria in their habitats appears to be determined by a mix of ecological (host filtering) and evolutionary (host adaptation) forces (Moran, Ochman, and Hammer 2019). Bifidobacteria thus offers a model to study these processes in animal microbiomes.

METHODS

Genome sequences and annotation

Genome sequences of all *Bifidobacterium* strains were downloaded from the Pathosystems Resource Integration Center (PATRIC) and the National Center for Biotechnology Information (NCBI) databases on March 14th, 2018 (n=497). Duplicate sequences were removed from further analysis. We identified the hosts for each of the strains by searching the PATRIC and NCBI databases or associated publications (n=449). Based on the concatenation of 107 core

genes (see phylogenetic analysis below for details), we removed sequences with many gaps in the core genes from further analysis and only kept unique strains (n=400). The vast majority of the strains in the databases were derived from human hosts followed by primates, cattle, pigs and bees. For strains isolated from humans (n=272), we assigned each strain to the most specific category possible, acknowledging that some categories are subsets of other categories: infant (n=117), adult (n=20), human blood (n=13), human milk (n=10), urogenital (n=9), elderly (n=5), child (n=4), probiotic (n=3), oral (n=2), human unspecified (n=89). Child refers to 2-6 years old while infant usually refers to children anywhere from birth to 1 year old (or reported as infant in their respected studies). A subset of 60 human-derived strains from diverse environments were used for genomic comparisons based on their descriptive isolation source.

To compare strains among hosts, we focused on a subset of 129 bifidobacteria strains. These strains included the majority of the non-human bifidobacteria strains in addition to a subset of human strains from adult and infant feces (n=13), blood (n=1), vagina (n=1), and mouth (n=1). The categories were the following: primate (n=18), human (n=16), cattle (n=15), pig (n=16), bee (n=16), rodent (n=12), probiotic (n=8), wastewater (n=7), rabbit (n=7), chicken (n=6), other mammals (n=4; including giraffe, hippopotamus, llama, and wallaby), dairy products (n=3), soil-plant-associated (n=1). We recognize that not all the host categories are at the same phylogenetic level.

To ensure uniform annotation, we reannotated all the genomes using Prodigal v2.6.3 in Normal Mode to predict Open Reading Frames (ORF) (Hyatt et al. 2010). We then used Prokka v1.13 (Seemann 2014) to annotate the sequences.

Phylogenetic analysis

Multilocus phylogenetic trees were constructed using the bcgTree pipeline (Ankenbrand and Keller 2016) with the protein fasta files (.*faa) derived from Prodigal v2.6.3. Each of the genome sequences was searched for 107 conserved single-copy genes defined by Dupont et al. 2012 (Dupont et al. 2012) using hmmsearch v3.1b2. The extracted genes were then each aligned using muscle v3.8.31 (Edgar 2004) and polished using Gblocks v0.91b (Castresana 2000) by eliminating poorly aligned areas. The 107 genes were then concatenated, and a phylogenetic tree was built using RAxML v8.2.10 with PROTGAMMABLOSUM62 substitution model and 100 rapid Bootstrap searches (Stamatakis 2014). We visualized the phylogenetic trees using the iTOL v3 interactive tool (Letunic and Bork 2016).

Comparative genomic analysis

We next tested whether some of the variation in the traits encoded by bifidobacteria genomes could be explained by the host or environment from which they were isolated. We used the genome size values provided by the PATRIC metadata to compare the genome size among isolates. For human-derived strains we used the same 60 sequences used in the phylogenetic analysis since they were carefully chosen to encompass variable human environments and tried to keep similar samples sizes when possible between categories; however, for the comparison among multiple hosts and environments we used a subset of the 129 strains to keep sample sizes the same for each category (n=6); hence, we did not include isolates from the dairy, mammal, and soil categories since their sample sizes were less than 6 strains.

The pan-genome and gene ontology of the 129 selected bifidobacteria strains were established with Roary v3.12.0 (Page et al. 2015) using the annotated genome assemblies obtained from Prokka v1.13 (.gff files). To account for the relatively high diversity of this genus, we used a 50% sequence identity for the blastp cutoff (Chase et al. 2018). The Roary software

was able to detect core genes (present in 99%-100% of the strains), soft core genes (present in 95%-99% of the strains), shell genes (present in 15%-95% of the strains), and cloud genes (present in 0%-15%). The presence-absence table given by Roary, depicting the 26,905 gene clusters, was curated by deleting the following genes: core genes present in all 129 strains (minus 352 = total: 26,553), singletons (minus 10,967 = total: 15,586), genes with an average sequence per isolate higher than 1, due to splitting errors (minus 189 = total: 15,397), and genes with hypothetical annotation with no identifiable gene name (minus 9,000 = total: 6,397). The final table containing 6,397 accessory genes was converted into a matrix for further comparisons between core genes and phylogenetic distance against accessory gene composition. We used Phandango (Hadfield et al. 2018) to construct the pan-genome alignment by incorporating the RAxML inferred tree and the presence-absence table given by Roary.

To assess the abundance (number of genes) and diversity (number of different genes) of amino acid biosynthesis genes, the automatic annotation server Ghostkoala was used to obtain gene function assignments based on the KEGG Orthology (Kanehisa, Sato, and Morishima 2016). To identify the CAZymes encoded in each genome, we used the dbCAN2 meta server based on the CAZy database updated on July 13th, 2018 (Yin et al. 2012; Zhang et al. 2018). The input files for the webserver were protein fasta files (.*faa) derived from Prodigal v2.6.3. This server has the option to utilize three tools to predict CAZymes: i) HMMER search against the dbCAN HMM (hidden MArkov model) database; ii) DIAMOND search against pre-annotated CAZyme sequence database; iii) Hotpep search against the CAZyme short peptide database. We used all three tools at the default parsing thresholds and only considered the CAZymes found by all three tools.

Statistical Analyses

We used ANOSIM in PRIMER-6 Software (Clarke and Gorley 2006) to test whether the isolation source categories were associated with phylogenetic relatedness and accessory genes of the bifidobacteria strains. To test for a correlation between the similarity in accessory and core gene content, we used the Relate test in PRIMER-6. We used the Tree and reticulogram REConstruction (T-REX) web server (Boc, Diallo, and Makarenkov 2012) to create the distance matrices used in the ANOSIM and Relate tests using the Netwick phylogenetic tree from RAxML. We assessed normality of data using Shapiro-Wilk normality test and its variance with Levene's test incorporated in RStudio version 1.1.453. To account for the non-normal data and non-equal sample sizes, we used the Kruskal-Wallis (with a calculated significance level of $p > 0.05$) and Dunn's post hoc tests (RStudio version 1.1.453) to compare genome size, amino acid biosynthesis genes, and CAZymes between the different strains belonging to varying hosts and environments. To construct heatmaps and boxplots, RStudio version 1.1.453 (http://www.rstudio.com/) was implemented and to help with the optimization of the images created, Adobe® Acrobat® Pro 2017 was used.

ACKNOWLEDGEMENTS

We would like to thank and acknowledge Katrine Whiteson, Brandon Gaut, and the members of the Martiny lab for their feedback while working on the manuscript.

REFERENCES

Ankenbrand, Markus J., and Alexander Keller. 2016. "BcgTree: Automatized Phylogenetic Tree Building from Bacterial Core Genomes." Edited by Frédéric Chain. *Genome* 59 (10): 783–91. https://doi.org/10.1139/gen-2015-0175.

Arboleya, Silvia, Claire Watkins, Catherine Stanton, and R. Paul Ross. 2016. "Gut Bifidobacteria Populations in Human Health and Aging." *Frontiers in Microbiology* 7 (August). https://doi.org/10.3389/fmicb.2016.01204.

Berlemont, Renaud, and Adam C. Martiny. 2016. "Glycoside Hydrolases across Environmental Microbial Communities." *PLoS Computational Biology* 12 (12). https://doi.org/10.1371/journal.pcbi.1005300.

Boc, Alix, Alpha Boubacar Diallo, and Vladimir Makarenkov. 2012. "T-REX: A Web Server for Inferring, Validating and Visualizing Phylogenetic Trees and Networks." *Nucleic Acids Research* 40 (W1): W573–79. https://doi.org/10.1093/nar/gks485.

Bode, Lars. 2012. "Human Milk Oligosaccharides: Every Baby Needs a Sugar Mama." *Glycobiology* 22 (9): 1147–62. https://doi.org/10.1093/glycob/cws074.

Bottacini, Francesca, Christian Milani, Francesca Turroni, Borja Sánchez, Elena Foroni, Sabrina Duranti, Fausta Serafini, et al. 2012. "Bifidobacterium Asteroides PRL2011 Genome Analysis Reveals Clues for Colonization of the Insect Gut." *PLOS ONE* 7 (9): e44229. https://doi.org/10.1371/journal.pone.0044229.

Cantarel, Brandi L., Vincent Lombard, and Bernard Henrissat. 2012. "Complex Carbohydrate Utilization by the Healthy Human Microbiome." *PLoS ONE* 7 (6). https://doi.org/10.1371/journal.pone.0028742.

Castresana, J. 2000. "Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis." *Molecular Biology and Evolution* 17 (4): 540–52. https://doi.org/10.1093/oxfordjournals.molbev.a026334.

Chase, Alexander B., Zulema Gomez-Lunar, Alberto E. Lopez, Junhui Li, Steven D. Allison, Adam C. Martiny, and Jennifer B. H. Martiny. 2018. "Emergence of Soil Bacterial Ecotypes along a Climate Gradient." *Environmental Microbiology* 20 (11): 4112–26. https://doi.org/10.1111/1462-2920.14405.

Clarke, K. R., and R. N. Gorley. 2006. *Primer V6: User Manual - Tutorial*. Plymouth Marine Laboratory.

Cobo-Simón, Marta, and Javier Tamames. 2017. "Relating Genomic Characteristics to Environmental Preferences and Ubiquity in Different Microbial Taxa." *BMC Genomics* 18 (June). https://doi.org/10.1186/s12864-017-3888-y.

David, Lawrence A., Arne C. Materna, Jonathan Friedman, Maria I. Campos-Baptista, Matthew C. Blackburn, Allison Perrotta, Susan E. Erdman, and Eric J. Alm. 2014. "Host Lifestyle Affects Human Microbiota on Daily Timescales." *Genome Biology* 15 (7): R89. https://doi.org/10.1186/gb-2014-15-7-r89.

David, Lawrence A., Corinne F. Maurice, Rachel N. Carmody, David B. Gootenberg, Julie E. Button, Benjamin E. Wolfe, Alisha V. Ling, et al. 2014. "Diet Rapidly and Reproducibly Alters the Human Gut Microbiome." *Nature* 505 (7484): 559–63. https://doi.org/10.1038/nature12820.

Dini-Andreote, Francisco, Fernando Dini Andreote, Welington Luiz Araújo, Jack T. Trevors, and Jan Dirk van Elsas. 2012. "Bacterial Genomes: Habitat Specificity and Uncharted Organisms." *Microbial Ecology* 64 (1): 1–7. https://doi.org/10.1007/s00248-012-0017-y.

Dupont, Chris L, Douglas B Rusch, Shibu Yooseph, Mary-Jane Lombardo, R Alexander Richter, Ruben Valas, Mark Novotny, et al. 2012. "Genomic Insights to SAR86, an Abundant and Uncultivated Marine Bacterial Lineage." *The ISME Journal* 6 (6): 1186–99. https://doi.org/10.1038/ismej.2011.189.

Edgar, Robert C. 2004. "MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput." *Nucleic Acids Research* 32 (5): 1792–97. https://doi.org/10.1093/nar/gkh340.

Esaiassen, Eirin, Erik Hjerde, Jorunn Pauline Cavanagh, Gunnar Skov Simonsen, and Claus Klingenberg. 2017. "Bifidobacterium Bacteremia: Clinical Characteristics and a Genomic Approach To Assess Pathogenicity." *Journal of Clinical Microbiology* 55 (7): 2234–48. https://doi.org/10.1128/JCM.00150-17.

Freitas, Aline C., and Janet E. Hill. 2018. "Bifidobacteria Isolated from Vaginal and Gut Microbiomes Are Indistinguishable by Comparative Genomics." *PLoS ONE* 13 (4). https://doi.org/10.1371/journal.pone.0196290.

Graf, Joerg, and Edward G. Ruby. 1998. "Host-Derived Amino Acids Support the Proliferation of Symbiotic Bacteria." *Proceedings of the National Academy of Sciences* 95 (4): 1818–22. https://doi.org/10.1073/pnas.95.4.1818.

Groussin, Mathieu, Florent Mazel, Jon G. Sanders, Chris S. Smillie, Sébastien Lavergne, Wilfried Thuiller, and Eric J. Alm. 2017. "Unraveling the Processes Shaping Mammalian Gut Microbiomes over Evolutionary Time." *Nature Communications* 8 (February): 14319. https://doi.org/10.1038/ncomms14319.

Hadfield, James, Nicholas J. Croucher, Richard J. Goater, Khalil Abudahab, David M. Aanensen, and Simon R. Harris. 2018. "Phandango: An Interactive Viewer for Bacterial Population Genomics." *Bioinformatics* 34 (2): 292–93. https://doi.org/10.1093/bioinformatics/btx610.

Hall, James P. J., Michael A. Brockhurst, and Ellie Harrison. 2017. "Sampling the Mobile Gene Pool: Innovation via Horizontal Gene Transfer in Bacteria." *Philosophical Transactions of the Royal Society B: Biological Sciences* 372 (1735). https://doi.org/10.1098/rstb.2016.0424.

Hyatt, Doug, Gwo-Liang Chen, Philip F LoCascio, Miriam L Land, Frank W Larimer, and Loren J Hauser. 2010. "Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification." *BMC Bioinformatics* 11 (March): 119. https://doi.org/10.1186/1471-2105-11-119.

Kamilar, Jason M., and Natalie Cooper. 2013. "Phylogenetic Signal in Primate Behaviour, Ecology and Life History." *Philosophical Transactions of the Royal Society B: Biological Sciences* 368 (1618). https://doi.org/10.1098/rstb.2012.0341.

Kanehisa, Minoru, Yoko Sato, and Kanae Morishima. 2016. "BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences." *Journal of Molecular Biology* 428 (4): 726–31. https://doi.org/10.1016/j.jmb.2015.11.006.

Kato, Kumiko, Toshitaka Odamaki, Eri Mitsuyama, Hirosuke Sugahara, Jin-zhong Xiao, and Ro Osawa. 2017. "Age-Related Changes in the Composition of Gut Bifidobacterium Species." *Current Microbiology* 74 (8): 987–95. https://doi.org/10.1007/s00284-017-1272-4.

Klijn, Adrianne, Annick Mercenier, and Fabrizio Arigoni. 2005. "Lessons from the Genomes of Bifidobacteria." *FEMS Microbiology Reviews* 29 (3): 491–509. https://doi.org/10.1016/j.fmrre.2005.04.010.

Lamendella, Regina, Jorge W. Santo Domingo, Catherine Kelty, and Daniel B. Oerther. 2008. "Bifidobacteria in Feces and Environmental Waters." *Appl. Environ. Microbiol.* 74 (3): 575–84. https://doi.org/10.1128/AEM.01221-07.

Letunic, Ivica, and Peer Bork. 2016. "Interactive Tree of Life (ITOL) v3: An Online Tool for the Display and Annotation of Phylogenetic and Other Trees." *Nucleic Acids Research* 44 (W1): W242-245. https://doi.org/10.1093/nar/gkw290.

Liu, Songling, Fazheng Ren, Liang Zhao, Lu Jiang, Yanling Hao, Junhua Jin, Ming Zhang, et al. 2015. "Starch and Starch Hydrolysates Are Favorable Carbon Sources for Bifidobacteria in the Human Gut." *BMC Microbiology* 15 (1): 54. https://doi.org/10.1186/s12866-015-0362-3.

LoCascio, Riccardo G., Milady R. Ninonuevo, Samara L. Freeman, David A. Sela, Rudolf Grimm, Carlito B. Lebrilla, David A. Mills, and J. Bruce German. 2007. "Glycoprofiling of Bifidobacterial Consumption of Human Milk Oligosaccharides Demonstrates Strain Specific, Preferential Consumption of Small Chain Glycans Secreted in Early Human Lactation." *Journal of Agricultural and Food Chemistry* 55 (22): 8914–19. https://doi.org/10.1021/jf0710480.

Lugli, Gabriele Andrea, Sabrina Duranti, Korin Albert, Leonardo Mancabelli, Stefania Napoli, Alice Viappiani, Rosaria Anzalone, et al. 2019. "Unveiling Genomic Diversity among Members of the Species Bifidobacterium Pseudolongum, a Widely Distributed Gut Commensal of the Animal Kingdom." *Applied and Environmental Microbiology* 85 (8): e03065-18. https://doi.org/10.1128/AEM.03065-18.

Lugli, Gabriele Andrea, Christian Milani, Francesca Turroni, Sabrina Duranti, Chiara Ferrario, Alice Viappiani, Leonardo Mancabelli, et al. 2014. "Investigation of the Evolutionary

Development of the Genus Bifidobacterium by Comparative Genomics." *Applied and Environmental Microbiology* 80 (20): 6383–94. https://doi.org/10.1128/AEM.02004-14.

Marcobal, A., and J. L. Sonnenburg. 2012. "Human Milk Oligosaccharide Consumption by Intestinal Microbiota." *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 18 (0 4): 12–15. https://doi.org/10.1111/j.1469-0691.2012.03863.x.

Martín, Rocío, Esther Jiménez, Hans Heilig, Leonides Fernández, María L. Marín, Erwin G. Zoetendal, and Juan M. Rodríguez. 2009. "Isolation of Bifidobacteria from Breast Milk and Assessment of the Bifidobacterial Population by PCR-Denaturing Gradient Gel Electrophoresis and Quantitative Real-Time PCR." *Applied and Environmental Microbiology* 75 (4): 965–69. https://doi.org/10.1128/AEM.02063-08.

Martiny, Adam C, Kathleen Treseder, and Gordon Pusch. 2013. "Phylogenetic Conservatism of Functional Traits in Microorganisms." *The ISME Journal* 7 (4): 830–38. https://doi.org/10.1038/ismej.2012.160.

McCutcheon, John P., and Nancy A. Moran. 2012. "Extreme Genome Reduction in Symbiotic Bacteria." *Nature Reviews Microbiology* 10 (1): 13–26. https://doi.org/10.1038/nrmicro2670.

McFall-Ngai, Margaret, Michael G. Hadfield, Thomas C. G. Bosch, Hannah V. Carey, Tomislav Domazet-Lošo, Angela E. Douglas, Nicole Dubilier, et al. 2013. "Animals in a Bacterial World, a New Imperative for the Life Sciences." *Proceedings of the National Academy of Sciences of the United States of America* 110 (9): 3229–36. https://doi.org/10.1073/pnas.1218525110.

Milani, Christian, Gabriele Andrea Lugli, Sabrina Duranti, Francesca Turroni, Francesca Bottacini, Marta Mangifesta, Borja Sanchez, et al. 2014. "Genomic Encyclopedia of Type Strains of the Genus Bifidobacterium." *Applied and Environmental Microbiology* 80 (20): 6290–6302. https://doi.org/10.1128/AEM.02308-14.

Milani, Christian, Marta Mangifesta, Leonardo Mancabelli, Gabriele A Lugli, Kieran James, Sabrina Duranti, Francesca Turroni, et al. 2017. "Unveiling Bifidobacterial Biogeography across the Mammalian Branch of the Tree of Life." *The ISME Journal* 11 (12): 2834–47. https://doi.org/10.1038/ismej.2017.138.

Milani, Christian, Francesca Turroni, Sabrina Duranti, Gabriele Andrea Lugli, Leonardo Mancabelli, Chiara Ferrario, Douwe van Sinderen, and Marco Ventura. 2016. "Genomics of the Genus Bifidobacterium Reveals Species-Specific Adaptation to the Glycan-Rich Gut Environment." *Applied and Environmental Microbiology* 82 (4): 980–91. https://doi.org/10.1128/AEM.03500-15.

Moeller, Andrew H., Alejandro Caro-Quintero, Deus Mjungu, Alexander V. Georgiev, Elizabeth V. Lonsdorf, Martin N. Muller, Anne E. Pusey, Martine Peeters, Beatrice H. Hahn, and

Howard Ochman. 2016. “Cospeciation of Gut Microbiota with Hominids.” *Science* 353 (6297): 380–82. https://doi.org/10.1126/science.aaf3951.

Moran, Nancy A., Howard Ochman, and Tobin J. Hammer. 2019. “Evolutionary and Ecological Consequences of Gut Microbial Communities.” *Annual Review of Ecology, Evolution, and Systematics* 50 (1): null. https://doi.org/10.1146/annurev-ecolsys-110617-062453.

Muegge, Brian D., Justin Kuczynski, Dan Knights, Jose C. Clemente, Antonio González, Luigi Fontana, Bernard Henrissat, Rob Knight, and Jeffrey I. Gordon. 2011. “Diet Drives Convergence in Gut Microbiome Functions Across Mammalian Phylogeny and Within Humans.” *Science* 332 (6032): 970–74. https://doi.org/10.1126/science.1198719.

Münkemüller, Tamara, Sébastien Lavergne, Bruno Bzeznik, Stéphane Dray, Thibaut Jombart, Katja Schiffers, and Wilfried Thuiller. 2012. “How to Measure and Test Phylogenetic Signal.” *Methods in Ecology and Evolution* 3 (4): 743–56. https://doi.org/10.1111/j.2041-210X.2012.00196.x.

Neis, Evelien P. J. G., Cornelis H. C. Dejong, and Sander S. Rensen. 2015. “The Role of Microbial Amino Acid Metabolism in Host Metabolism.” *Nutrients* 7 (4): 2930–46. https://doi.org/10.3390/nu7042930.

O’Callaghan, Amy, and Douwe van Sinderen. 2016. “Bifidobacteria and Their Role as Members of the Human Gut Microbiota.” *Frontiers in Microbiology* 7 (June). https://doi.org/10.3389/fmicb.2016.00925.

Page, Andrew J., Carla A. Cummins, Martin Hunt, Vanessa K. Wong, Sandra Reuter, Matthew T. G. Holden, Maria Fookes, Daniel Falush, Jacqueline A. Keane, and Julian Parkhill. 2015. “Roary: Rapid Large-Scale Prokaryote Pan Genome Analysis.” *Bioinformatics* 31 (22): 3691–93. https://doi.org/10.1093/bioinformatics/btv421.

Rivière, Audrey, Frédéric Moens, Marija Selak, Dominique Maes, Stefan Weckx, and Luc De Vuyst. 2014. “The Ability of Bifidobacteria To Degrade Arabinoxylan Oligosaccharide Constituents and Derived Oligosaccharides Is Strain Dependent.” *Applied and Environmental Microbiology* 80 (1): 204–17. https://doi.org/10.1128/AEM.02853-13.

Ruiz-Moyano, Santiago, Sarah M. Totten, Daniel A. Garrido, Jennifer T. Smilowitz, J. Bruce German, Carlito B. Lebrilla, and David A. Mills. 2013. “Variation in Consumption of Human Milk Oligosaccharides by Infant Gut-Associated Strains of Bifidobacterium Breve.” *Appl. Environ. Microbiol.* 79 (19): 6040–49. https://doi.org/10.1128/AEM.01843-13.

Seemann, Torsten. 2014. “Prokka: Rapid Prokaryotic Genome Annotation.” *Bioinformatics (Oxford, England)* 30 (14): 2068–69. https://doi.org/10.1093/bioinformatics/btu153.

Sharma, Vikas, Fauzul Mobeen, and Tulika Prakash. 2018. “Exploration of Survival Traits, Probiotic Determinants, Host Interactions, and Functional Evolution of Bifidobacterial

Genomes Using Comparative Genomics." *Genes* 9 (10): 477. https://doi.org/10.3390/genes9100477.

Slavin, Joanne. 2013. "Fiber and Prebiotics: Mechanisms and Health Benefits." *Nutrients* 5 (4): 1417–35. https://doi.org/10.3390/nu5041417.

Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30 (9): 1312–13. https://doi.org/10.1093/bioinformatics/btu033.

Sun, Zhihong, Wenyi Zhang, Chenyi Guo, Xianwei Yang, Wenjun Liu, Yarong Wu, Yuqin Song, et al. 2015. "Comparative Genomic Analysis of 45 Type Strains of the Genus Bifidobacterium: A Snapshot of Its Genetic Diversity and Evolution." Edited by Christian U. Riedel. *PLOS ONE* 10 (2): e0117912. https://doi.org/10.1371/journal.pone.0117912.

Tao, Nannan, Shuai Wu, Jaehan Kim, Hyun Joo An, Katie Hinde, Michael L. Power, Pascal Gagneux, J. Bruce German, and Carlito B. Lebrilla. 2011. "Evolutionary Glycomics: Characterization of Milk Oligosaccharides in Primates." *Journal of Proteome Research* 10 (4): 1548–57. https://doi.org/10.1021/pr1009367.

Thomson, Pamela, Daniel A. Medina, and Daniel Garrido. 2018. "Human Milk Oligosaccharides and Infant Gut Bifidobacteria: Molecular Strategies for Their Utilization." *Food Microbiology*, New Tools to Detect and Prevent Foodborne Outbreaks from Farm to Fork, 75 (October): 37–46. https://doi.org/10.1016/j.fm.2017.09.001.

Toh, Hidehiro, Kenshiro Oshima, Akiyo Nakano, Naoko Yamashita, Erica Iioka, Rina Kurokawa, Hidetoshi Morita, and Masahira Hattori. 2015. "Complete Genome Sequence of Bifidobacterium Scardovii Strain JCM 12489T, Isolated from Human Blood." *Genome Announcements* 3 (2). https://doi.org/10.1128/genomeA.00285-15.

Turroni, Francesca, David Berry, and Marco Ventura. 2017. *Bifidobacteria and Their Role in the Human Gut Microbiota*. Frontiers Media SA.

Turroni, Francesca, Christian Milani, Sabrina Duranti, Chiara Ferrario, Gabriele Andrea Lugli, Leonardo Mancabelli, Douwe van Sinderen, and Marco Ventura. 2018. "Bifidobacteria and the Infant Gut: An Example of Co-Evolution and Natural Selection." *Cellular and Molecular Life Sciences* 75 (1): 103–18. https://doi.org/10.1007/s00018-017-2672-0.

Turroni, Francesca, Douwe van Sinderen, and Marco Ventura. 2011. "Genomics and Ecological Overview of the Genus Bifidobacterium." *International Journal of Food Microbiology*, 3rd International Symposium on Propionibacteria and Bifidobacteria: Dairy and Probiotic applications, Oviedo 1-4 June 2010, 149 (1): 37–44. https://doi.org/10.1016/j.ijfoodmicro.2010.12.010.

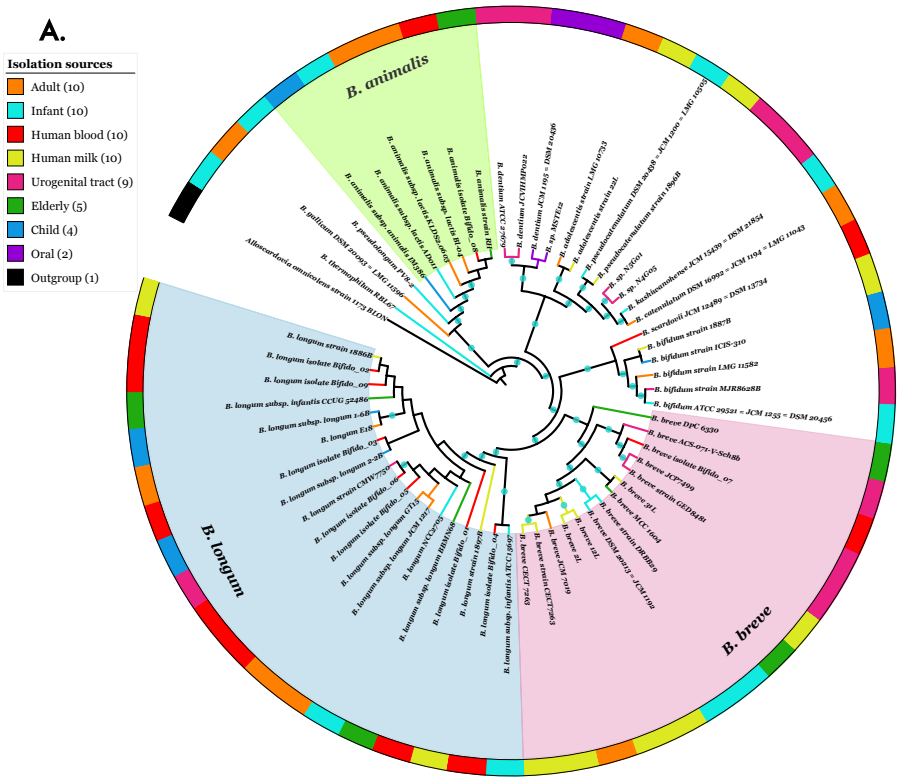
Ventura, Marco, Carlos Canchaya, Antonio Del Casale, Franco Dellaglio, Erasmo Neviani, Gerald F. Fitzgerald, and Douwe van Sinderen. 2006. "Analysis of Bifidobacterial

Evolution Using a Multilocus Approach." *International Journal of Systematic and Evolutionary Microbiology* 56 (12): 2783–92. https://doi.org/10.1099/ijs.0.64233-0.

Yin, Yanbin, Xizeng Mao, Jincai Yang, Xin Chen, Fenglou Mao, and Ying Xu. 2012. "DbCAN: A Web Resource for Automated Carbohydrate-Active Enzyme Annotation." *Nucleic Acids Research* 40 (W1): W445–51. https://doi.org/10.1093/nar/gks479.

Zhang, Han, Tanner Yohe, Le Huang, Sarah Entwistle, Peizhi Wu, Zhenglu Yang, Peter K. Busk, Ying Xu, and Yanbin Yin. 2018. "DbCAN2: A Meta Server for Automated Carbohydrate-Active Enzyme Annotation." *Nucleic Acids Research* 46 (W1): W95–101. https://doi.org/10.1093/nar/gky418.

FIGURES



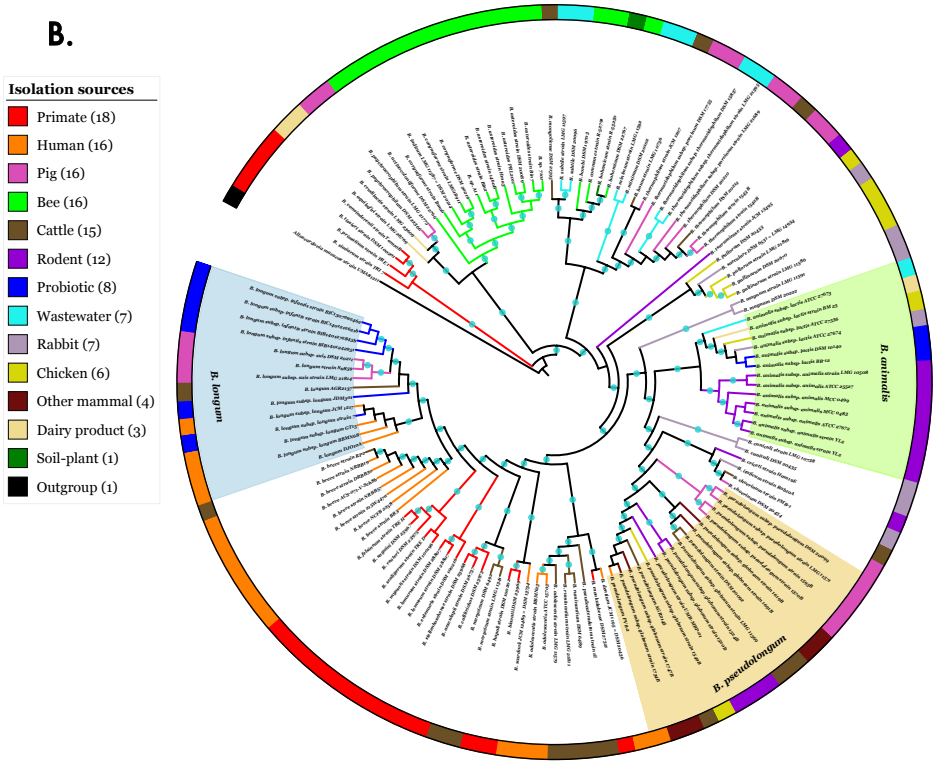


Figure 1.1: Cladograms of bifidobacteria strains harbored by A) human hosts and B) multiple hosts and environments. 1.1A|Phylogenetic relationship of human-derived *Bifidobacterium* strains based on 107 marker genes (n=60 + outgroup). The prominent species, *B. longum* (31.7%), *B. breve* (23.3%), and *B. animalis* (10%), are shaded in different colors. **1B**|Phylogenetic relationship of *Bifidobacterium* strains harbored by multiple hosts based on 107 marker genes (n=129 + outgroup). The prominent species, *B. pseudolongum* (12.4%), *B. longum* (10.9%) and *B. animalis* (10.1%), are shaded in different colors. For both cladograms, the outermost ring represents the different isolation sources. Bootstrap values higher than 70% are represented with blue circles. Strains from the *Alloscardovia* genus were used as outgroups for both phylogenetic trees (accession numbers JWAI01000000 and NEKB01000000). Note that the "child" category refers to ages 2 through 6 years old while "infant" is younger. The "mammal" category indicates a mammal with only 1 sample size, including giraffe (n=1), hippopotamus (n=1), llama (n=1), and wallaby (n=1). Also, the "primate" category indicates non-human primates, and "probiotic" had an original, unknown isolation source that may overlap with the other categories.

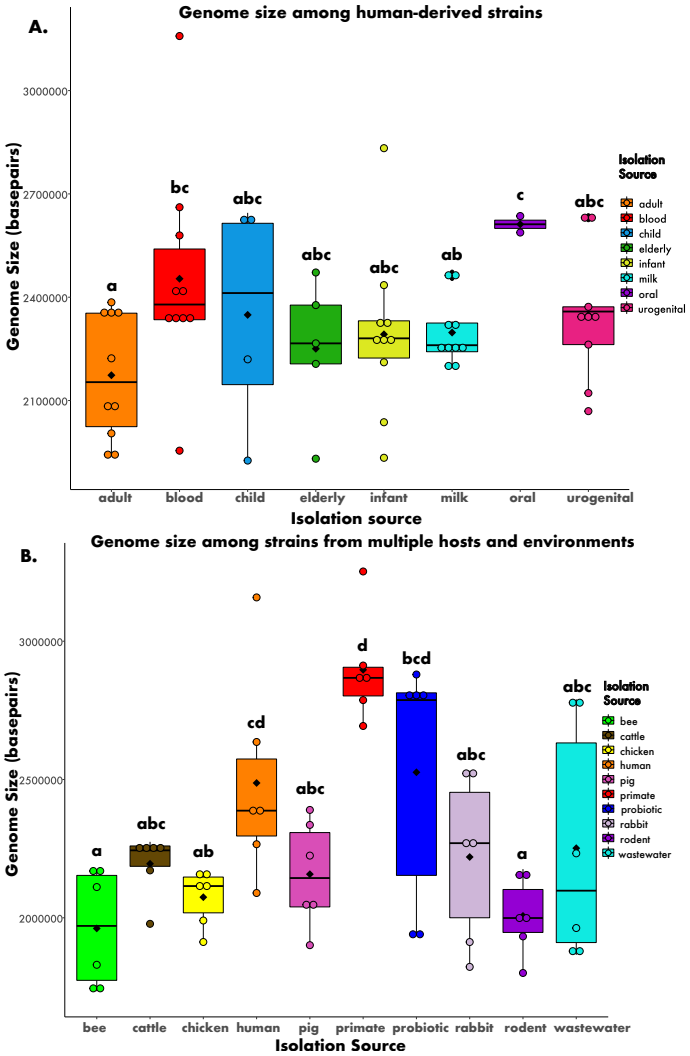


Figure 1.2: Bifidobacteria genome sizes for strains derived from A) humans and B) multiple hosts and environments. The circles depict the data points, and the black diamonds represent the mean of each boxplot. The letters above each box represent the *post hoc* comparisons using Dunn's test where groups sharing a letter are not significantly different. See Methods and Figure 1.1 legend for more information about the isolation categories.

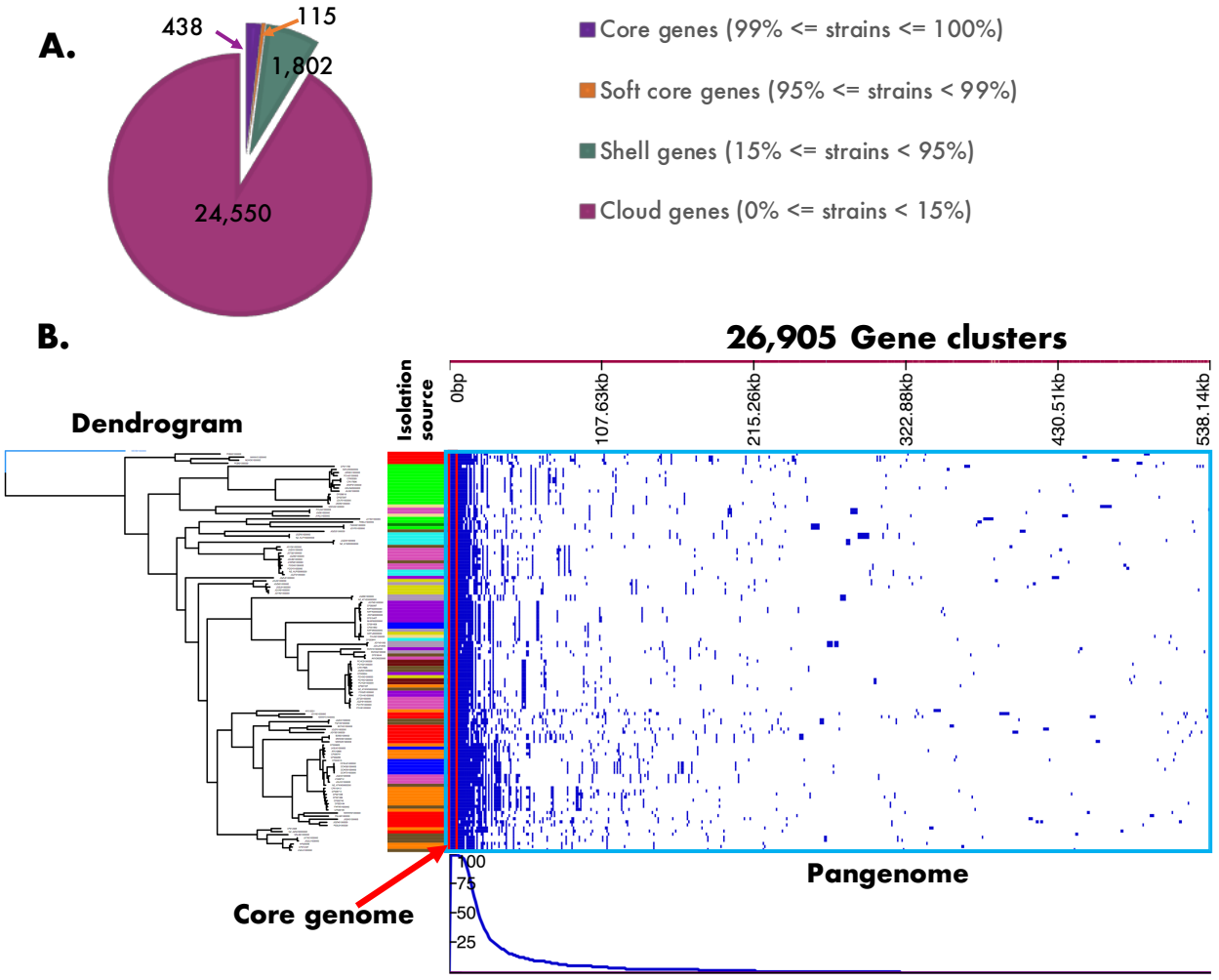


Figure 1.3: Pan-genome analysis of 129 *Bifidobacterium* strains harbored by multiple hosts. **2A**|The pan-genome of the 129 bifidobacteria strains is summarized in a pie chart showing the core genes (438), the soft genes (115), the shell genes (1,802), and the cloud genes (24,550). **2B**|Pan-genome alignment of 129 bifidobacteria strains is depicted by combining the phylogenetic tree inferred by RAxML 8.2.10 and the pan-genome heatmap showing gene presence (royal blue) or absence (white) in each of the strains obtained with Roary 3.11.2. There was a total of 26,905 gene clusters (of orthologous proteins) from which 438 were present in all strains. The line graph at the bottom shows the frequency of genes present within samples. The core-genome and pan-genome are boxed in red and light blue, respectively. The color strip next to the alignment depicts the isolation sources described in Fig. 1.1.

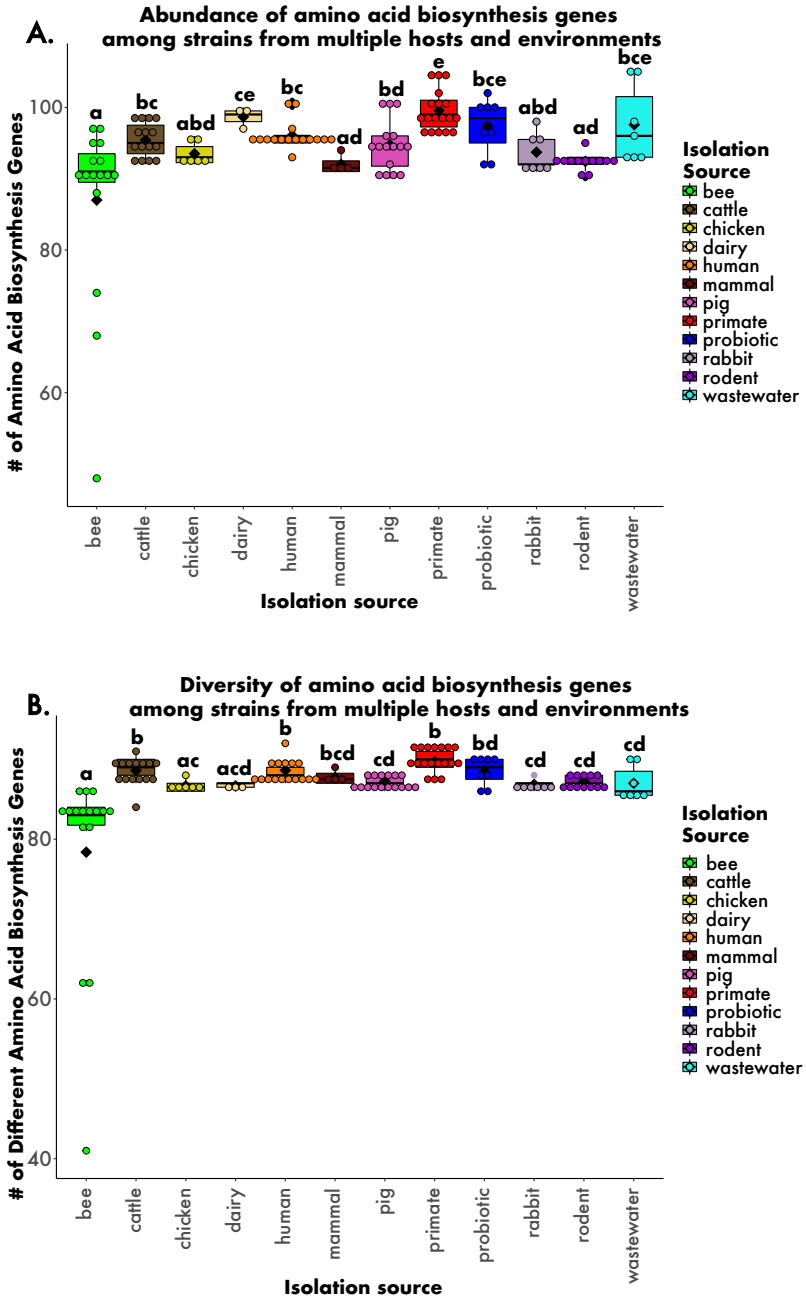


Figure 1.4: Amino acid biosynthesis gene A) abundances and B) diversity among different hosts and environments. Abundance refers to the number of total gene count and diversity refers to the number of different genes found. The circles depict the data points and the black diamonds represent the mean of each boxplot. The letters above each box represent the *post hoc* comparisons using Dunn's test where groups sharing a letter are not significantly different. See Methods and Figure 1.1 legend for more information about the isolation categories.

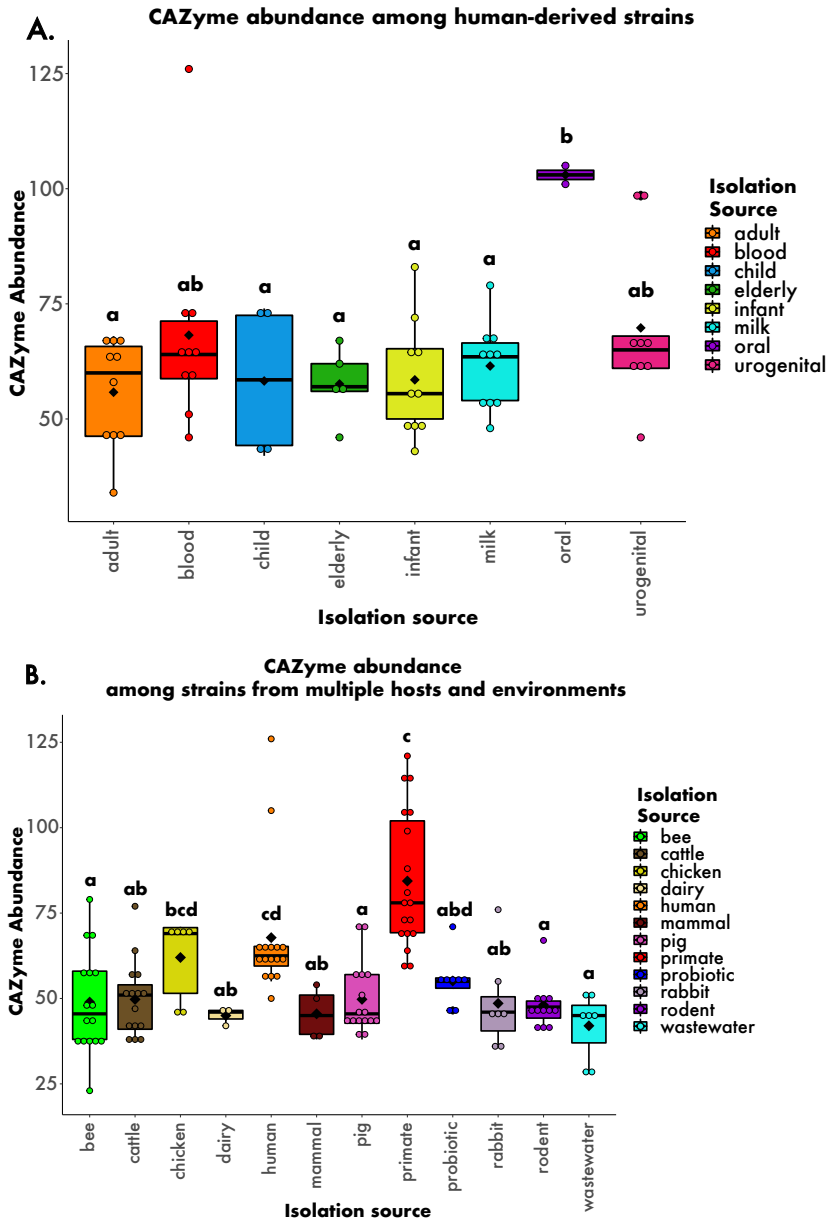


Figure 1.5: CAZyme abundances for strains isolated from A) humans and B) multiple hosts and environments. The circles depict the data points and the black diamonds represent the mean of each boxplot. The letters above each box represent the *post hoc* comparisons using Dunn's test. See Methods and Figure 1.1 legend for more information about the isolation categories.

CHAPTER 2

Short-term dietary fiber interventions produce consistent gut microbiome responses across studies

ABSTRACT

The composition of the human gut microbiome varies tremendously among individuals (high inter-individual variation), making intervention effects such as changes in diet difficult to detect and characterize. For instance, the consumption of fiber is known to be important for gut health yet the specific effects of increased fiber intake on the gut microbiome vary. Moreover, the details of dietary fiber interventions across studies differ greatly in the types of fiber, length of study, size of cohort, and molecular approaches, which might also contribute to different study outcomes. To better understand fiber-induced changes in the gut microbiome of healthy individuals, we re-analyzed 16S rRNA sequencing data from 21 dietary fiber interventions from 12 human studies. We found that short-term increases in dietary fiber consumption resulted in highly consistent gut microbiome responses across studies. In particular, increased fiber consumption explained an average of 1.5% of compositional variation (compared to 82% of variation attributed to the individual), reduced alpha diversity, and resulted in phylogenetically conserved responses in relative abundances among bacterial taxa. Additionally, we identified bacterial clades, at approximately the genus level, that increased in relative abundance in the majority of the studies. Therefore, by synthesizing and reanalyzing microbiome data from many studies we identify consistent responses to initial changes of the gut microbiome to increased fiber intake.

INTRODUCTION

The human gut microbiome plays an important role in human health and disease (Jovel et al. 2018; Shreiner, Kao, and Young 2015). Resident gut bacteria help maintain host homeostasis not only in the gut but also throughout the human body. They do so by maintaining the integrity of the gut barrier, producing beneficial metabolites, and much more while receiving essential nutrients for survival from the human host (Durack and Lynch 2018). One question that has been popularly raised is whether it is possible to purposely engineer the composition of the gut microbiome towards a healthier state, for instance, by providing beneficial gut bacteria with the nutrients they need to flourish (prebiotics). These compounds are typically found in high-fiber foods or supplements that contain complex carbohydrates (Singh et al. 2017; Del Chierico et al. 2014).

Dietary fibers are carbohydrates that resist digestion by the small intestine and have a positive health impacts on humans (Jones 2014). High-fiber diets are associated with health benefits such as increased nutrient absorption, production of beneficial metabolites, improved immune responses, and amelioration of various diseases including obesity, diabetes, allergies, and others (Carlson et al. 2018; Yang et al. 2012; Hosseini-Esfahani et al. 2017; Makki et al. 2018). To understand the influence of dietary fiber on the gut microbiota, researchers have performed dietary interventions using different fiber compounds on humans (Sawicki et al. 2017; Holscher 2017; Simpson and Campbell 2015).

Experiments that increase fiber intake in humans often result in changes in the gut microbiome which can be consistent across studies. For example, fiber interventions including inulins, fructans, and gala-oligosaccharides often report an increase of *Bifidobacterium* and *Lactobacillus* taxa in the gut, genera known as lactic acid producers and carbohydrate degraders (S. Wang et al. 2020; So et al. 2018; Davis et al. 2011). In addition, plant-based diets (known to

be high in fiber content) have also shown to enrich the gut microbiome with *Ruminococcus* and *Prevotella* genera, which are known to degrade and ferment complex dietary carbohydrates (Abell et al. 2008; Hooda et al. 2012; Wu et al. 2011; Cronin et al. 2021; Schnorr et al. 2014; Kovatcheva-Datchary et al. 2015). Although some bacterial responses seem to be consistent across interventions, other studies have reported contradictory trends in bacterial responses to fiber (Holscher 2017; Tian et al. 2021; Whisner et al. 2018). For instance, Tian *et al*. 2021 reported no increases in the taxa mentioned above in response to fiber intake but instead showed a decrease in the *Ruminococcus* genus. Whisner and colleagues also found that *Ruminococcus* were more abundant in a group of college students that consumed low fiber foods (Whisner et al. 2018). However, such contradictions are not surprising; comparing results across any type of microbiome intervention, including increased fiber, comes with many challenges.

The first obstacle that researchers encounter when trying to compare across fiber-diet interventions is heterogeneity in study design and technical approaches. For instance, studies vary widely in the types of fiber compounds used, intervention lengths, and population sizes. Moreover, differences in molecular approaches and in downstream bioinformatic pipelines could add technical variation to the characterization of microbiome responses that potentially obscures biological patterns across studies.

Second, a challenge for any type of microbiome intervention is that individual human variation in gut microbiomes is very high. High inter-individual variation makes comparing microbiome responses across individuals difficult, let alone across studies. Not only does the starting, pre-intervention composition of the gut microbiome vary widely between individuals, but many operational taxonomic units (OTUs) are not shared among individuals within a study.

As a result, the variation in gut composition explained by an intervention will typically be small relative to inter-individual variation.

Finally, studies report their results at different taxonomic levels, making it difficult to draw conclusions about which taxa are consistently responding to the fiber treatment. Not only can OTUs of bacterial sequences be defined differently across studies (e.g., at different cutoffs such as 100%, 99%, and 97% sequence similarity), but the results are often summarized differently. For instance, some studies may report changes in relative abundance in terms of phyla (e.g., *Actinobacteria*), whereas others by family or genus (e.g., *Bifidobacteriaceae*/ *Bifidobacterium*). Moreover, it is often unclear whether all bacteria within a reported taxonomic level respond similarly to an intervention. For example, an individual's gut typically contains several strains and/or species within the same genus (Garud et al. 2019). Indeed, a study of healthy adults found that individuals can harbor up to 6 species of *Bifidobacterium* at any one time (Oliver et al. 2021). However, most studies report only the most responsive OTUs and/or changes in relative abundances lumped on a particular taxonomic level. If some taxa within a broader group respond differently to increased fiber, then this might contribute to the inconsistent results among fiber intervention studies.

Although some of the above-mentioned discrepancies are impossible to fix for past studies (e.g., study design and sequencing processes), there are avenues to make the results more comparable across interventions). One approach is to reanalyze the data in a consistent manner and use phylogenetic information to organize biological variation. To make direct comparisons of bacterial shifts across studies, the raw data (e.g., 16S rRNA sequencing reads) can be uniformly processed using similar bioinformatic pipelines, threshold parameters, and statistical analyses. In addition, phylogenetic information, rather than taxonomic bins, can allow for direct

comparison in relative abundances among taxa and variation observed after fiber consumption making it easier to measure bacterial changes across studies. For instance, there can be a mixture of responses within a taxonomic clade, and to be able to compare microbial responses across studies it is necessary to place the taxonomic responses on a phylogenetic tree. Furthermore, this approach could shed light on the depth of taxonomic responses and help clarify whether large phylogenetic clades respond in a positive or negative manner (phylogenetic signal), which can then reduce the number of bacterial groups to examine. A phylogenetic signal would be detected when related taxa respond in a similar manner to the fiber intervention.

Here, we investigated the consistency of fiber-induced changes in the gut microbiome of healthy individuals by re-analyzing 16S rRNA sequencing data from 21 dietary fiber interventions. We hypothesized that short-term increases in fiber intake would result in consistent changes in microbiome composition across different types of interventions. To test this hypothesis, we assessed three features of each intervention: 1) changes in bacterial alpha-diversity after the fiber intervention, 2) the amount of compositional variation (beta-diversity) explained by the fiber intervention relative to that of between individuals, and 3) bacterial responses in a phylogenetic context to identify consistent fiber-responding taxa.

METHODS

Study inclusion criteria

We conducted a keyword search of published literature through the PubMed search engine (keywords: dietary, fiber, and microbiome) under the Best Match algorithm recommended by PubMed on May 9th, 2020. The search yielded 977 abstract hits from 2010 to 2020 (https://pubmed.ncbi.nlm.nih.gov/). We also searched through all the records available in the database of open-source microbial management site Qiita (Gonzalez et al. 2018) on April 7th,

2020 and found 528 microbiome studies including human and animal studies (https://qiita.ucsd.edu). From both sources, each abstract was carefully read to select studies with fiber interventions in healthy humans that included 16S rRNA amplicon sequencing data from fecal microbial communities (n=34). We excluded studies in animals and unhealthy humans. Corresponding authors and first authors were contacted up to 4 times requesting their sequencing data and metadata when not publicly available. We were able to obtain 16S rRNA amplicon sequencing and their corresponding metadata from 12 studies (**Fig. 2.1**).

From the 12 collected studies, 5 conducted diet interventions with different types of dietary fibers and/or food sources. When this was the case, the samples in each study were divided by the fiber intervention, resulting in a total of 21 intervention experiments. (For example, if one study conducted separate interventions with inulin and psyllium, the dataset was divided into two.) We named each of the interventions as: Last name of the first author in the publication, followed by the year the study was published, continuing with the region of the 16S rRNA bacterial gene that was amplified, with the addition of the fiber used in the study (e.g., Baxter_2019_V4_potato; **Table 2.1**).

Sequencing processing

Individual studies used different methods for sequencing processing and bioinformatic pipelines, and such differences can influence the diversity and composition of microorganisms detected in a sample as well as the variation observed across samples. Thus, to compare the sequences directly across studies, we obtained the raw sequencing reads for each study and processed them in a similar manner.

First, we assessed the quality of the 16S sequencing data using FastQC software version 0.11.8 (Andrews 2010). The sequencing reads were cleaned from poor quality sequences using

the Fastp program version 0.20.0 (Chen et al. 2018). The cleaned sequences were imported into the QIIME2 platform version 2020.11.1 (Estaki et al. 2020), and primers were removed using Cutadapt plugin (Martin 2011) when necessary. We then denoised the reads using DADA2 plugin (Callahan et al. 2016), obtaining an OTU table with exact sequence variants (ESVs) depicting the number of reads per sample for each taxonomic unit.

Next, the taxonomic classification of the reads was also performed in the QIIME2 platform by training the SILVA version 132_99_16S (Quast et al. 2013) and the Genome Taxonomy Database (GTDB) version bac120_ssu_reps_r95 (Parks et al. 2018) databases to each respective study based on the primers that were originally used. The SILVA database was used to remove chloroplast and mitochondrial DNA. Then, the cleaned reads were assigned to a final taxonomic group using the GTDB trained database. Only reads classified to the phylum level were kept in the OTU tables. The sequencing processing and taxonomic classification was performed with both the forward and reverse reads (when available) and the forward and/or single reads only.

For uniformity, we only utilized the forward reads from all the studies and imported the data into R (version 4.0.2) for rarefaction to normalize samples and to perform alpha- and beta-diversity analyses using the vegan package version 2.6-2. We calculated rarefied OTU tables through randomized sampling sequences without replacement for 1000 iterations, using the highest sequencing depth possible for each dataset (**Table 2.2**). For each study, we only used samples from the before and after fiber intervention treatments, excluding samples from other treatments (e.g., maltodextrin-controls).

Bacterial community composition responses to individual fiber interventions

We tested differences in alpha diversity metrics (Shannon and Simpson indices) before and after fiber interventions using vegan package, version 2.6-2, and paired-t tests in R, version 4.0.2. When multiple timepoints for before and after the fiber intervention were collected, we used only two timepoints (the earliest timepoint from before and the latest sample from after the intervention) to allow for paired analyses. To test for differences in bacterial community composition between before versus after fiber intervention, we ran permutational multivariate analysis of variance (PERMANOVA) using Bray-Curtis dissimilarity matrices that were rarefied and square root transformed using all timepoints available. The PERMANOVA formula used in R through the vegan package (version 2.6-2) was: adonis2.(bray.dist.matrix~subject_id + timepoint, data= metadata, method= "bray", by= "term", permutations=999).

Phylogenetic responses to dietary fiber

Since the majority of the studies (8/12) shared the V4 region of the 16S rRNA gene, we selected this region to conduct phylogenetic analyses. For the studies that shared this region, we merged the forward and reverse reads, when available, using BBmerge from BBMap Tools, version 38.95, (Bushnell, Rood, and Singer 2017). Then, we extracted the same V4 region across the studies with Cutadapt version 3.5 using the V4 primer sequences (forward:GTGYCAGCMGCCGCGGTAA; reverse:GGACTACNVGGGTWTCTAAT) from the Earth Microbiome Project (Thompson et al. 2017). To ensure that the sequences were properly extracted (e.g., read size = 250bp) we visualized them using Geneious prime (version 2020.2.4; https://www.geneious.com/), FastQC version 0.11.9 and summarized the results with Multiqc, version 1.11. Then, the extracted reads across studies were imported into QIIME2, version 2020.11, as a single artifact. The q2-vsearch plugin in QIIME2 was used to dereplicate the sequences, and cluster them at 97% identity. Since these studies were used to make in-depth

phylogenetic comparisons, the 97% dereplication identity was used, instead of ESVs to simplify the complexity of the gut bacterial responses across studies using different collection and sequencing methods. Moreover, based on previous research (Isobe et al. 2020) a finer-scale assignment of OTUs (ESVs) results in too few overlaps in OTUs among the studies making it difficult to make comparisons across interventions.

We then filtered the OTU table and sequences clustered at 97% identity by removing the OTUs with a total abundance (summed across all samples) of less than 10 in QIIME2 using the `filter-features` option. We also used contingency-based filtering in QIIME2 to remove OTUs that appeared in less than 3 samples based on the assumption that these may not represent real biological sequences but rather are sequencing errors or PCR chimeras.

As performed for the studies individually, we trained the SILVA (v132_99_16S) and the Genome Taxonomy Database (GTDB; vbac120_ssu_reps_r95) databases based on the V4 primer sequences from the Earth Microbiome Project to conduct taxonomic classification for the V4 sequences in QIIME2. The reads were first taxonomically classified with the trained SILVA database to remove chloroplast and mitochondrial DNA. The cleaned reads were then assigned a taxonomic group using the trained GTDB database. Only reads that were classified to the phylum level were kept for further analysis.

The merged data were then divided into the different studies while filtering out the "control" samples (e.g. maltodextrin intervention) when present and only keeping the fiber intervention samples (before versus after). Finally, to focus on taxa distributed widely among individuals, we excluded OTUs that were present in less than 50% of the samples per study.

Once again, the data was exported from QIIME2 to be used in R to conduct an in-depth phylogenetic analysis. We first used Phyloseq version 1.34.0 (McMurdie and Holmes 2013) to

format the data to be used in DESeq2. We used the non-rarefied data in DESeq2 to normalize the data per study and to infer what were the main taxonomic groups affected by fiber consumption. We used an alpha value of 0.05 and a $\log_2$-fold change cutoff of 0 and 0.58 (which equals 1.5 fold change) to select the OTUs that significantly were affected by the fiber treatment.

We then selected "widespread" OTUs present in at least 3 studies and aligned their sequences using the Biostrings version 2.58.0 and DECIPHER version 2.18 (Wright 2016) packages to create a neighbor-joining (NJ) tree using phangorn version 2.5.3 package (Schliep 2011).

We averaged the $\log_2$-fold change responses from the widespread OTUs obtained by DESeq2. OTUs with a $\log_2$-fold change higher than zero were considered to be positive responding taxa, whereas the OTUs with a negative $\log_2$-fold change were considered negatively responding taxa. The positive and negative responding taxa were assigned a 1 and a 0 respectively as the ConsenTRAIT program requires the data to be in binary form. We ran ConsenTRAIT analysis using the castor package version 1.3.5 (Louca and Doebeli 2018) to identify phylogenetic clades that respond to fiber intervention in the same direction across studies (consensus clades) and to calculate the average depth (τ_D) of the conserved clades from the NJ phylogenetic tree we created.

Given that we observed a significant phylogenetic signal between the clades that were responding to fiber treatment using the widespread OTUs, we conducted a similar analysis for the individual studies as described above; for each study, a phylogenetic tree was built (using all OTUs present) and a ConsenTRAIT analysis was run to confirm the overall results.

RESULTS

We screened over 1,500 abstracts and obtained data for 21 fiber diet interventions (from 12 studies) performed in healthy humans, for a total of 2,564 samples from 538 subjects (**Fig. 2.1; Table 2.1**). The duration of interventions ranged from 3 days to 84 days with a minimum of two fecal collection timepoints (before and after the diet intervention) but some collected up to 8 times. The types of fibers also varied across fiber interventions, with starches derived from potato being the most common fiber intervention used (**Table 2.2** ; **Table S2.1**). Moreover, the most sequenced region of the 16S rRNA gene was the V4 section, as 8 studies selected it to decipher microbial composition.

Alpha-diversity responses

Short-term increases in dietary fiber consumption resulted in highly consistent gut microbiome responses across studies. To investigate if fiber intervention altered alpha-diversity in the human gut microbiome, we calculated the Shannon and Simpson indices before and after the fiber intervention. Individually, only 5 studies showed a significant decline in bacterial alpha-diversity with both indices (paired-t-test $p<0.05$; **Table 2.2**). However, in 20 out of 21 studies, alpha-diversity tended to decrease at least with one alpha diversity metric (e.g., Shannon). On average, diversity decreased by 3.9% and 1.1% in terms of Shannon and Simpson metrics, respectively (**Fig. 2.2**)

Beta-diversity responses

Increased fiber intake also had a consistent effect on gut microbiome beta-diversity in healthy humans. As expected, inter-individual variation in microbiome composition was high. Microbiome composition differed significantly among individuals in every study, on average explaining 82% of the compositional variation observed (PERMANOVA: $p <0.05$; **Table 2.2**). Despite this variability, in 14 out of 21 studies, it was still possible to detect a significant effect

of the fiber intervention on microbiome composition. Further, the different interventions explained a relatively small but consistent amount of microbiome variation across studies, ranging from 0.2 – 4.6%, for an average of 1.5% of compositional variation (PERMANOVA: $p < 0.05$; **Table 2.2**).

Phylogenetic responses

To detect bacterial taxa that consistently shifted after fiber interventions, we conducted an in-depth re-analysis of 8 studies that amplified the same 16S rRNA genetic region (V4). To be able to make direct comparisons, we combined the sequence data from all studies and then assigned 97% OTUs.

We then analyzed the studies in two ways. First, we identified bacterial taxa that shifted positively or negatively in relative abundance in each study by a standard differential abundance method and then compared the responses of the same OTUs across the studies. After averaging the $\log_2$-fold change responses for the widespread OTUs, we identified 5 bacterial taxa within the Actinomycetia, Clostridia, and Gammaproteobacteria classes that displayed highly positive responses to fiber interventions ($\log_2$-fold change > 1). These positive responding taxa belonged to the families *Bifidobacteriaceae* (three from *Bifidobacterium* genus), *Burkholderiaceae* (one from *Sutterella* genus), and *Ruminococcaceae* (one from *Faecalibacterium* genus). Among these taxa, OTUs belonging to the *Bifidobacteriaceae* family had the highest positive response to fiber with an average of 1.3 positive $\log_2$-fold change, followed by *Burkholderiaceae* and *Ruminococcaceae* with 1.2 and 1.1 $\log_2$-fold change, respectively. We also identified 8 bacterial taxa that showed a highly negative response to fiber treatment ($\log_2$-fold change < -1.0). These taxa all fell within the class Clostridia and belonged to the following families: *CAG-508* (three from *UMGS1994, CAG-354*, and unidentified genus), *Lachnospiraceae* (one from

Mediterraneibacter and three from unidentified genus), and *Ruminococcaceae* (one from *Negativacillus* genus). The OTUs belonging to the *Lachnospiraceae* family had the strongest negative $\log_2$-fold change with an average of -1.4, followed by *CAG-508* and *Ruminococcaceae* with -1.2 and -1.1 $\log_2$-fold change, respectively (**Fig. 2.3**).

Second, we constructed a phylogenetic tree of the OTUs that were represented in at least 3 studies and conducted a ConsenTRAIT analysis. This analysis serves as a metric to identify phylogenetic groups in which a trait or response is conserved (in this case a response to fiber intervention) and calculates the average phylogenetic depth (τ_D) of conservation from a phylogenetic tree (Martiny, Treseder, and Pusch 2013; Isobe et al. 2020). Bacterial responses, positive and negative, to fiber treatment were significantly conserved with an average phylogenetic depth, τ_D, of 0.019 and a 0.020, respectively (permutation test; $p < 0.05$, **Fig. 2.4**). To confirm this significant phylogenetic signal between the "widespread OTUs" across all fiber interventions, we repeated this ConsenTRAIT analysis on the individual studies. This analysis revealed that the degree of conservation for positively responding taxa was in average 0.021 when only considering the studies that gave a significant result (n=6). Conversely, the degree of conservation across negatively responding taxa was slightly lower, 0.019, and it was only significant in 4 out of the 8 studies (**Table 2.3**).

DISCUSSION

To test the hypothesis that short-term increases in fiber consumption would result in consistent changes in microbiome composition, we re-analyzed bacterial 16S rRNA data from different fiber interventions in healthy humans. We found evidence supporting this hypothesis as we observed consistent gut microbiome responses in alpha- and beta-diversity as well as in a phylogenetic context. Overall, fiber consumption decreased alpha diversity and explained an

average of 1.5% of bacterial compositional variation (compared to 82% of variation explained by individuals). Moreover, when conducting an in-depth phylogenetic analysis, we observed that bacterial responses (positive and negative) were phylogenetically conserved, and the depth of conservation was similar across studies.

Diet is known to play a major role in the gut microbiota composition including diversity and richness (Graf et al. 2015). It is often thought that a diet rich in fiber increases both diversity (e.g., richness) and functioning (e.g., production of short chain fatty acids [SCFAs]), which have been used as markers of intestinal health (Makki et al. 2018). However, we observed that fiber intake consistently reduced bacterial diversity in 20 out of 21 studies, with this pattern being significant in 5 interventions. This is in line with previous work that has assessed the impact of increased consumption of specific fiber components or an overall high-fiber diet (Hooda et al. 2012; Oliver et al. 2021; So et al. 2018; Wastyk et al. 2021; Deehan et al. 2020; Liu et al. 2017). As it has been stated before, this could be due to the length of the interventions as the majority are short term (some being as short as 3 days). Thus, we might only be capturing the transitional period of microbial responses to fiber, where we observe the loss of microbes that are not well adapted to the changing environment (e.g., decreased pH due to increased fermentation) and cannot consume the carbohydrates found in fiber. Perhaps, the cascading effects of a sustained high-fiber diet on bacterial diversity are not observed in these short time frames, as this might be a relative slow process. Furthermore, when increasing fiber intake, humans are not ingesting more microbes that can colonize the human gut and instead an increase in bacterial diversity is most likely observed in interventions with fermented aliments, where the presence of microbes and their byproducts in the foods have a faster impact in overall bacterial richness that can be captured in short-term interventions (Wastyk et al. 2021).

Even though an increase in bacterial diversity was not detected in short-term fiber interventions, we were able to identify conserved responses in relative abundances among bacterial taxa. In particular, OTUs belonging to the genus *Bifidobacterium* showed the strongest positive response across fiber interventions (**Fig. 2.3**). This bacterial clade is known for its carbohydrate degrading capabilities and its positive associations with health benefits in humans (O'Callaghan and van Sinderen 2016). Indeed, this bacterial genus is generally found to increase in abundance following an increased fiber intake (So et al. 2018; Oliver et al. 2021; Davis et al. 2011). *Bifidobacterium* species can act as primary fiber degraders as they are known to possess carbohydrate active enzymes (CAZymes) that allow the degradation of various plant carbohydrates (Kelly, Munoz-Munoz, and van Sinderen 2021; Fushinobu and Abou Hachem 2021; Baxter et al. 2019).

In addition, the *Sutterella* genus also demonstrated a significant increase after fiber intervention. This pattern had been previously reported once in a fiber intervention with women suffering with hypertensive disorders during pregnancy (Tomsett et al. 2020). However, *Sutterella* species have been associated with health disorders such as autism and metabolic syndrome (L. Wang et al. 2013; Lim et al. 2017). Nonetheless, the *Sutterella* genus is also present in healthy humans, and their ability to adhere to intestinal epithelial cells might indicate a mutualistic relationship with its host (Hiippala et al. 2016). Although its functional role is not well understood, this might be an important microbe to consider in future fiber interventions. Overall, the ability for bacteria to survive the changing environment (e.g., low pH due to increased fermentation) and attach to fiber substrates grants them access to these carbohydrates, promoting rapid increases in abundance following a fiber intervention, while negative responding taxa might lack these abilities.

Although microbial responses to fiber interventions are thought to be highly individualized (Martínez et al. 2010; Baxter et al. 2019), we found a significant phylogenetic signal of bacterial taxa that respond to fiber interventions. Specifically, we found that bacterial taxa that respond positively or negatively to fiber intake showed a significant average phylogenetic depth of conservation (τ_D= 0.019 and τ_D= 0.020 ; $p<0.05$; **Fig. 2.4**). This depth is similar to that previously found for nitrogen fixation traits (τ_D= 0.018-0.020), but was more deeply conserved than that of simple carbon utilization (τ_D= 0.011) (Martiny, Treseder, and Pusch 2013; Amend et al. 2016; Isobe et al. 2019) and the ability to produce extracellular enzymes (τ_D= 0.008–0.01) (Zimmerman, Martiny, and Allison 2013). Moreover, the average depth of bacterial responses to fiber intervention was consistent across studies (τ_D= 0.014–0.028; **Table 2.3**). Though phylogenetic trees built with 16S rRNA amplicon sequences are not as reliable as multi-locus trees, they are still useful to estimate the depth of the response to fiber interventions and to compare this response with other traits that have been analyzed previously.

It is important to note that our results come with certain limitations inherent to the re-analysis of publicly available data and time-course-gut microbiome samples. Perhaps, the most notable caveat is that we were able to obtain sequencing data from 12 studies and only 8 of them shared the same 16S rRNA region for an in-depth phylogenetic analysis. Although they represent different populations and fiber interventions, the inclusion of more studies with more diverse populations and different fiber types would increase our ability to predict bacterial responses due to fiber intake. Additionally, since we are looking at temporal variation in the gut microbiome, this change could be due to external factors rather than the intervention itself. However, we found trends suggesting that at least some of this change is due to the intervention. For example,

we observed consistent shifts in alpha-diversity and similar responses in bacterial taxa across studies and individuals.

In the future, similar data syntheses for other types of interventions or diseases could be useful to get a more detailed understanding of which bacterial clades respond more readily to perturbations. Also, if more studies report that the individual variation is generally as high as 82%, as found here, then it cannot be expected to see an intervention effect higher than single digits. This is not to say that these small effects are not important – within a person, those changes could be quite consequential. For these reasons, it is crucial to elucidate the overall changes in microbial composition, as this gives a more general view on the effects of interventions in the human gut and help decipher the bacterial taxa that are influenced by drastic changes in diet or gut perturbations.

CONCLUSION

The research presented here allows us to understand the impact dietary fiber has on the human gut microbiome. We found that short-term increases in fiber intake resulted in consistent changes in microbiome composition across different types of interventions, such as decreased alpha diversity and similar amount of compositional variation explained by fiber intake. Moreover, we documented that the subject where the fecal sample comes from explained a similarly high variation in bacterial composition across all interventions (inter-individual variation), allowing for future studies to have a baseline expectation of the variation that should be explained by the subject itself. Additionally, we showed that a phylogenetic approach, that has been previously used to test bacterial trait conservation in environmental samples (Martiny, Treseder, and Pusch 2013; Isobe et al. 2019; 2020), can be useful to disentangle the bacterial responses to a dietary change in the human gut microbiome. Indeed, we demonstrated that

despite the high microbial variation in human subjects, this method can be applied to human related microbiomes to identify bacterial clades that are responsive to dietary changes and their average phylogenetic depth of conservation.

ACKNOWLEDGMENTS

We would like to thank all the authors of the studies mentioned here for making their data available for re-analysis. We would also like to thank Kristin Barbour for her feedback on this manuscript and the members of the J. Martiny lab for their encouragement. This work was supported by a Faculty Mentor Program (FMP) fellowship, UC President's Dissertation Year Fellowship, and the Rose Hills Foundation Science & Engineering Fellowship.

REFERENCES

Abell, Guy C. J., Caroline M. Cooke, Corinna N. Bennett, Michael A. Conlon, and Alexandra L. McOrist. 2008. "Phylotypes Related to Ruminococcus Bromii Are Abundant in the Large Bowel of Humans and Increase in Response to a Diet High in Resistant Starch." *FEMS Microbiology Ecology* 66 (3): 505–15. https://doi.org/10.1111/j.1574-6941.2008.00527.x.

Amend, Anthony S., Adam C. Martiny, Steven D. Allison, Renaud Berlemont, Michael L. Goulden, Ying Lu, Kathleen K. Treseder, Claudia Weihe, and Jennifer B. H. Martiny. 2016. "Microbial Response to Simulated Global Change Is Phylogenetically Conserved and Linked with Functional Potential." *The ISME Journal* 10 (1): 109–18. https://doi.org/10.1038/ismej.2015.96.

Andrews, Simon. 2010. "Babraham Bioinformatics - FastQC A Quality Control Tool for High Throughput Sequence Data." 2010. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Baxter, Nielson T., Alexander W. Schmidt, Arvind Venkataraman, Kwi S. Kim, Clive Waldron, and Thomas M. Schmidt. 2019. "Dynamics of Human Gut Microbiota and Short-Chain Fatty Acids in Response to Dietary Interventions with Three Fermentable Fibers." *MBio* 10 (1): e02566-18. https://doi.org/10.1128/mBio.02566-18.

Bushnell, Brian, Jonathan Rood, and Esther Singer. 2017. "BBMerge – Accurate Paired Shotgun Read Merging via Overlap." *PLoS ONE* 12 (10): e0185056. https://doi.org/10.1371/journal.pone.0185056.

Callahan, Benjamin J., Paul J. McMurdie, Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, and Susan P. Holmes. 2016. "DADA2: High-Resolution Sample Inference from

Illumina Amplicon Data." *Nature Methods* 13 (7): 581–83. https://doi.org/10.1038/nmeth.3869.

Carlson, Justin L, Jennifer M Erickson, Beate B Lloyd, and Joanne L Slavin. 2018. "Health Effects and Sources of Prebiotic Dietary Fiber." *Current Developments in Nutrition* 2 (3): nzy005. https://doi.org/10.1093/cdn/nzy005.

Chen, Shifu, Yanqing Zhou, Yaru Chen, and Jia Gu. 2018. "Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor." *Bioinformatics* 34 (17): i884–90. https://doi.org/10.1093/bioinformatics/bty560.

Cronin, Peter, Susan A. Joyce, Paul W. O'Toole, and Eibhlís M. O'Connor. 2021. "Dietary Fibre Modulates the Gut Microbiota." *Nutrients* 13 (5): 1655. https://doi.org/10.3390/nu13051655.

Davis, Lauren M. G., Inés Martínez, Jens Walter, Caitlin Goin, and Robert W. Hutkins. 2011. "Barcoded Pyrosequencing Reveals That Consumption of Galactooligosaccharides Results in a Highly Specific Bifidogenic Response in Humans." *PLOS ONE* 6 (9): e25200. https://doi.org/10.1371/journal.pone.0025200.

Deehan, Edward C., Chen Yang, Maria Elisa Perez-Muñoz, Nguyen K. Nguyen, Christopher C. Cheng, Lucila Triador, Zhengxiao Zhang, Jeffrey A. Bakal, and Jens Walter. 2020. "Precision Microbiome Modulation with Discrete Dietary Fiber Structures Directs Short-Chain Fatty Acid Production." *Cell Host & Microbe* 27 (3): 389-404.e6. https://doi.org/10.1016/j.chom.2020.01.006.

Del Chierico, Federica, Pamela Vernocchi, Bruno Dallapiccola, and Lorenza Putignani. 2014. "Mediterranean Diet and Health: Food Effects on Gut Microbiota and Disease Control." *International Journal of Molecular Sciences* 15 (7): 11678–99. https://doi.org/10.3390/ijms150711678.

Durack, Juliana, and Susan V. Lynch. 2018. "The Gut Microbiome: Relationships with Disease and Opportunities for Therapy." *Journal of Experimental Medicine* 216 (1): 20–40. https://doi.org/10.1084/jem.20180448.

Estaki, Mehrbod, Lingjing Jiang, Nicholas A. Bokulich, Daniel McDonald, Antonio González, Tomasz Kosciolek, Cameron Martino, et al. 2020. "QIIME 2 Enables Comprehensive End-to-End Analysis of Diverse Microbiome Data and Comparative Studies with Publicly Available Data." *Current Protocols in Bioinformatics* 70 (1): e100. https://doi.org/10.1002/cpbi.100.

Fushinobu, Shinya, and Maher Abou Hachem. 2021. "Structure and Evolution of the Bifidobacterial Carbohydrate Metabolism Proteins and Enzymes." *Biochemical Society Transactions* 49 (2): 563–78. https://doi.org/10.1042/BST20200163.

Garud, Nandita R., Benjamin H. Good, Oskar Hallatschek, and Katherine S. Pollard. 2019. "Evolutionary Dynamics of Bacteria in the Gut Microbiome within and across Hosts." *PLOS Biology* 17 (1): e3000102. https://doi.org/10.1371/journal.pbio.3000102.

Gonzalez, Antonio, Jose A. Navas-Molina, Tomasz Kosciolek, Daniel McDonald, Yoshiki Vázquez-Baeza, Gail Ackermann, Jeff DeReus, et al. 2018. "Qiita: Rapid, Web-Enabled Microbiome Meta-Analysis." *Nature Methods* 15 (10): 796–98. https://doi.org/10.1038/s41592-018-0141-9.

Graf, Daniela, Raffaella Di Cagno, Frida Fåk, Harry J. Flint, Margareta Nyman, Maria Saarela, and Bernhard Watzl. 2015. "Contribution of Diet to the Composition of the Human Gut Microbiota." *Microbial Ecology in Health and Disease* 26 (February): 10.3402/mehd.v26.26164. https://doi.org/10.3402/mehd.v26.26164.

Hiippala, Kaisa, Veera Kainulainen, Marko Kalliomäki, Perttu Arkkila, and Reetta Satokari. 2016. "Mucosal Prevalence and Interactions with the Epithelium Indicate Commensalism of Sutterella Spp." *Frontiers in Microbiology* 7. https://www.frontiersin.org/articles/10.3389/fmicb.2016.01706.

Holscher, Hannah D. 2017. "Dietary Fiber and Prebiotics and the Gastrointestinal Microbiota." *Gut Microbes* 8 (2): 172–84. https://doi.org/10.1080/19490976.2017.1290756.

Hooda, Seema, Brittany M. Vester Boler, Mariana C. Rossoni Serao, Jennifer M. Brulc, Michael A. Staeger, Thomas W. Boileau, Scot E. Dowd, George C. Fahey Jr, and Kelly S. Swanson. 2012. "454 Pyrosequencing Reveals a Shift in Fecal Microbiota of Healthy Adult Men Consuming Polydextrose or Soluble Corn Fiber." *The Journal of Nutrition* 142 (7): 1259–65. https://doi.org/10.3945/jn.112.158766.

Hosseini-Esfahani, Firoozeh, Gelareh Koochakpoor, Maryam S. Daneshpour, Parvin Mirmiran, Bahareh Sedaghati-khayat, and Fereidoun Azizi. 2017. "The Interaction of Fat Mass and Obesity Associated Gene Polymorphisms and Dietary Fiber Intake in Relation to Obesity Phenotypes." *Scientific Reports* 7 (December): 18057. https://doi.org/10.1038/s41598-017-18386-8.

Isobe, Kazuo, Steven D. Allison, Banafshe Khalili, Adam C. Martiny, and Jennifer B. H. Martiny. 2019. "Phylogenetic Conservation of Bacterial Responses to Soil Nitrogen Addition across Continents." *Nature Communications* 10 (1): 2499. https://doi.org/10.1038/s41467-019-10390-y.

Isobe, Kazuo, Nicholas J. Bouskill, Eoin L. Brodie, Erika A. Sudderth, and Jennifer B. H. Martiny. 2020. "Phylogenetic Conservation of Soil Bacterial Responses to Simulated Global Changes." *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (1798): 20190242. https://doi.org/10.1098/rstb.2019.0242.

Jones, Julie Miller. 2014. "CODEX-Aligned Dietary Fiber Definitions Help to Bridge the 'Fiber Gap.'" *Nutrition Journal* 13 (1): 34. https://doi.org/10.1186/1475-2891-13-34.

Jovel, Juan, Levinus A. Dieleman, Dina Kao, Andrew L. Mason, and Eytan Wine. 2018. "Chapter 10 - The Human Gut Microbiome in Health and Disease." In *Metagenomics*, edited by Muniyandi Nagarajan, 197–213. Academic Press. https://doi.org/10.1016/B978-0-08-102268-9.00010-0.

Kelly, Sandra M., Jose Munoz-Munoz, and Douwe van Sinderen. 2021. "Plant Glycan Metabolism by Bifidobacteria." *Frontiers in Microbiology* 12. https://www.frontiersin.org/articles/10.3389/fmicb.2021.609418.

Kovatcheva-Datchary, Petia, Anne Nilsson, Rozita Akrami, Ying Shiuan Lee, Filipe De Vadder, Tulika Arora, Anna Hallen, Eric Martens, Inger Björck, and Fredrik Bäckhed. 2015. "Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of Prevotella." *Cell Metabolism* 22 (6): 971–82. https://doi.org/10.1016/j.cmet.2015.10.001.

Lim, Mi Young, Hyun Ju You, Hyo Shin Yoon, Bomi Kwon, Jae Yoon Lee, Sunghee Lee, Yun-Mi Song, Kayoung Lee, Joohon Sung, and GwangPyo Ko. 2017. "The Effect of Heritability and Host Genetics on the Gut Microbiota and Metabolic Syndrome." *Gut* 66 (6): 1031–38. https://doi.org/10.1136/gutjnl-2015-311326.

Liu, Feitong, Pan Li, Muxuan Chen, Yuemei Luo, M Prabhakar, Huimin Zheng, Yan He, et al. 2017. "Fructooligosaccharide (FOS) and Galactooligosaccharide (GOS) Increase Bifidobacterium but Reduce Butyrate Producing Bacteria with Adverse Glycemic Metabolism in Healthy Young Population." *Scientific Reports* 7 (September): 11789. https://doi.org/10.1038/s41598-017-10722-2.

Louca, Stilianos, and Michael Doebeli. 2018. "Efficient Comparative Phylogenetics on Large Trees." *Bioinformatics* 34 (6): 1053–55. https://doi.org/10.1093/bioinformatics/btx701.

Makki, Kassem, Edward C. Deehan, Jens Walter, and Fredrik Bäckhed. 2018. "The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease." *Cell Host & Microbe* 23 (6): 705–15. https://doi.org/10.1016/j.chom.2018.05.012.

Martin, Marcel. 2011. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads." *EMBnet.Journal* 17 (1): 10–12. https://doi.org/10.14806/ej.17.1.200.

Martínez, Inés, Jaehyoung Kim, Patrick R. Duffy, Vicki L. Schlegel, and Jens Walter. 2010. "Resistant Starches Types 2 and 4 Have Differential Effects on the Composition of the Fecal Microbiota in Human Subjects." *PLOS ONE* 5 (11): e15046. https://doi.org/10.1371/journal.pone.0015046.

Martiny, Adam C, Kathleen Treseder, and Gordon Pusch. 2013. "Phylogenetic Conservatism of Functional Traits in Microorganisms." *The ISME Journal* 7 (4): 830–38. https://doi.org/10.1038/ismej.2012.160.

McMurdie, Paul J., and Susan Holmes. 2013. "Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data." *PLOS ONE* 8 (4): e61217. https://doi.org/10.1371/journal.pone.0061217.

O'Callaghan, Amy, and Douwe van Sinderen. 2016. "Bifidobacteria and Their Role as Members of the Human Gut Microbiota." *Frontiers in Microbiology* 7 (June). https://doi.org/10.3389/fmicb.2016.00925.

Oliver, Andrew, Alexander B. Chase, Claudia Weihe, Stephanie B. Orchanian, Stefan F. Riedel, Clark L. Hendrickson, Mi Lay, Julia Massimelli Sewall, Jennifer B. H. Martiny, and Katrine Whiteson. 2021. "High-Fiber, Whole-Food Dietary Intervention Alters the Human Gut Microbiome but Not Fecal Short-Chain Fatty Acids." *MSystems* 6 (2): e00115-21. https://doi.org/10.1128/mSystems.00115-21.

Parks, Donovan H., Maria Chuvochina, David W. Waite, Christian Rinke, Adam Skarshewski, Pierre-Alain Chaumeil, and Philip Hugenholtz. 2018. "A Standardized Bacterial Taxonomy Based on Genome Phylogeny Substantially Revises the Tree of Life." *Nature Biotechnology* 36 (10): 996–1004. https://doi.org/10.1038/nbt.4229.

Quast, Christian, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, and Frank Oliver Glöckner. 2013. "The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools." *Nucleic Acids Research* 41 (D1): D590–96. https://doi.org/10.1093/nar/gks1219.

Sawicki, Caleigh M., Kara A. Livingston, Martin Obin, Susan B. Roberts, Mei Chung, and Nicola M. McKeown. 2017. "Dietary Fiber and the Human Gut Microbiota: Application of Evidence Mapping Methodology." *Nutrients* 9 (2): 125. https://doi.org/10.3390/nu9020125.

Schliep, Klaus Peter. 2011. "Phangorn: Phylogenetic Analysis in R." *Bioinformatics* 27 (4): 592–93. https://doi.org/10.1093/bioinformatics/btq706.

Schnorr, Stephanie L., Marco Candela, Simone Rampelli, Manuela Centanni, Clarissa Consolandi, Giulia Basaglia, Silvia Turroni, et al. 2014. "Gut Microbiome of the Hadza Hunter-Gatherers." *Nature Communications* 5 (April). https://doi.org/10.1038/ncomms4654.

Shreiner, Andrew B., John Y. Kao, and Vincent B. Young. 2015. "The Gut Microbiome in Health and in Disease." *Current Opinion in Gastroenterology* 31 (1): 69–75. https://doi.org/10.1097/MOG.0000000000000139.

Simpson, H. L., and B. J. Campbell. 2015. "Review Article: Dietary Fibre–Microbiota Interactions." *Alimentary Pharmacology & Therapeutics* 42 (2): 158–79. https://doi.org/10.1111/apt.13248.

Singh, Rasnik K., Hsin-Wen Chang, Di Yan, Kristina M. Lee, Derya Ucmak, Kirsten Wong, Michael Abrouk, et al. 2017. "Influence of Diet on the Gut Microbiome and Implications for Human Health." *Journal of Translational Medicine* 15 (April): 73. https://doi.org/10.1186/s12967-017-1175-y.

So, Daniel, Kevin Whelan, Megan Rossi, Mark Morrison, Gerald Holtmann, Jaimon T. Kelly, Erin R. Shanahan, Heidi M. Staudacher, and Katrina L. Campbell. 2018. "Dietary Fiber Intervention on Gut Microbiota Composition in Healthy Adults: A Systematic Review and Meta-Analysis." *The American Journal of Clinical Nutrition* 107 (6): 965–83. https://doi.org/10.1093/ajcn/nqy041.

Thompson, Luke R., Jon G. Sanders, Daniel McDonald, Amnon Amir, Joshua Ladau, Kenneth J. Locey, Robert J. Prill, et al. 2017. "A Communal Catalogue Reveals Earth's Multiscale Microbial Diversity." *Nature* 551 (7681): 457–63. https://doi.org/10.1038/nature24621.

Tian, Tian, Xiaobo Zhang, Tao Luo, Duolao Wang, Yuping Sun, and Jianghong Dai. 2021. "Effects of Short-Term Dietary Fiber Intervention on Gut Microbiota in Young Healthy People." *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy* 14 (August): 3507–16. https://doi.org/10.2147/DMSO.S313385.

Tomsett, Kate I., Helen L. Barrett, Evelyn E. Dekker, Leonie K. Callaway, David H. McIntyre, and Marloes Dekker Nitert. 2020. "Dietary Fiber Intake Alters Gut Microbiota Composition but Does Not Improve Gut Wall Barrier Function in Women with Future Hypertensive Disorders of Pregnancy." *Nutrients* 12 (12): 3862. https://doi.org/10.3390/nu12123862.

Wang, Lv, Claus T. Christophersen, Michael J. Sorich, Jacobus P. Gerber, Manya T. Angley, and Michael A. Conlon. 2013. "Increased Abundance of Sutterella Spp. and Ruminococcus Torques in Feces of Children with Autism Spectrum Disorder." *Molecular Autism* 4 (1): 42. https://doi.org/10.1186/2040-2392-4-42.

Wang, Shumin, Yue Xiao, Fengwei Tian, Jianxin Zhao, Hao Zhang, Qixiao Zhai, and Wei Chen. 2020. "Rational Use of Prebiotics for Gut Microbiota Alterations: Specific Bacterial Phylotypes and Related Mechanisms." *Journal of Functional Foods* 66 (March): 103838. https://doi.org/10.1016/j.jff.2020.103838.

Wastyk, Hannah C., Gabriela K. Fragiadakis, Dalia Perelman, Dylan Dahan, Bryan D. Merrill, Feiqiao B. Yu, Madeline Topf, et al. 2021. "Gut-Microbiota-Targeted Diets Modulate Human Immune Status." *Cell* 184 (16): 4137-4153.e14. https://doi.org/10.1016/j.cell.2021.06.019.

Whisner, Corrie M., Juan Maldonado, Brandon Dente, Rosa Krajmalnik-Brown, and Meg Bruening. 2018. "Diet, Physical Activity and Screen Time but Not Body Mass Index Are Associated with the Gut Microbiome of a Diverse Cohort of College Students Living in University Housing: A Cross-Sectional Study." *BMC Microbiology* 18 (1): 210. https://doi.org/10.1186/s12866-018-1362-x.

Wright, Erik, S. 2016. "Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R." *The R Journal* 8 (1): 352. https://doi.org/10.32614/RJ-2016-025.

Wu, Gary D., Jun Chen, Christian Hoffmann, Kyle Bittinger, Ying-Yu Chen, Sue A. Keilbaugh, Meenakshi Bewtra, et al. 2011. "Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes." *Science (New York, N.y.)* 334 (6052): 105–8. https://doi.org/10.1126/science.1208344.

Yang, Jing, Hai-Peng Wang, Li Zhou, and Chun-Fang Xu. 2012. "Effect of Dietary Fiber on Constipation: A Meta Analysis." *World Journal of Gastroenterology : WJG* 18 (48): 7378–83. https://doi.org/10.3748/wjg.v18.i48.7378.

Zimmerman, Amy E, Adam C Martiny, and Steven D Allison. 2013. "Microdiversity of Extracellular Enzyme Genes among Sequenced Prokaryotic Genomes." *The ISME Journal* 7 (6): 1187–99. https://doi.org/10.1038/ismej.2012.176.

FIGURES

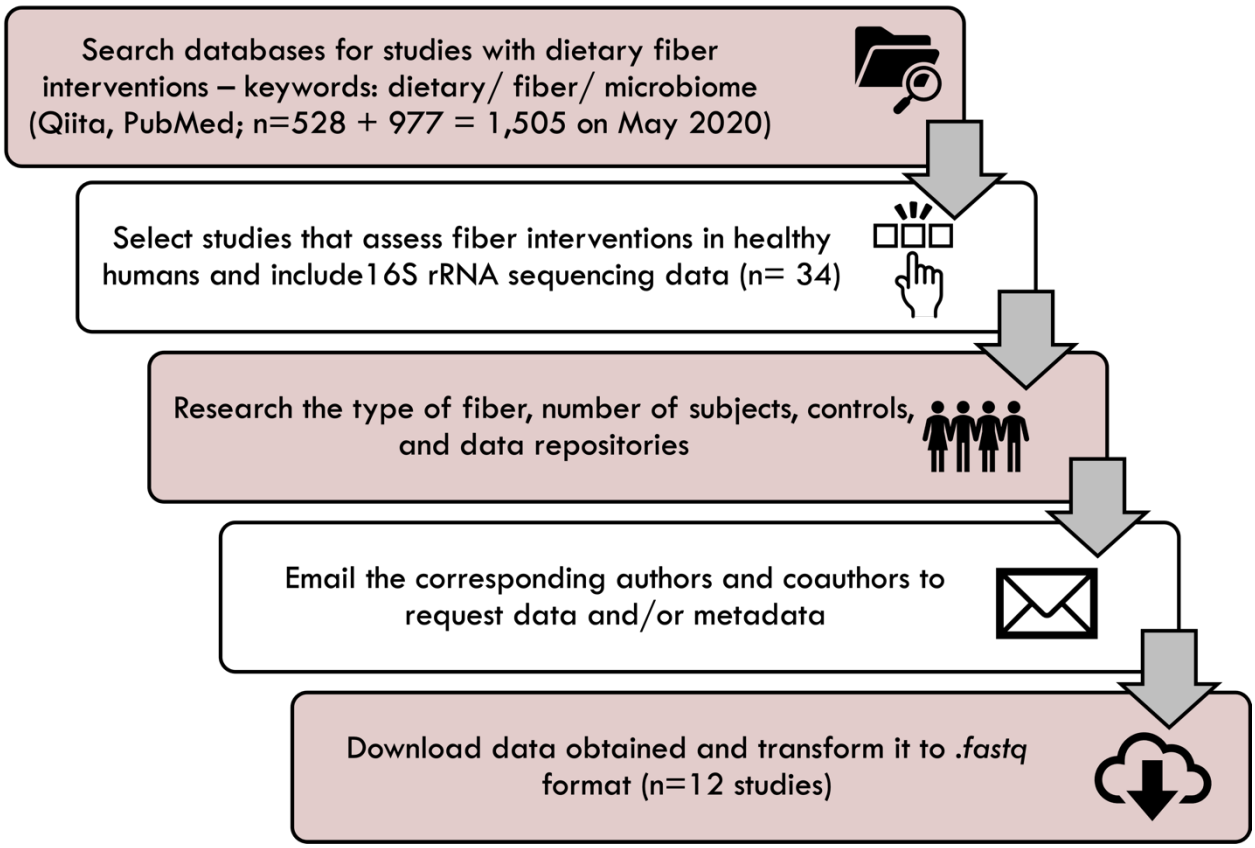


Figure 2.1 Data collection overview. The step-by-step workflow for search and data collection of fiber intervention studies.

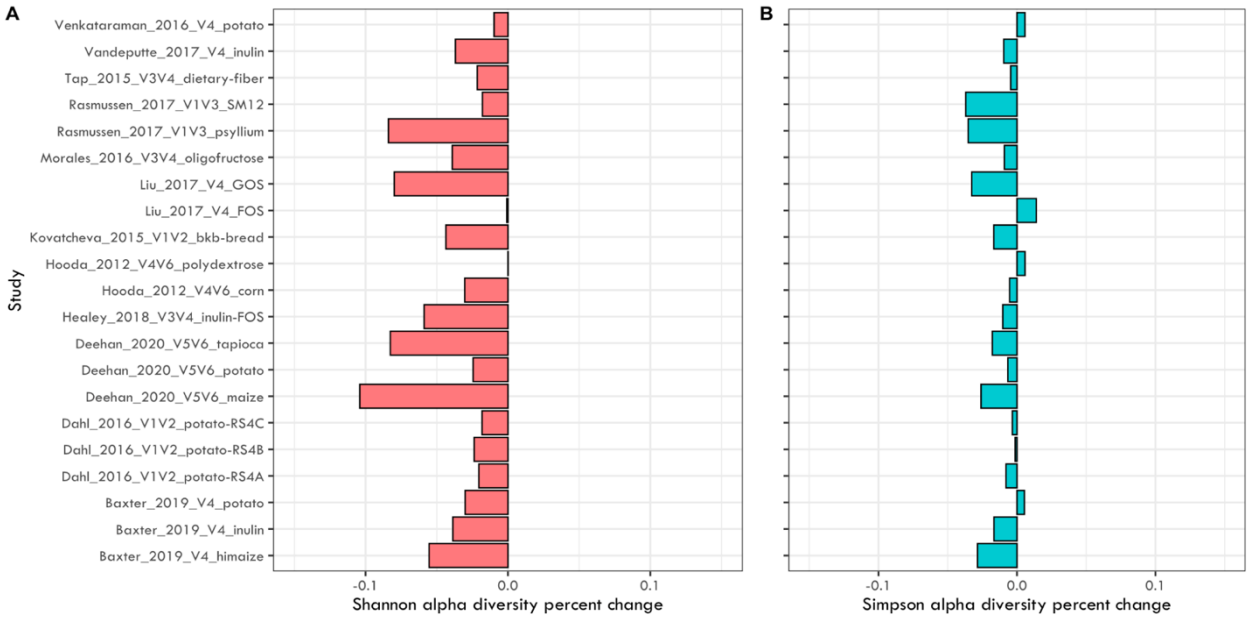


Figure 2.2 Percent change for alpha diversity metrics A) Shannon and B) Simpson indices. Alpha diversity metrics were calculated using ESVs and rarefied data (See Methods for details). Percent change was measured by subtracting the before-fiber intervention alpha diversity mean from the after-fiber intervention mean using only two timepoints. When multiple timepoints where available, only the first and the last were used for alpha diversity analysis. Coral bars represent the Shannon while blue bars represent the Simpson percent change.

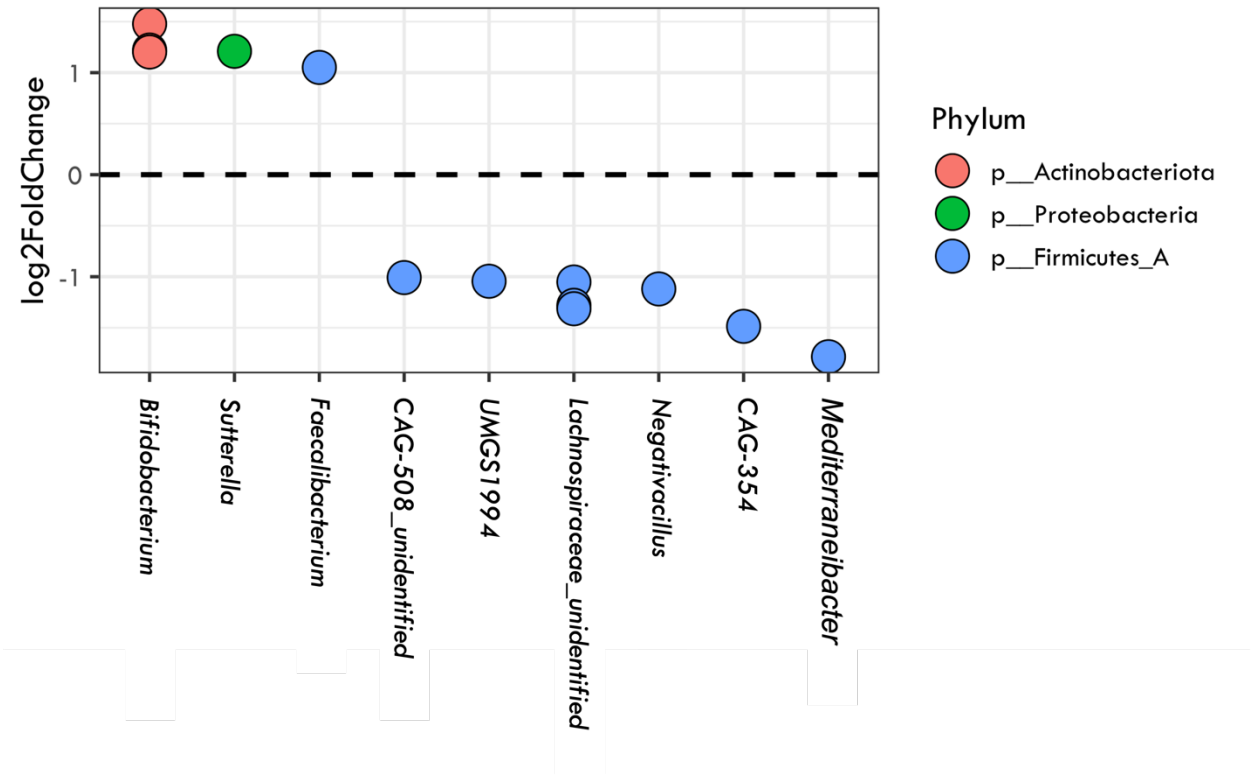


Figure 2.3 Averaged bacterial responses across fiber interventions sharing the V4 region. The points represent bacterial clades that had a strong response in abundance based on the log_2-fold change using the widespread OTUs (present in at least 3 studies) obtained by DESeq2. We only included taxa that had an average log_2-fold response higher than 1 and lower than -1. The data points are named based on their taxonomic classification (GTDB) at the genus level. When there was no genus identified we added their Family assignment followed by "unidentified."

Bacterial response	Average depth (τ_D)	*p*-value
Positive	0.019	0.010
Negative	0.020	0.033

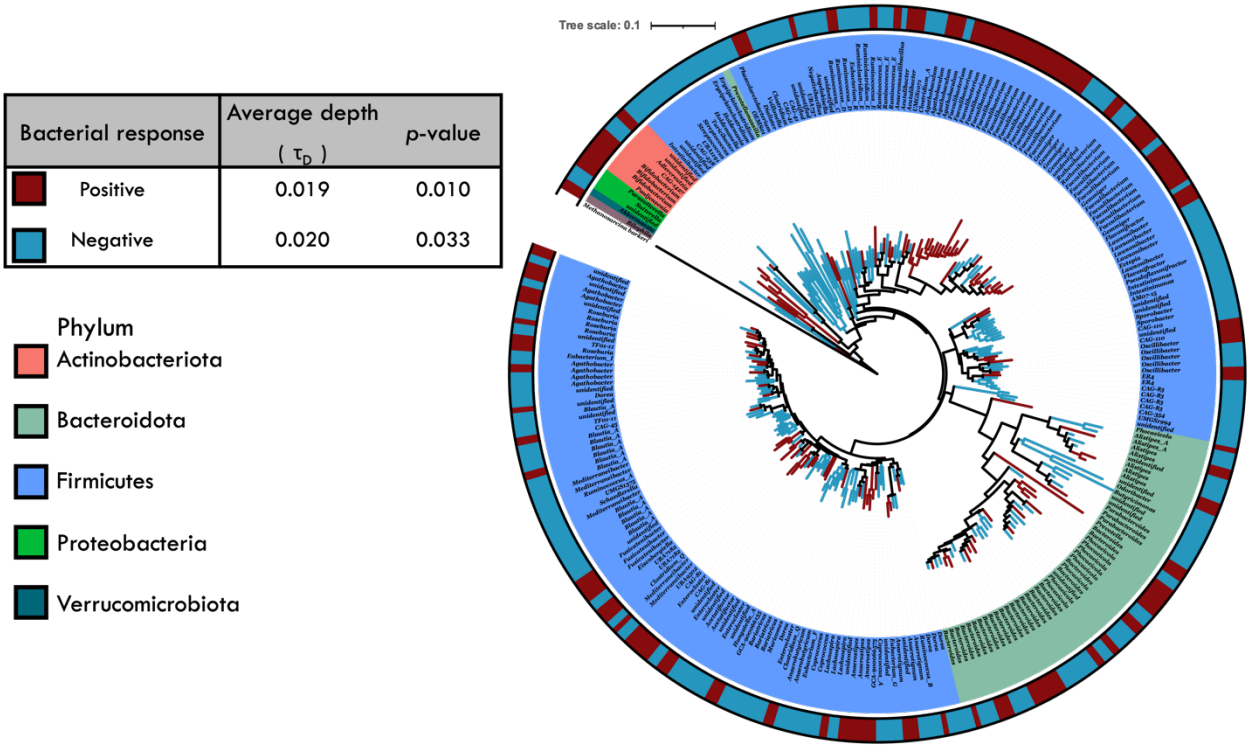


Figure 2.4 Phylogenetic distribution of the averaged responses to fiber intervention. The widespread OTUs (present in at least 3 studies) are colored based on their response to fiber (positive or negative). The inner ring represents the phylum-level taxonomy of the OTUs determined using the GTDB trained database. The average depth of conservation is shown on the right panel next to their respective p-values.

Table 2.1. Summary of datasets collected including fiber type, grams of fiber used, duration of the intervention, collection timepoints and number of subjects per study.

Study Name	Number of interventions	Fibers used in intervention + control when applicable	Amount of fiber or control given in intervention (grams)	Duration of intervention (days)	Collection timepoints	Number of subjects	Number of samples
Baxter_2019_V4	4	Resistant starch from potatoes (RPS), resistant starch from maize (RMS), inulin from chicory root, and an accessible corn starch control	20-40	14	8	175	1,205
Dahl_2016_V1V2	3	RS-4-A, RS-4-B, RS-4-C - Resistant potato starches (RS type 4)	30	14	4	53	212
Deehan_2020_V5V6	4	Tapioca, potato, and maize- Resistant starches (RS type 4) + corn starch control	increasing from 0 to 10 , 20, 35, and 50	28	5	40	200
Healey_2018_V3V4	2	50:50 inulin to fructo-oligosaccharide and maltodextrin control	16	21	4	34	134
Hooda_2012_V4V6	2	Polydextrose and soluble corn fiber control	21	21	3	10	28
Kovatcheva_2015_V1V2	2	Kernel-based bread (BKB) and white-wheat-bread (WWB)	37.6 & 9.1	3	3	20	60
Liu_2017_V4	2	Fructooligosaccharides (FOS) and galactooligosaccharides (GOS)	16	14	4	35	132
Morales_2016_V3V4	2	Oligofructose and maltodextrin control (extra treatments of Orlistat were also given)	16	7	2	41	82
Rasmussen_2017_V1V3	2	Starch-entrapped microspheres and psyllium	9 & 12	84	2	41	82
Tap_2015_V3V4	NA	Dietary fiber meals	10 & 40	5	4	19	76
Vandeputte_2017_V4	2	Inulin and maltodextrin control	12	28	4	50	196
Venkataraman_2016_V4	1	Resistant starch (unmodified potato starch; RS type 2)	48	17	8	20	157

Table 2.2. Summary of samples included per fiber intervention and the alpha- and beta-diversity results.

Study Name	Number of samples included in analysis	Number of subjects	Rarefaction depth	Shift in alpha diversity after fiber treatment: Shannon & Simpson metrics *ǂ	Beta diversity: Subject variation explained (%) ǂ	Beta diversity: Fiber treatment variation explained (%) ǂ
Baxter_2019_V4_himaize	313	43	4891	↓ **significant**	**86**	**0.2**
Baxter_2019_V4_inulin	365	50	4546	↓ not significant	**84**	**0.6**
Baxter_2019_V4_potato	273	43	4622	↓ not significant – only in Shannon	**86**	**0.7**
Dahl_2016_V1V2_potato-RS4A	34	17	16289	↓ not significant	**89**	**1**
Dahl_2016_V1V2_potato-RS4B	36	18	15957	↓ not significant	**88**	**1**
Dahl_2016_V1V2_potato-RS4C	36	18	8135	↓ not significant	**87**	0.8
Deehan_2020_V5V6_maize	50	10	27488	↓ **significant**	**87**	**1**
Deehan_2020_V5V6_potato	50	10	18744	↓ not significant	**85**	0.5
Deehan_2020_V5V6_tapioca	50	10	7157	↓ **significant**	**79**	**1**
Healey_2018_V3V4_inulin-FOS	68	34	6014	↓ **significant**	**85**	**1.6**
Hooda_2012_V4V6_corn	19	10	3689	↓ not significant	**75**	**4.4**
Hooda_2012_V4V6_polydextrose	19	10	2966	↑ not significant	**76**	**4.6**
Kovatcheva_2015_V1V2_bkb-bread	40	20	3642	↓ not significant	**86**	0.6
Liu_2017_V4_FOS	66	34	1929	↓ not significant – only in Shannon	**80**	0.9
Liu_2017_V4_GOS	66	34	1465	↓ not significant	**78**	**1.5**
Morales_2016_V3V4_oligofructose	22	11	66061	↓ not significant	**88**	**1.6**
Rasmussen_2017_V1V3_SM12	30	15	3217	↓ not significant	**74**	1.6
Rasmussen_2017_V1V3_psyllium	24	12	1197	↓ not significant	**76**	2.6
Tap_2015_V3V4_dietary-fiber	38	19	1021	↓ not significant	**71**	1.4
Vandeputte_2017_V4_inulin	96	49	7912	↓ **significant**	**83**	**0.7**
Venkataraman_2016_V4_potato	157	20	2167	↓ not significant – only in Shannon	**84**	**0.9**
Average					82	1.4
Average - significant only					**82**	**1.5**

* = comparison done through paired-t test
ǂ = bolded signifies significance $p < 0.05$

Table 2.3. ConsenTRAIT analysis for individual studies.

Study	# of samples	# OTUs at 97% identity	# of samples used	Min. # of samples the OTUs needed to be present	OTUs after 50% presence filtering	#Sig. OTUs p.adj <0.05 & >1.5 fold change	#Sig. OTUs p.adj <0.05 & >0 fold change	Positive responding OTUs	Negative responding OTUs	Positive τ_D ǂ	Negative τ_D ǂ
Baxter_2019_V4	1,205	22,222	954	477	137	4	40	52	85	**0.028**	0.020
Healey_2018_V3V4	134	14,211	68	34	312	23	24	140	172	**0.017**	**0.019**
Hooda_2012_V4V6	28	4,957	28	14	208	23	23	115	93	0.018	**0.020**
Liu_2017_V4	132	4,523	132	66	86	12	12	31	55	0.019	0.029
Morales_2016_V3V4	82	24,943	22	11	1,044	11	11	494	550	**0.014**	**0.015**
Tap_2015_V3V4	74	5,377	37	19	128	4	4	82	46	**0.026**	0.015
Vandeputte_2017_V4	196	22,702	96	48	463	22	22	136	327	**0.016**	**0.021**
Venkataraman_2016_V4	157	10,847	157	79	179	11	23	88	91	**0.023**	0.021
Average										0.020	0.020
Average -significant only										**0.021**	**0.019**

ǂ = bolded signifies significance $p < 0.05$

Supplemental Table 2.1. Summary of fibers used in intervention studies.

Fiber type	Description/manufacturer	Study
Resistant starch from potatoes (RPS)	Bob's Red Mill, Milwaukee, OR	Baxter, 2019
Inulin from chicory root	Swanson Health Products, Fargo, ND	Baxter, 2019
Hi-Maize 260 resistant corn starch (RMS)	manufactured by Ingredion Inc., Westchester, IL, and distributed by myworldhut.com	Baxter, 2019
Amylase-accessible corn starch (placebo)	Amioca powder; Skidmore Sales and Distribution, West Chester, OH	Baxter, 2019
Resistant potato starch RS4-A	PenFibe® RO – 170; phosphorylated, soluble fibre with high viscosity - Penford Food Ingredients Inc., Denver, CO, USA	Dahl, 2016
Resistant potato starch RS4-B	PenFibe® RO – 177; hydrolysed, phosphorylated, soluble fibre with low viscosity - Penford Food Ingredients Inc., Denver, CO, USA	Dahl, 2016
Resistant potato starch RS4-C	PenFibe® RS; insoluble fibre with low viscosity - Penford Food Ingredients Inc., Denver, CO, USA	Dahl, 2016
AMIOCA™ Powder TF (Placebo)	Ingredion Inc, Bridgewater, NJ 08807, USA	Deehan, 2020
VERSAFIBE™ 2470 (Maize RS4)	Ingredion Inc, Bridgewater, NJ 08807, USA	Deehan, 2020
VERSAFIBE™ 1490 (Potato RS4)	Ingredion Inc, Bridgewater, NJ 08807, USA	Deehan, 2020
VERSAFIBE™ 3490 (Tapioca RS4)	Ingredion Inc, Bridgewater, NJ 08807, USA	Deehan, 2020
Orafti® Synergy1 – 50:50 inulin to fructo-oligosaccharide mix	Beneo GmbH	Healey, 2018
Glucidex® 29 Premium-digestible maltodextrin; placebo	Roquette Worldwide	Healey, 2018
Polydextrose	PDX; Litesse II, Danisco	Hooda, 2012
Soluble corn fiber (placebo)	SCF; PROMITOR, Tate and Lyle Ingredients	Hooda, 2012
Kernel-based bread (KBB)	NA	Kovatcheva-Datchary, 2015
White-wheat-bread (WWB)	NA	Kovatcheva-Datchary, 2015
Fructooligosaccharide- FOS (QHT-Purity95%)	source: Sucrose; Quantum Hi-Tech (China) Biological company, Guangdong, China	Liu, 2017
Galactooligosaccharide- GOS (QHT- Purity95%)	source: lactose; Quantum Hi-Tech (China) Biological company, Guangdong, China	Liu, 2017
Maltodextrin (placebo)	NA	Morales, 2016
Oligofructose	NA	Morales, 2016
Starch-entrapped microspheres (SM)	A suspension of sodium alginate (2% w/v) and normal corn starch (9% w/v) was made in water through a special recipe	Rasmussen, 2017
Psyllium	Natural Foods Inc (Toledo, OH)	Rasmussen, 2017
Dietary fiber meals (different foods)	NA	Tap, 2015
Chicory-derived inulin	Orafti inulin	Vandeputte, 2017
Maltodextrin (placebo)	NA	Vandeputte, 2017
Raw unmodified potato starch	Bob's Red Mill, Milwaukie, OR. This potato starch contains approximately 50 % resistant starch (type 2) by weight.	Venkataraman, 2016

CHAPTER 3

The consequences of bifidobacteria diversity for ecosystem functioning

INTRODUCTION

To understand the relationship between biodiversity and ecosystem functioning (BEF), plant and animal ecologists often experimentally manipulate diversity and compare how functioning varies. These studies have demonstrated that for an aggregate of functional measurements, species richness increases functioning in an array of ecosystems (Guerrero-Ramírez et al. 2017; Hooper and Vitousek 1997; David Tilman, Isbell, and Cowles 2014). Fewer BEF experiments have been conducted with bacterial communities, but those that have, focus on environmental habitats including soil, phyllosphere, aquatic, and industrial ecosystems such as waste bioreactors (Awasthi et al. 2014; Bell et al. 2009; 2005; Evans et al. 2017; Gravel et al. 2011; Johnson et al. 2015). However, tests that experimentally control species or trait richness of human gut bacteria are lacking (Allen-Vercoe 2013; Venema and van den Abbeele 2013; Reese and Dunn 2018).

It is important to study the BEF relationships in human associated microbiomes, particularly the human gut, as the loss of bacterial richness in this ecosystem (a metric of alpha diversity) has been correlated with many health disorders such as obesity, insulin resistance, and inflammatory diseases (Le Chatelier et al. 2013; Lloyd-Price, Abu-Ali, and Huttenhower 2016). Furthermore, healthy adults from rural areas and hunter-gatherer communities generally exhibit higher gut bacterial richness when compared to populations in metropolitan or urban societies (Schnorr et al. 2014; Tasnim et al. 2017; Yatsunenko et al. 2012). In addition, obesity and allergic diseases that develop in childhood have been linked with low bacterial richness, and in particular, a low abundance of bifidobacteria (Akay et al. 2014; Kalliomäki et al. 2008; 2001;

Liwen et al. 2018; O'Neill, Schofield, and Hall 2017). Thus, in the present study, we tested whether a BEF relationship might underlie some of the diversity patterns observed in human studies while focusing on gut-isolated bifidobacteria (genus *Bifidobacterium*), a widespread common genus in gut microbiomes that has been correlated with a wide array of health benefits (Hidalgo-Cantabrana et al. 2017).

Bifidobacteria are often associated with human newborns and infants as specific species of bifidobacteria can degrade human milk oligosaccharides derived from breastmilk and this appears to aid in the development of a healthy immune system (O'Neill, Schofield, and Hall 2017; Katayama 2016). Although bifidobacteria abundance decreases with age (Arboleya et al. 2016), their presence in the adult human gut contributes to protection from pathogen colonization, production of beneficial metabolites, and the degradation of fiber (Turroni, Milani, Van Sinderen, et al. 2018). Due to these positive associations, bifidobacteria are a popular probiotic added to yogurt and dietary supplements (Hidalgo-Cantabrana et al. 2017). While human bifidobacteria have been well studied, most studies focus on one strain at a time, without considering the variety that often coexist. Indeed, humans can harbor a diversity of bifidobacteria species in their gut at any one time (Milani, Mancabelli, et al. 2015; Turroni et al. 2009; 2012; Oliver et al. 2021), and the composition and abundance of this diversity varies over their lifetime (Arboleya et al. 2016; Turroni et al. 2009). However, the impacts of such diversity have been unexplored.

Bifidobacteria richness might increase gut functioning (particularly digestion) via carbohydrate degradation. As observed in Chapter 1 and other studies (Egan and Van Sinderen 2018), the genes involved in carbohydrate breakdown (e.g., CAZymes) are abundant and variable even within the same species of bifidobacteria (Arboleya et al. 2018). As such,

increased bifidobacteria diversity might allow for a broader use of different carbohydrates given that strains might partition these resources. Higher richness might further increase functioning because of cross-feeding between strains (Turroni, Milani, Duranti, et al. 2018). Indeed, co-cultivation of two bifidobacteria species has been shown to increase the metabolic activity of strains, suggesting that bifidobacterial isolates can utilize the byproducts from carbohydrate digestion of others (Egan et al. 2014; Milani, Lugli, et al. 2015; Turroni, Milani, Duranti, et al. 2018; Turroni et al. 2015). To our knowledge, no one has tested whether the richness of bifidobacteria alters their overall functioning.

To test how the richness and diversity of bifidobacteria influences their coexistence and ecosystem functioning (i.e., biomass and production of lactate), we conducted an *in vitro* microcosm experiment varying richness (1-8 strains) and functional groups (1-2 groups). We characterized the functioning of the bacterial communities using flow cytometry (to count bacterial cells), lactate assays (as a proxy to fiber degradation), and next-generation sequencing (to decipher the bacterial community's composition). We hypothesized a positive BEF relationship where increasing bacterial richness and diversity would increase functioning (measured by cell biomass and fiber degradation). As for plant communities (D. Tilman 2001; David Tilman, Reich, and Knops 2006; David Tilman, Isbell, and Cowles 2014), this positive BEF relationship might occur because of resource specialization, synergistic species interactions, and/or sampling effects, where more diverse communities are more likely to contain higher functioning strains. Specifically, we expected that: 1) different strains of bifidobacteria would co-exist; 2) monocultures would have lower functioning (biomass and lactate production) in comparison to co-cultures and would increase further with co-culture richness; and 3)

functioning would increase further in co-cultures with more functional diversity (two versus one functional group).

METHODS

Strain collection and verification

We obtained all the available bifidobacteria isolates from the Human Microbiome Project (HMP) through the Biodefense and Emerging Infections Research Resources Repository (BEI Resources; https://www.beiresources.org). From the 13 strains obtained, eight were isolated from the human gut (**Table S3.1**). We also selectively isolated new potential strains of bifidobacteria from fecal samples from 16 different healthy adults collected for a fiber intervention conducted at the University of California in Irvine (Oliver et al. 2021) using *Bifidobacterium* selective media (BSM). To verify that the bacteria isolated belong to the *Bifidobacterium* genus, we extracted their DNA using ZymoBIOMICS™ DNA Miniprep Kit and amplified the *groEL* gene with bifidobacteria specific primers (Hu et al. 2017). As expected, all 13 HMP strains amplified the *groEL* gene at the correct amplicon size (**Fig. S3.1**). From the 52 strains we isolated from fecal samples, 32 positively amplified the *groEL* gene (**Fig. S3.1**). Using the whole genomic DNA extracted, we prepared a shotgun sequencing library using a low-volume Illumina procedure (Weihe and Avelar-Barragan 2021) and sequenced it using Illumina HiSeq 4000 (2 X150 bp) at the Genomics High Throughput Facility, UC Irvine, CA, USA.

Genome assembly and taxonomic classification

Sequenced reads were visualized with FastQC version 0.11.8 to check quality parameters (Andrews 2010), and then Fastp version 0.20.0 was used to remove reads with an average quality score lower than 34 and sequencing adapters. We also enabled the *--correction* parameter to find the overlap across the paired-end reads (Chen et al. 2018). To confirm that the reads were

processed correctly by Fastp, we visualized the reads once again with FastQC. We then assembled genomes using SPAdes (3.13.0) with *--careful* and *--cov-cutoff-auto* parameters (Nurk et al. 2017). To assess the quality of the contigs assembled, we used QUAST version 5 (Gurevich et al. 2013). We map the reads back to the assembled contigs using Bowtie version 2.2.7 and Samtools version 1.9. The output of these tools was used to create Taxon-Annotated-GC-Coverage plots (TAGC plots) with Blobtools version 1.1.1 and Blastn version 2.3.1. After selecting the genomes that were positively assigned to the *Bifidobacterium* genus based on Blobtools, we polished the assembled contigs of these isolates by removing contigs shorter than 100 bp using seqkit version 0.11.0 (Shen et al. 2016). Lastly, we estimated genome completeness using BUSCO (Simão et al. 2015).

Phylogenetic, pangenome, and CAZyme analyses

To predict open reading frames (ORF) in the assembled bifidobacterial genomes we used Prodigal version 2.6.3 with default parameters (Hyatt et al. 2010). We then constructed a multi-locus phylogenetic tree using the bcgTree pipeline (Ankenbrand and Keller 2016) with the protein fasta files (.*faa) obtained with Prodigal. Briefly, the phylogeny is built based on 107 conserved genes that are aligned with Muscle software version 3.8.31 and polished using Gblocks version 0.91b. The 107 genes were concatenated, and a phylogenetic tree was built using an isolate from the *Alloscardovia* genus as an outgroup with RAxML version 8.2.10 with the GTRGAMMA substitution model and 100 interactions. We then visualized this tree using iTOL (Letunic and Bork 2016) to select 25 genetically-diverse isolates for further analysis.

To find functional genes that differentiated the bifidobacteria isolates, we performed a pangenome analysis and identified their carbon utilization enzymes. Pangenome information was derived from Roary version 3.12.0 (Page et al. 2015) using .gff files of the annotated genomes

obtained from Prokka version 1.13.3 (Seemann 2014). We used a 70% Blastp cutoff sequence identity. To visualize the pangenome we created a heatmap using the presence-absence of gene clusters given by Roary in Phandango (Hadfield et al. 2018). To identify Carbohydrate Active Enzymes (CAZymes) in the selected bifidobacteria isolates, we used the dbCAN2 server (Zhang et al. 2018). We only considered CAZymes that were identified by three prediction tools within the dbCAN2 program using default parameters. From the 25 isolates, we selected 16 strains from adults that were distinct, based on the phylogenetic, pangenome, and CAZyme analyses (**Fig. 3.1**; **Fig. 3.2**; **Fig. 3.3**). We measured the average nucleotide identity (ANI) of the selected isolates using the FastANI software version 1.32 with default parameters (Jain et al. 2018).

Microcosms set-up

The experiment consisted of 75 microcosms, including treatments varying in isolate richness (1, 2, 4, 8 strains) and functional groups (A, B, or AB), together with 3 replicates for each combination (**Fig. 3.4;** the results section explains how we identified the functional groups). The microcosms were constructed in 50 ml sterile conical vials with lids and incubated in anaerobic conditions (H2 5%, CO2 5%, N2 90%) at 37 °C. Standard liquid media used to grow bifidobacteria, De Man-Rogosa-Sharpe (MRS) broth (Sigma-Aldrich), was modified by adding 1 mL of Tween 80 per liter and 1.5% w/v of lentil flour (Red Lentil Flour by Hearthy Foods) as a source of complex carbohydrates. We also included 3 sterile media controls.

Before starting the microcosms, the isolates (previously stored in glycerol stocks at -70 °C) were revived on MRS agar plates in anaerobic conditions (H2 5%, CO2 5%, N2 90%) at 37 °C for 72 h. From these plates, one colony per isolate was picked and individually grown in MRS + 1.5% w/v lentil flour broth for 48h to acclimate the strains to liquid broth. After this, the

corresponding strains were mixed in a standardized manner based on optical density into the 50 ml sterile vials with MRS + 1.5% lentil media, marking the start of the microcosm experiment.

Throughout the experiment, samples were collected for analysis at 0h, 6h, 15h, 24h, 39h, 48h, 63h, and 72h with the experiment concluding at this timepoint. The period of 72h was chosen for three reasons: 1) based on previous growth curve data, which indicated that all strains grew to stationary phase by 48 hours; 2) past bifidobacteria *in vitro* experiments have used this timeframe (Egan et al. 2014); and 3) food's transit time in the human intestine as it takes between 24 and 72 h for food to move through the entire gastrointestinal tract (Read et al. 1980).

Cell abundance and lactate production

Total productivity of the microcosms was assayed by flow cytometry and lactate production. We used the NovoCyte® flow cytometer to obtain the bacterial cell counts in each microcosm over time (0hr, 6hr, 15hr, 24hr, 39hr, 48hr, 63hr, and 72hr). Cell cultures from microcosms were stored with 1% final concentration of glutaraldehyde (GTA) at 4 °C until ready to proceed with flow cytometry counts (~ 1 week). The day of flow cytometry counts, samples were brought out from the 4 °C fridge to room temperature and diluted 1:125 for timepoints at 0hr and 6hr, the remaining samples for the latter timepoints were diluted to 1:2500, as they had higher cell densities. Cells were stained with 1X final dilution of SYBR green 200X (200x, Invitrogen Life Science Technologies, S756) and incubated for at least 10 minutes before proceeding with counts. Samples were run for 30 seconds at 20 μL/min, using a SYBR-Green-H threshold value of 2,000.

We used lactate concentration as a proxy for carbohydrate fermentation. Theoretically, bifidobacteria produce 1 mol of lactate for every mol of fermented glucose as part of their bifid shunt metabolic pathway (Pokusaeva, Fitzgerald, and van Sinderen 2011; Palframan, Gibson,

and Rastall 2003). To measure lactate in the media, we filtered the microcosm cultures through a 0.22 μm syringe-filter and stored the supernatant at -80 °C until ready to perform the assay. We performed a colorimetric assay using the Lactate Assay Kit (MAK064 Millipore Sigma) and measured the absorbance at 570nm.

Sequence analysis

We extracted DNA from the microcosm communities following the ZymoBIOMICS™ DNA Miniprep Kit protocol and prepared the shotgun sequencing library using a low-volume Illumina procedure (Weihe and Avelar-Barragan 2021). To examine community assembly, we sequenced the extracted DNA from the microcosms using NovaSeq S4 technology (2 X 150) with a 5% phiX spike at 15h, 24h, 48h, and 72h collection time points at the Genomics High Throughput Facility, UC Irvine, Irvine, CA, USA.

To characterize the bacterial community in the microcosms, we first visualized the reads using FastQC software (version 0.11.9) for quality control purposes. Then, the Fastp software version 0.20.0 was used to detect and remove: sequencing adapters *(--detect_adapter_for_pe*), reads with an average quality lower than 30 *(--average_qual 30*), and reads shorter than 50bp *(--length_required 50*). We also enabled the base pair correction parameter which attempts to find an overlap of each pair of reads to improve accuracy of base calls (*--correction*). To verify the identity of the isolates in the monocultures, we assembled the genomes using SPAdes version 3.14.0 with high-coverage data (*--isolate*) and coverage cut off value on auto as parameters.

To define the community composition of the microcosms containing co-cultures we used the MIDAS program using the monoculture genomes as reference for mapping reads and calling SNPs (Nayfach et al. 2016). Briefly, we created our own database using the 16 bifidobacteria isolate genomes through the build_midasdb.py script. We provided the fasta genomic files (.fna),

the protein sequences (.faa), the gene sequences (.ffn), and a tab delimited file with the genomic coordinates of the genes (.genes) by converting the .gff extension files using the csvtk tool. All the mentioned files were obtained from Prodigal version 2.6.3. Lastly, we ran MIDAS with our custom database using the run_midas.py species script. Isolates that had less than 200 reads assigned were considered extinct.

Statistical analyses

To detect differences in the productivity metrics (biomass and lactate concentration) across richness and functional groups and their possible interaction, we ran two-way ANOVA tests. To confirm our results and account for unevenness in variance, we also ran Welch one-way ANOVA and Games Howell Post-hoc tests with the package 'rstatix' in R. To compare total biomass and lactate data between monocultures and co-cultures across functional groups and richness levels, we conducted Welch two sample t-tests. We ran Spearman's correlations for lactate concentration at 24h versus 72h, between biomass data and lactate data at 72h, and across the coefficient of variation and productivity metrics (biomass and lactate) at 72h using. All tests were conducted using R version 4.0.2. We used a significance level threshold of $p < 0.05$ across all tests.

RESULTS

Genomic features of bifidobacteria isolates

On average the sequenced bifidobacterial genomes had 35 contigs (>1000 bp), a total genome length of 2.36 Mb (>1000 bp), and a GC content of 59.5% (**Table S3.1**), similar to other bifidobacteria genomes (Lee and O'Sullivan 2010; Milani et al. 2014; Turroni, van Sinderen, and Ventura 2011). From here, we selected 25 genomes isolated from the human gut. A phylogenetic tree constructed with 107 concatenated core genes shows that most of the isolates in our pool of

bifidobacteria were from the *B. longum* species (n=15), followed by *B. adolescentis* (n=5), *B. pseudocatenulatum* (n=2), *B. animalis* (n=1), *B. angulatum* (n=1), and *B. breve* (n=1; **Fig. 3.1**).

The pangenome analysis revealed that this pool of bifidobacteria has 550 core genes (present in 100% of the genomes), 128 soft core genes (present in > 95% and less than 100%), 2,568 shell genes (present in >15% and less than 95%), and 4,770 cloud genes (present in > 0% and less than 15%) (**Fig. 3.2**). A CAZyme analysis identified GT2_Glycos_transf_2 as the most abundant CAZyme at the gene subfamily level. The number of carbohydrate-degrading genes in a genome ranged from 47 (isolate CRM19 from the *B. animalis* species) to 87 (isolate CRM24 from the *B. longum* species). Both the pangenome and CAZyme analyses supported a clustering of two distinct groups among the genomes (**Fig. 3.2** and **3.3**); for simplicity, we will hereafter refer to these groupings as Functional group A and Functional group B. We then selected 16 isolates evenly from these functional groups to be included in our microcosm experiment (**Fig 3.4**). These isolates had an average nucleotide identity ranging from 80-98.6% (**Fig. S3.2**). Here, we define a bacterial strain as a non-clonal bacterial isolate within the genus that could be from the same or different species, belonging to either Functional Group A or B.

Microcosm community composition over time

We sequenced the microcosms at 15h, 24h, 48h, and 72h timepoints to assess the bacterial community composition over time. As predicted, the majority of strains persisted over the course of the experiment, with seven of eight strains persisting at the highest richness level (**Fig. 3.5**). The exceptions were the following strains that went extinct in all the communities they were included: CRM26 *B. pseudocatenulatum* in 3 microcosms (i.e., 4_A, 8_A, 8_AB), CRM03 *B. longum* in 2 microcosms (4_B and 8_B), and CRM14 *B. longum* in one microcosm (8_B). Note that all these strains persisted in monoculture during the experiment (**Figure S3.3**).

By 15h, the bacterial communities already displayed an uneven distribution of the remaining strains, but bacterial composition remained relatively stable through the rest of the experiment (**Fig. 3.5**). Lastly, community composition was remarkably consistent, with similar relative abundances for each particular community among the replicates (**Fig. 3.6**).

Community biomass

We measured cell abundance (a proxy for total biomass) of all microcosms at 8 timepoints over the 3 days of the experiment. The majority of the communities reached stationary phase by 24h (**Fig. S3.3**). Neither richness nor functional group diversity significantly affected community biomass as determined by cell counts (two-way ANOVA, $p > 0.05$). Moreover, there was not a statistically significant interactive effect of richness and functional group on biomass (two-way ANOVA, $F(5, 64) = 0.878$, $p = 0.501$); **Fig. 3.7A**).

Further, bifidobacteria interactions amongst the co-cultures did not generally increase overall biomass at any richness level; co-culture biomass was similar to that predicted by the individual monocultures at both 24 and 72 hours (**Fig. 3.8A and B**). However, variation across microcosms tended to decrease as richness increased, although the correlation between variation and richness was not significant (**Fig. S3.4A;** Spearman's $\varrho = -0.8$, $p>0.05$).

Lactate concentration as a measure of ecosystem functioning

As a more direct proxy for carbohydrate fermentation, we measured lactate concentration in the microcosms at 24 and 72 hours. Indeed, there was no significant correlation between biomass and lactic acid production at 72 hours (Spearman's $\varrho = 0.099$, $p> 0.05$: **Fig. S3.5**). Although most microcosms reached stationary phase by 24h (**Fig. S3.3**), lactate production continued to increase from 24h to 72h in all microcosms (**Fig. S3.6**) and there was a strong positive correlation between the timepoints (Spearman's $\varrho= 0.73 < 0.05$; **Fig. S3.7**). In contrast

to biomass, lactate production by 72 hours increased with richness level (two-way ANOVA, $p < 0.05$). However as for biomass, there was not a statistically significant interaction between the effects of richness and functional group on lactate concentration (two-way ANOVA, $F(5, 64) = 0.608$, $p = 0.694$). To account for unevenness in variance, we corroborated these results with a Welch one-way ANOVA and richness was still significant (Welch ANOVA, $F(3,22.42) = 6.74$, $p = 0.002$, $n = 75$; **Fig. 3.7B**).

Finally, bifidobacteria interactions amongst the co-cultures appeared to increase overall lactate concentration relative to the concentration predicted from the monocultures, particularly for co-cultures of 4 and 8 strains (Welch two-sample-t-test, $p < 0.05$; **Fig. 3.8C and D**). As with biomass, variation among the cultures decreased as richness increased although again, this correlation was not significant (**Fig. S3.4B;** Spearman's $\varrho = -0.1$, $p > 0.05$).

DISCUSSION

To test how diversity of bifidobacteria potentially influences their gut ecosystem functioning, we conducted an *in vitro* microcosm experiment varying bacterial richness and functional group diversity. Up to 7 bifidobacteria strains were able to coexist in this *in vitro* system. We found a positive BEF relationship as measured by lactate concentration, but not overall biomass. Further, this relationship seemed to be driven by resource specialization or synergistic interactions amongst the strains, as the co-cultures at the highest richness levels produced more lactate than expected based on the monocultures alone. In contrast, the relationship did not appear to be solely due to sampling effects, as lactate concentrations of the co-cultures were higher than that of all the monocultures. Finally, there was no evidence that functional diversity (the presence of strains from two versus one functional group) increased overall functioning.

Why might the BEF trends for biomass and lactate production differ? First, cell abundance might not accurately reflect total biomass. Flow cytometry is an accurate method of counting, but it is possible that cell size could have also been changing and our metric does not account for that. Alternatively, lactate production might better reflect cell activity/functioning than overall biomass. Indeed, that cultures appeared to be growth limited by 24 hours of the experiment as they leveled off in cell counts, and the cultures might further have been growth limited because of the acidity of the cultures. Thus, the cells might have continued to be active without additional growth. Overall, lactate concentration is a better metric for ecosystem functioning in this system as it can serve as a measure for fiber degradation – a direct productivity metric. As seen in plant ecosystems (Kohyama et al. 2020; Hofmann and Jahufer 2011), primary productivity is not always correlated with total biomass. Hence, the contrasting results between biomass and lactate metrics could be explained by a trade-off between bacterial growth and overall activity. Perhaps a greater productivity (higher fiber degradation) in more rich and diverse communities comes at a cost for the individual growth of bacteria. Future research can aim to address this by incorporating metabolomics and identifying possible cross-feeding compounds, measuring what carbohydrates are being spent from the media across different diversity levels, and corroborating that higher bacterial richness results in higher fiber degradation.

Bifidobacteria are known to exhibit social behavior through carbohydrate resource sharing (Turroni, Milani, Duranti, et al. 2018). Previous research has demonstrated syntrophic interactions among the bifidobacteria species. For instance, *B. breve* species show syntrophic interaction with members of the *Bifidobacterium, Bacteroides* and *Lactobacillus* genera (Egan et al. 2014; Munoz et al. 2020; Cheng et al. 2020). Similarly, it has been shown that lactate produced by bifidobacteria and their partial broken down carbohydrates serve as cross-feeding

molecules for other bacteria (Belenguer et al. 2006; Turroni et al. 2015; Turroni, Milani, Duranti, et al. 2018; Turroni et al. 2016). We further expected that functional diversity based on distinction functional gene differences would lead to clearer resource partitioning and thereby, higher functioning. For example, the HM-856 isolate showed the most distinct collection of genes (and CAZymes) in comparison to the other strains in our pool, suggesting that this isolate has a different functional capacity that allows it to feed from the byproducts of the surrounding bacteria. However, we did not observe evidence that functional diversity (the presence of strains from two versus one functional group) increased overall functioning.

CONCLUSION

Our study shows that bifidobacterial strains are able to co-exist in *in vitro* conditions allowing for future investigation of the impacts that bifidobacteria diversity can have on the human gut microbiome. This diversity appears to have a functional impact in fiber degradation, suggesting that it is not just the presence and total abundance of the genus *Bifidobacterium* that will affect gut functioning, but also its strain diversity. Future work should further aim to disentangle the specifics of resource partitioning and cross-feeding interactions. Ultimately, these results will begin to uncover the ecological mechanisms behind the observed correlations between gut bacterial diversity and human health.

ACKNOWLEDGEMENTS

We thank Andrew Oliver and the UCI Microbiome Center for providing the fecal samples for bifidobacteria isolation. Moreover, we want to express our gratitude to Tania Rodriguez for her invaluable help in assembling the microcosm experiments. We also thank Claudia Weihe for her help in setting up the anaerobic chamber and Dr. Matthew Bracken for advice on the experimental design. We further thank Dr. Katrine Whiteson and Dr. Brandon

Gaut for their encouragement and feedback throughout the project and the J. Martiny lab for their support and critique on figures.

REFERENCES

Akay, Hatice Kubra, Hrisi Bahar Tokman, Nevin Hatipoglu, Huseyin Hatipoglu, Rengin Siraneci, Mehmet Demirci, Baris Ata Borsa, et al. 2014. "The Relationship between Bifidobacteria and Allergic Asthma and/or Allergic Dermatitis: A Prospective Study of 0-3 Years-Old Children in Turkey." *Anaerobe* 28 (August): 98–103. https://doi.org/10.1016/j.anaerobe.2014.05.006.

Allen-Vercoe, Emma. 2013. "Bringing the Gut Microbiota into Focus through Microbial Culture: Recent Progress and Future Perspective." *Current Opinion in Microbiology* 16 (5). https://doi.org/10.1016/j.mib.2013.09.008.

Andrews, Simon. 2010. "Babraham Bioinformatics - FastQC A Quality Control Tool for High Throughput Sequence Data." 2010. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Ankenbrand, Markus J., and Alexander Keller. 2016. "BcgTree: Automatized Phylogenetic Tree Building from Bacterial Core Genomes." Edited by Frédéric Chain. *Genome* 59 (10): 783–91. https://doi.org/10.1139/gen-2015-0175.

Arboleya, Silvia, Francesca Bottacini, Mary O'Connell-Motherway, C. Anthony Ryan, R. Paul Ross, Douwe van Sinderen, and Catherine Stanton. 2018. "Gene-Trait Matching across the Bifidobacterium Longum Pan-Genome Reveals Considerable Diversity in Carbohydrate Catabolism among Human Infant Strains." *BMC Genomics* 19 (January). https://doi.org/10.1186/s12864-017-4388-9.

Arboleya, Silvia, Claire Watkins, Catherine Stanton, and R. Paul Ross. 2016. "Gut Bifidobacteria Populations in Human Health and Aging." *Frontiers in Microbiology* 7 (August). https://doi.org/10.3389/fmicb.2016.01204.

Awasthi, Ashutosh, Mangal Singh, Sumit K Soni, Rakshapal Singh, and Alok Kalra. 2014. "Biodiversity Acts as Insurance of Productivity of Bacterial Communities under Abiotic Perturbations." *The ISME Journal* 8 (12): 2445–52. https://doi.org/10.1038/ismej.2014.91.

Belenguer, Alvaro, Sylvia H. Duncan, A. Graham Calder, Grietje Holtrop, Petra Louis, Gerald E. Lobley, and Harry J. Flint. 2006. "Two Routes of Metabolic Cross-Feeding between Bifidobacterium Adolescentis and Butyrate-Producing Anaerobes from the Human Gut." *Applied and Environmental Microbiology* 72 (5): 3593–99. https://doi.org/10.1128/AEM.72.5.3593-3599.2006.

Bell, Thomas, Andrew K. Lilley, Andy Hector, Bernhard Schmid, Lindsay King, and Jonathan A. Newman. 2009. "A Linear Model Method for Biodiversity–Ecosystem Functioning Experiments." *The American Naturalist* 174 (6): 836–49. https://doi.org/10.1086/647931.

Bell, Thomas, Jonathan A. Newman, Bernard W. Silverman, Sarah L. Turner, and Andrew K. Lilley. 2005. "The Contribution of Species Richness and Composition to Bacterial Services." *Nature* 436 (7054): 1157–60. https://doi.org/10.1038/nature03891.

Chen, Shifu, Yanqing Zhou, Yaru Chen, and Jia Gu. 2018. "Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor." *Bioinformatics* 34 (17): i884–90. https://doi.org/10.1093/bioinformatics/bty560.

Cheng, Christopher C., Rebbeca M. Duar, Xiaoxi Lin, Maria Elisa Perez-Munoz, Stephanie Tollenaar, Jee-Hwan Oh, Jan-Peter van Pijkeren, et al. 2020. "Ecological Importance of Cross-Feeding of the Intermediate Metabolite 1,2-Propanediol between Bacterial Gut Symbionts." *Applied and Environmental Microbiology* 86 (11): e00190-20. https://doi.org/10.1128/AEM.00190-20.

Egan, Muireann, Mary O'Connell Motherway, Michelle Kilcoyne, Marian Kane, Lokesh Joshi, Marco Ventura, and Douwe van Sinderen. 2014. "Cross-Feeding by Bifidobacterium Breve UCC2003 during Co-Cultivation with Bifidobacterium Bifidum PRL2010 in a Mucin-Based Medium." *BMC Microbiology* 14 (1): 282. https://doi.org/10.1186/s12866-014-0282-7.

Egan, Muireann, and Douwe Van Sinderen. 2018. "Chapter 8 - Carbohydrate Metabolism in Bifidobacteria." In *The Bifidobacteria and Related Organisms*, edited by Paola Mattarelli, Bruno Biavati, Wilhelm H. Holzapfel, and Brian J. B. Wood, 145–64. Academic Press. https://doi.org/10.1016/B978-0-12-805060-6.00008-9.

Evans, Rachael, Anna M Alessi, Susannah Bird, Simon J McQueen-Mason, Neil C Bruce, and Michael A Brockhurst. 2017. "Defining the Functional Traits That Drive Bacterial Decomposer Community Productivity." *The ISME Journal* 11 (7): 1680–87. https://doi.org/10.1038/ismej.2017.22.

Gravel, Dominique, Thomas Bell, Claire Barbera, Thierry Bouvier, Thomas Pommier, Patrick Venail, and Nicolas Mouquet. 2011. "Experimental Niche Evolution Alters the Strength of the Diversity–Productivity Relationship." *Nature* 469 (7328): 89–92. https://doi.org/10.1038/nature09592.

Guerrero-Ramírez, Nathaly R., Dylan Craven, Peter B. Reich, John J. Ewel, Forest Isbell, Julia Koricheva, John A. Parrotta, et al. 2017. "Diversity-Dependent Temporal Divergence of Ecosystem Functioning in Experimental Ecosystems." *Nature Ecology & Evolution* 1 (11): 1639. https://doi.org/10.1038/s41559-017-0325-1.

Gurevich, Alexey, Vladislav Saveliev, Nikolay Vyahhi, and Glenn Tesler. 2013. "QUAST: Quality Assessment Tool for Genome Assemblies." *Bioinformatics* 29 (8): 1072–75. https://doi.org/10.1093/bioinformatics/btt086.

Hadfield, James, Nicholas J. Croucher, Richard J. Goater, Khalil Abudahab, David M. Aanensen, and Simon R. Harris. 2018. "Phandango: An Interactive Viewer for Bacterial Population Genomics." *Bioinformatics* 34 (2): 292–93. https://doi.org/10.1093/bioinformatics/btx610.

Hidalgo-Cantabrana, Claudio, Susana Delgado, Lorena Ruiz, Patricia Ruas-Madiedo, Borja Sánchez, and Abelardo Margolles. 2017. "Bifidobacteria and Their Health-Promoting Effects." *Microbiology Spectrum* 5 (3): 5.3.21. https://doi.org/10.1128/microbiolspec.BAD-0010-2016.

Hofmann, Rainer W., and M. Z. Zulfiqhar Jahufer. 2011. "Tradeoff between Biomass and Flavonoid Accumulation in White Clover Reflects Contrasting Plant Strategies." *PLOS ONE* 6 (4): e18949. https://doi.org/10.1371/journal.pone.0018949.

Hooper, David U., and Peter M. Vitousek. 1997. "The Effects of Plant Composition and Diversity on Ecosystem Processes." *Science* 277 (5330): 1302–5. https://doi.org/10.1126/science.277.5330.1302.

Hu, Lujun, Wenwei Lu, Linlin Wang, Mingluo Pan, Hao Zhang, Jianxin Zhao, and Wei Chen. 2017. "Assessment of Bifidobacterium Species Using GroEL Gene on the Basis of Illumina MiSeq High-Throughput Sequencing." *Genes* 8 (11). https://doi.org/10.3390/genes8110336.

Hyatt, Doug, Gwo-Liang Chen, Philip F LoCascio, Miriam L Land, Frank W Larimer, and Loren J Hauser. 2010. "Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification." *BMC Bioinformatics* 11 (March): 119. https://doi.org/10.1186/1471-2105-11-119.

Jain, Chirag, Luis M. Rodriguez-R, Adam M. Phillippy, Konstantinos T. Konstantinidis, and Srinivas Aluru. 2018. "High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries." *Nature Communications* 9 (1): 5114. https://doi.org/10.1038/s41467-018-07641-9.

Johnson, David R., Damian E. Helbling, Tae Kwon Lee, Joonhong Park, Kathrin Fenner, Hans-Peter E. Kohler, and Martin Ackermann. 2015. "Association of Biodiversity with the Rates of Micropollutant Biotransformations among Full-Scale Wastewater Treatment Plant Communities." *Appl. Environ. Microbiol.* 81 (2): 666–75. https://doi.org/10.1128/AEM.03286-14.

Kalliomäki, Marko, Maria Carmen Collado, Seppo Salminen, and Erika Isolauri. 2008. "Early Differences in Fecal Microbiota Composition in Children May Predict Overweight." *The*

American Journal of Clinical Nutrition 87 (3): 534–38. https://doi.org/10.1093/ajcn/87.3.534.

Kalliomäki, Marko, Pirkka Kirjavainen, Erkki Eerola, Pentti Kero, Seppo Salminen, and Erika Isolauri. 2001. "Distinct Patterns of Neonatal Gut Microflora in Infants in Whom Atopy Was and Was Not Developing." *Journal of Allergy and Clinical Immunology* 107 (1): 129–34. https://doi.org/10.1067/mai.2001.111237.

Katayama, Takane. 2016. "Host-Derived Glycans Serve as Selected Nutrients for the Gut Microbe: Human Milk Oligosaccharides and Bifidobacteria." *Bioscience, Biotechnology, and Biochemistry* 80 (4): 621–32. https://doi.org/10.1080/09168451.2015.1132153.

Kohyama, Takashi S., Matthew D. Potts, Tetsuo I. Kohyama, Kaoru Niiyama, Tze Leong Yao, Stuart J. Davies, and Douglas Sheil. 2020. "Trade-off between Standing Biomass and Productivity in Species-Rich Tropical Forest: Evidence, Explanations and Implications." *Journal of Ecology* 108 (6): 2571–83. https://doi.org/10.1111/1365-2745.13485.

Le Chatelier, Emmanuelle, Trine Nielsen, Junjie Qin, Edi Prifti, Falk Hildebrand, Gwen Falony, Mathieu Almeida, et al. 2013. "Richness of Human Gut Microbiome Correlates with Metabolic Markers." *Nature* 500 (7464): 541–46. https://doi.org/10.1038/nature12506.

Lee, Ju-Hoon, and Daniel J. O'Sullivan. 2010. "Genomic Insights into Bifidobacteria." *Microbiology and Molecular Biology Reviews : MMBR* 74 (3): 378–416. https://doi.org/10.1128/MMBR.00004-10.

Letunic, Ivica, and Peer Bork. 2016. "Interactive Tree of Life (ITOL) v3: An Online Tool for the Display and Annotation of Phylogenetic and Other Trees." *Nucleic Acids Research* 44 (Web Server issue): W242–45. https://doi.org/10.1093/nar/gkw290.

Liwen, Zhang, Wan Yu, Ma Liang, Xu Kaihong, and Cheng Baojin. 2018. "A Low Abundance of Bifidobacterium but Not Lactobacillius in the Feces of Chinese Children with Wheezing Diseases." *Medicine* 97 (40). https://doi.org/10.1097/MD.0000000000012745.

Lloyd-Price, Jason, Galeb Abu-Ali, and Curtis Huttenhower. 2016. "The Healthy Human Microbiome." *Genome Medicine* 8 (April). https://doi.org/10.1186/s13073-016-0307-y.

Milani, Christian, Gabriele Andrea Lugli, Sabrina Duranti, Francesca Turroni, Francesca Bottacini, Marta Mangifesta, Borja Sanchez, et al. 2014. "Genomic Encyclopedia of Type Strains of the Genus Bifidobacterium." *Applied and Environmental Microbiology* 80 (20): 6290–6302. https://doi.org/10.1128/AEM.02308-14.

Milani, Christian, Gabriele Andrea Lugli, Sabrina Duranti, Francesca Turroni, Leonardo Mancabelli, Chiara Ferrario, Marta Mangifesta, et al. 2015. "Bifidobacteria Exhibit Social Behavior through Carbohydrate Resource Sharing in the Gut." *Scientific Reports* 5 (October): 15782. https://doi.org/10.1038/srep15782.

Milani, Christian, Leonardo Mancabelli, Gabriele Andrea Lugli, Sabrina Duranti, Francesca Turroni, Chiara Ferrario, Marta Mangifesta, et al. 2015. "Exploring Vertical Transmission of Bifidobacteria from Mother to Child." *Appl. Environ. Microbiol.* 81 (20): 7078–87. https://doi.org/10.1128/AEM.02037-15.

Munoz, Jose, Kieran James, Francesca Bottacini, and Douwe Van Sinderen. 2020. "Biochemical Analysis of Cross-Feeding Behaviour between Two Common Gut Commensals When Cultivated on Plant-Derived Arabinogalactan." *Microbial Biotechnology* 13 (6): 1733–47. https://doi.org/10.1111/1751-7915.13577.

Nayfach, Stephen, Beltran Rodriguez-Mueller, Nandita Garud, and Katherine S. Pollard. 2016. "An Integrated Metagenomics Pipeline for Strain Profiling Reveals Novel Patterns of Bacterial Transmission and Biogeography." *Genome Research* 26 (11): 1612–25. https://doi.org/10.1101/gr.201863.115.

Nurk, Sergey, Dmitry Meleshko, Anton Korobeynikov, and Pavel A. Pevzner. 2017. "MetaSPAdes: A New Versatile Metagenomic Assembler." *Genome Research* 27 (5): 824–34. https://doi.org/10.1101/gr.213959.116.

Oliver, Andrew, Alexander B. Chase, Claudia Weihe, Stephanie B. Orchanian, Stefan F. Riedel, Clark L. Hendrickson, Mi Lay, Julia Massimelli Sewall, Jennifer B. H. Martiny, and Katrine Whiteson. 2021. "High-Fiber, Whole-Food Dietary Intervention Alters the Human Gut Microbiome but Not Fecal Short-Chain Fatty Acids." *MSystems* 6 (2): e00115-21. https://doi.org/10.1128/mSystems.00115-21.

O'Neill, Ian, Zoe Schofield, and Lindsay J. Hall. 2017. "Exploring the Role of the Microbiota Member Bifidobacterium in Modulating Immune-Linked Diseases." *Emerging Topics in Life Sciences* 1 (4): 333–49. https://doi.org/10.1042/ETLS20170058.

Page, Andrew J., Carla A. Cummins, Martin Hunt, Vanessa K. Wong, Sandra Reuter, Matthew T. G. Holden, Maria Fookes, Daniel Falush, Jacqueline A. Keane, and Julian Parkhill. 2015. "Roary: Rapid Large-Scale Prokaryote Pan Genome Analysis." *Bioinformatics* 31 (22): 3691–93. https://doi.org/10.1093/bioinformatics/btv421.

Palframan, Richard J., Glenn R. Gibson, and Robert A. Rastall. 2003. "Carbohydrate Preferences of Bifidobacterium Species Isolated from the Human Gut." *Current Issues in Intestinal Microbiology* 4 (2): 71–75.

Pokusaeva, Karina, Gerald F. Fitzgerald, and Douwe van Sinderen. 2011. "Carbohydrate Metabolism in Bifidobacteria." *Genes & Nutrition* 6 (3): 285–306. https://doi.org/10.1007/s12263-010-0206-6.

Read, N. W., C. A. Miles, D. Fisher, A. M. Holgate, N. D. Kime, M. A. Mitchell, A. M. Reeve, T. B. Roche, and M. Walker. 1980. "Transit of a Meal through the Stomach, Small Intestine, and Colon in Normal Subjects and Its Role in the Pathogenesis of Diarrhea." *Gastroenterology* 79 (6): 1276–82. https://doi.org/10.1016/0016-5085(80)90925-7.

Reese, Aspen T., and Robert R. Dunn. 2018. "Drivers of Microbiome Biodiversity: A Review of General Rules, Feces, and Ignorance." *MBio* 9 (4): e01294-18. https://doi.org/10.1128/mBio.01294-18.

Schnorr, Stephanie L., Marco Candela, Simone Rampelli, Manuela Centanni, Clarissa Consolandi, Giulia Basaglia, Silvia Turroni, et al. 2014. "Gut Microbiome of the Hadza Hunter-Gatherers." *Nature Communications* 5 (April). https://doi.org/10.1038/ncomms4654.

Seemann, Torsten. 2014. "Prokka: Rapid Prokaryotic Genome Annotation." *Bioinformatics (Oxford, England)* 30 (14): 2068–69. https://doi.org/10.1093/bioinformatics/btu153.

Shen, Wei, Shuai Le, Yan Li, and Fuquan Hu. 2016. "SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation." *PLOS ONE* 11 (10): e0163962. https://doi.org/10.1371/journal.pone.0163962.

Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2015. "BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs." *Bioinformatics* 31 (19): 3210–12. https://doi.org/10.1093/bioinformatics/btv351.

Tasnim, Nishat, Nijiati Abulizi, Jason Pither, Miranda M. Hart, and Deanna L. Gibson. 2017. "Linking the Gut Microbial Ecosystem with the Environment: Does Gut Health Depend on Where We Live?" *Frontiers in Microbiology* 8 (October). https://doi.org/10.3389/fmicb.2017.01935.

Tilman, D. 2001. "Diversity and Productivity in a Long-Term Grassland Experiment." *Science* 294 (5543): 843–45. https://doi.org/10.1126/science.1060391.

Tilman, David, Forest Isbell, and Jane M. Cowles. 2014. "Biodiversity and Ecosystem Functioning." *Annual Review of Ecology, Evolution, and Systematics* 45 (1): 471–93. https://doi.org/10.1146/annurev-ecolsys-120213-091917.

Tilman, David, Peter B. Reich, and Johannes M. H. Knops. 2006. "Biodiversity and Ecosystem Stability in a Decade-Long Grassland Experiment." *Nature* 441 (7093): 629. https://doi.org/10.1038/nature04742.

Turroni, Francesca, Elena Foroni, Paola Pizzetti, Vanessa Giubellini, Angela Ribbera, Paolo Merusi, Patrizio Cagnasso, et al. 2009. "Exploring the Diversity of the Bifidobacterial Population in the Human Intestinal Tract." *Applied and Environmental Microbiology* 75 (6): 1534–45. https://doi.org/10.1128/AEM.02216-08.

Turroni, Francesca, Christian Milani, Sabrina Duranti, Jennifer Mahony, Douwe van Sinderen, and Marco Ventura. 2018. "Glycan Utilization and Cross-Feeding Activities by Bifidobacteria." *Trends in Microbiology* 26 (4): 339–50. https://doi.org/10.1016/j.tim.2017.10.001.

Turroni, Francesca, Christian Milani, Sabrina Duranti, Leonardo Mancabelli, Marta Mangifesta, Alice Viappiani, Gabriele Andrea Lugli, et al. 2016. "Deciphering Bifidobacterial-Mediated Metabolic Interactions and Their Impact on Gut Microbiota by a Multi-Omics Approach." *The ISME Journal* 10 (7): 1656–68. https://doi.org/10.1038/ismej.2015.236.

Turroni, Francesca, Christian Milani, Douwe Van Sinderen, and Marco Ventura. 2018. "Chapter 12 - Bifidobacteria: Ecology and Coevolution With the Host." In *The Bifidobacteria and Related Organisms*, edited by Paola Mattarelli, Bruno Biavati, Wilhelm H. Holzapfel, and Brian J. B. Wood, 213–20. Academic Press. https://doi.org/10.1016/B978-0-12-805060-6.00012-0.

Turroni, Francesca, Ezgi Özcan, Christian Milani, Leonardo Mancabelli, Alice Viappiani, Douwe van Sinderen, David Sela, and Marco Ventura. 2015. "Glycan Cross-Feeding Activities between Bifidobacteria under in Vitro Conditions." *Frontiers in Microbiology* 6. https://doi.org/10.3389/fmicb.2015.01030.

Turroni, Francesca, Clelia Peano, Daniel A. Pass, Elena Foroni, Marco Severgnini, Marcus J. Claesson, Colm Kerr, et al. 2012. "Diversity of Bifidobacteria within the Infant Gut Microbiota." *PloS One* 7 (5): e36957. https://doi.org/10.1371/journal.pone.0036957.

Turroni, Francesca, Douwe van Sinderen, and Marco Ventura. 2011. "Genomics and Ecological Overview of the Genus Bifidobacterium." *International Journal of Food Microbiology*, 3rd International Symposium on Propionibacteria and Bifidobacteria: Dairy and Probiotic applications, Oviedo 1-4 June 2010, 149 (1): 37–44. https://doi.org/10.1016/j.ijfoodmicro.2010.12.010.

Venema, Koen, and Pieter van den Abbeele. 2013. "Experimental Models of the Gut Microbiome." *Best Practice & Research Clinical Gastroenterology*, The Gut Microbiome, 27 (1): 115–26. https://doi.org/10.1016/j.bpg.2013.03.002.

Weihe, Claudia, Claudia, and Julio Avelar-Barragan. 2021. "Next Generation Shotgun Library Preparation for Illumina Sequencing - Low Volume." Protocols.Io. July 2, 2021. https://www.protocols.io/view/next-generation-shotgun-library-preparation-for-il-bvv8n69w.

Yatsunenko, Tanya, Federico E. Rey, Mark J. Manary, Indi Trehan, Maria Gloria Dominguez-Bello, Monica Contreras, Magda Magris, et al. 2012. "Human Gut Microbiome Viewed across Age and Geography." *Nature* 486 (7402): 222–27. https://doi.org/10.1038/nature11053.

Zhang, Han, Tanner Yohe, Le Huang, Sarah Entwistle, Peizhi Wu, Zhenglu Yang, Peter K Busk, Ying Xu, and Yanbin Yin. 2018. "DbCAN2: A Meta Server for Automated Carbohydrate-Active Enzyme Annotation." *Nucleic Acids Research* 46 (Web Server issue): W95–101. https://doi.org/10.1093/nar/gky418.

FIGURES

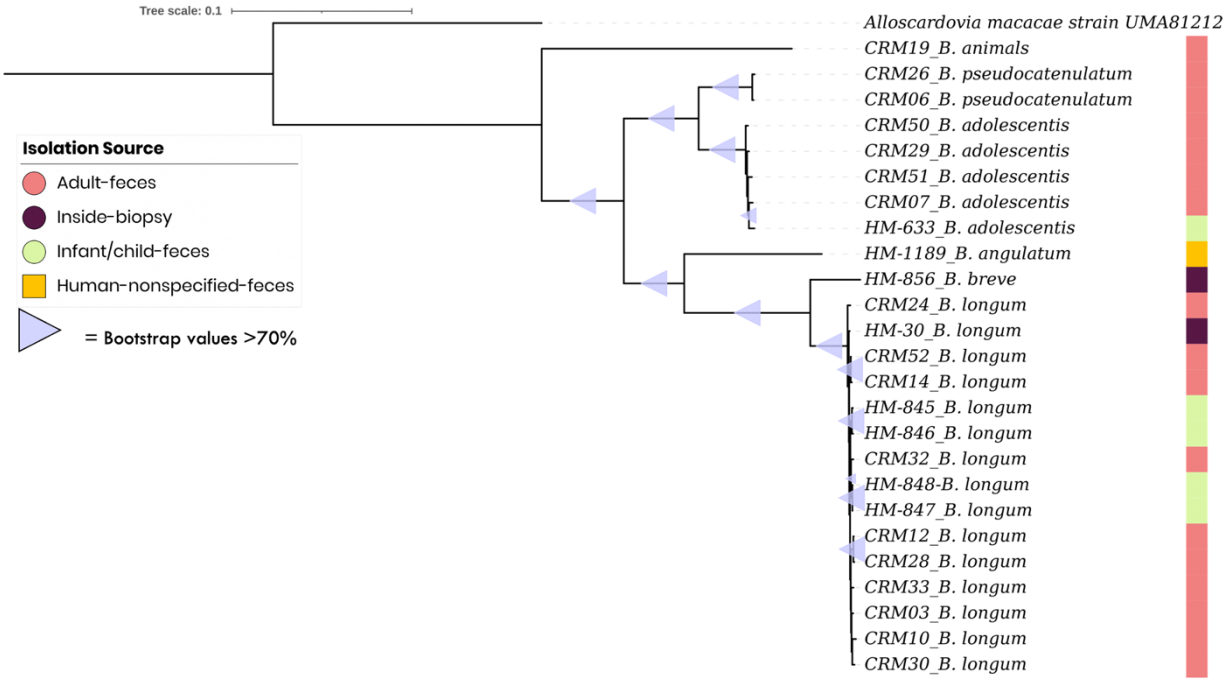


Figure 3.1: Phylogenetic tree of 25 gut-isolated bifidobacteria. The phylogenetic tree constructed with 107 concatenated core genes includes bifidobacteria isolates belonging to the following species: *B. longum* (n=15), *B. adolescentis* (n=5), *B. pseudocatenulatum* (n=2), *B. animalis* (n=1), *B. angulatum* (n=1), and *B. breve* (n=1). A strain from the *Alloscardovia* genus was used as an outgroup. Boostrap values higher than 70% are represented with purple triangles.

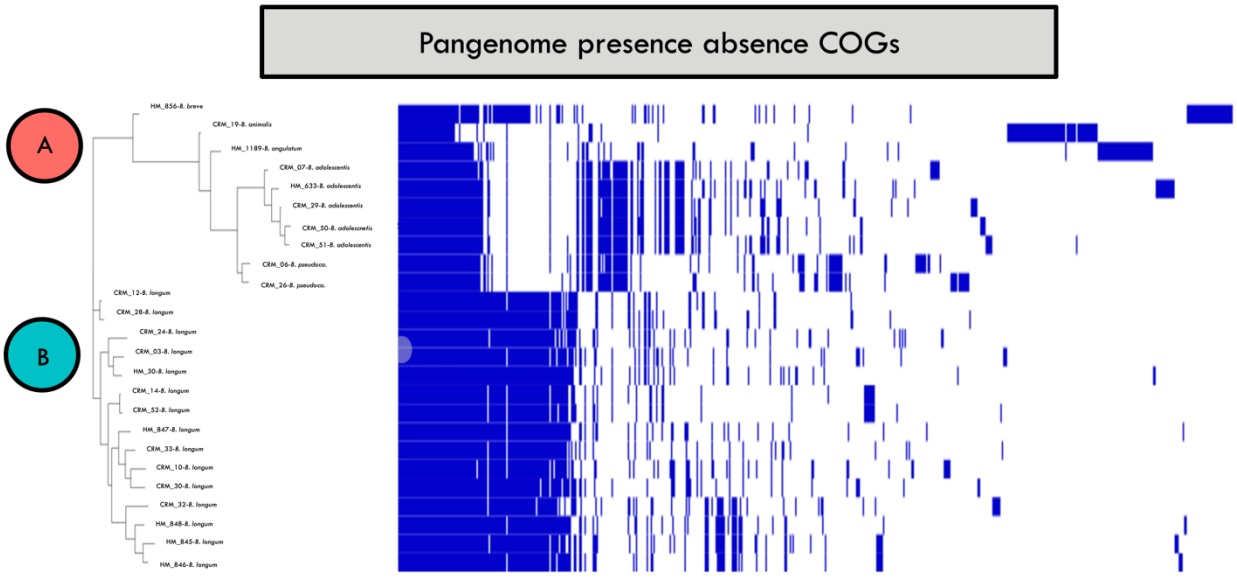


Figure 3.2: Pangemome analysis of 25 gut-isolated bifidobacteria. The pangenome is depicted as a heatmap showing gene presence (royal blue) or absence (white) in each of the isolates obtained from Roary 3.11.2. There was a total of 8,073 gene clusters of orthologous genes (COGs) from which 550 were present in all isolates. A pattern of two functional groups can be observed and is highlighted by letters A and B.

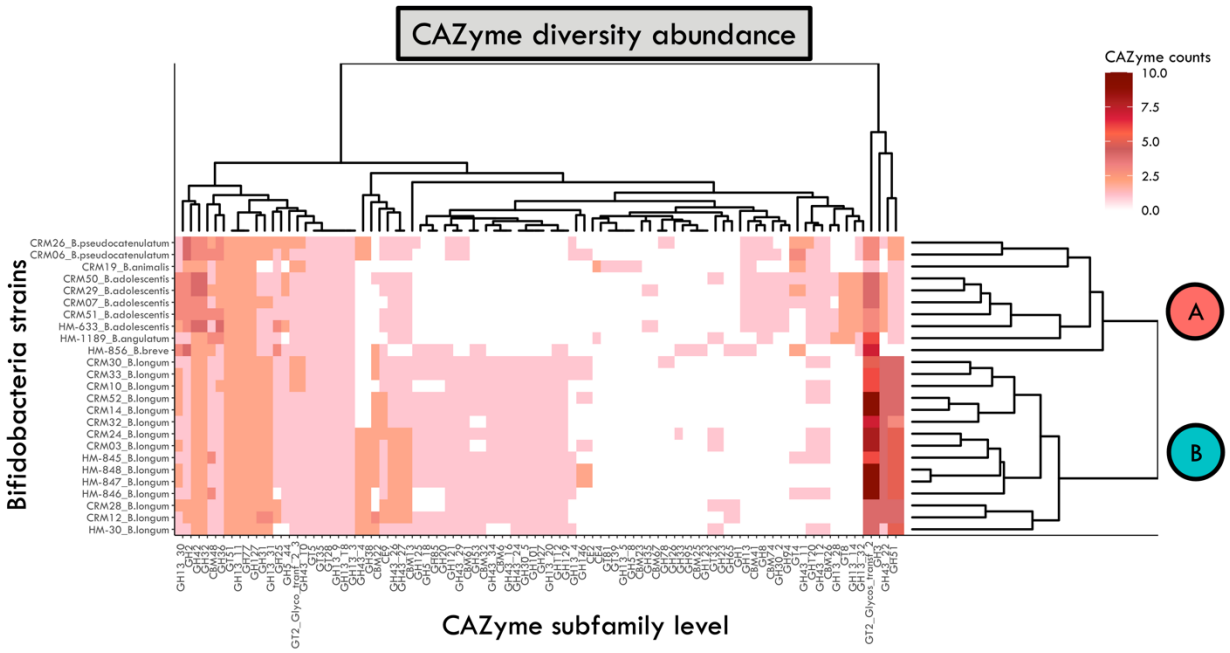


Figure 3.3: Heatmap of CAZymes present in 25 gut-isolates bifidobacteria. The number of carbohydrate-degrading genes per isolate was obtained through DbCAN2 and their gene counts were quantified; these counts are summarized by the color intensity chart. A pattern of two functional groups can be observed and is highlighted by letters A and B, which correspond to the same groupings formed by the pangenome analysis.

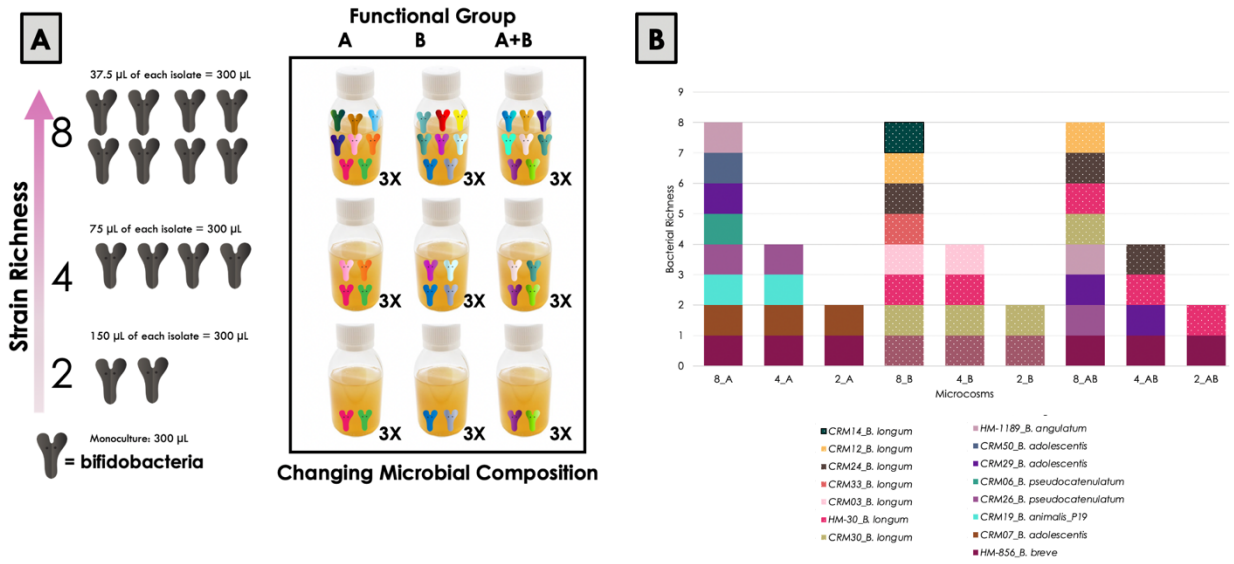


Figure 3.4: Experimental design for microcosm experiment. We constructed microcosms varying strain richness (2, 4, and 8) and functional composition (Functional groups A, AB, and B) with a nested design. We grew monocultures of each of the isolates included in the experiment plus media controls, for a total of 78 vials as each had 3 biological replicates. Bacterial cultures were standardized to 0.1 optical density units and added 300 μL of total standardized bacteria to the microcosm treatments and the monocultures in 30 mL of rich MRS media with 1.5% lentil flour.

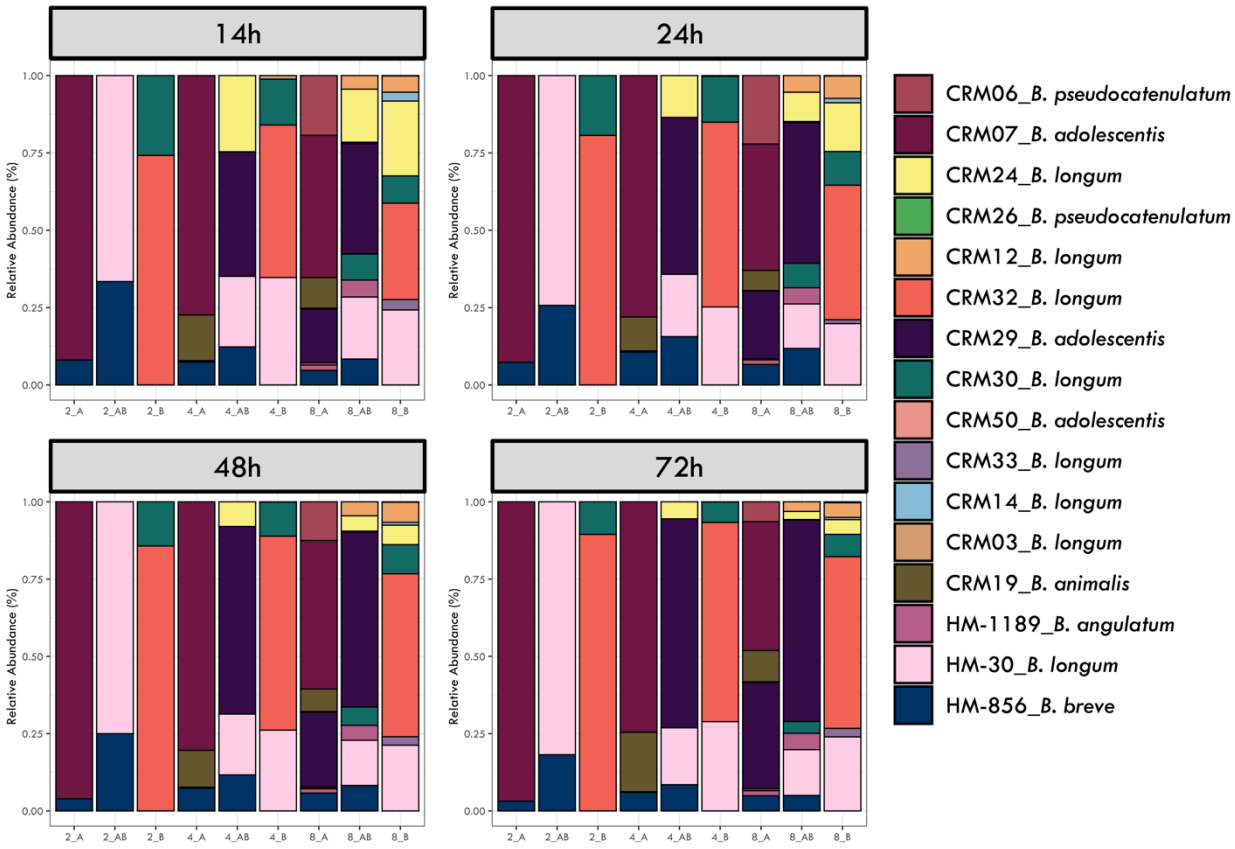


Figure 3.5: Taxa bar plots of co-cultures across time. Bacterial community composition was obtained using MIDAS. Each column represents the average community composition of the 3 biological replicates per microcosm. The columns are sorted by richness levels starting with the lowest level (2). Functional groups are displayed by naming the columns as "richness number" underscore "Functional group."

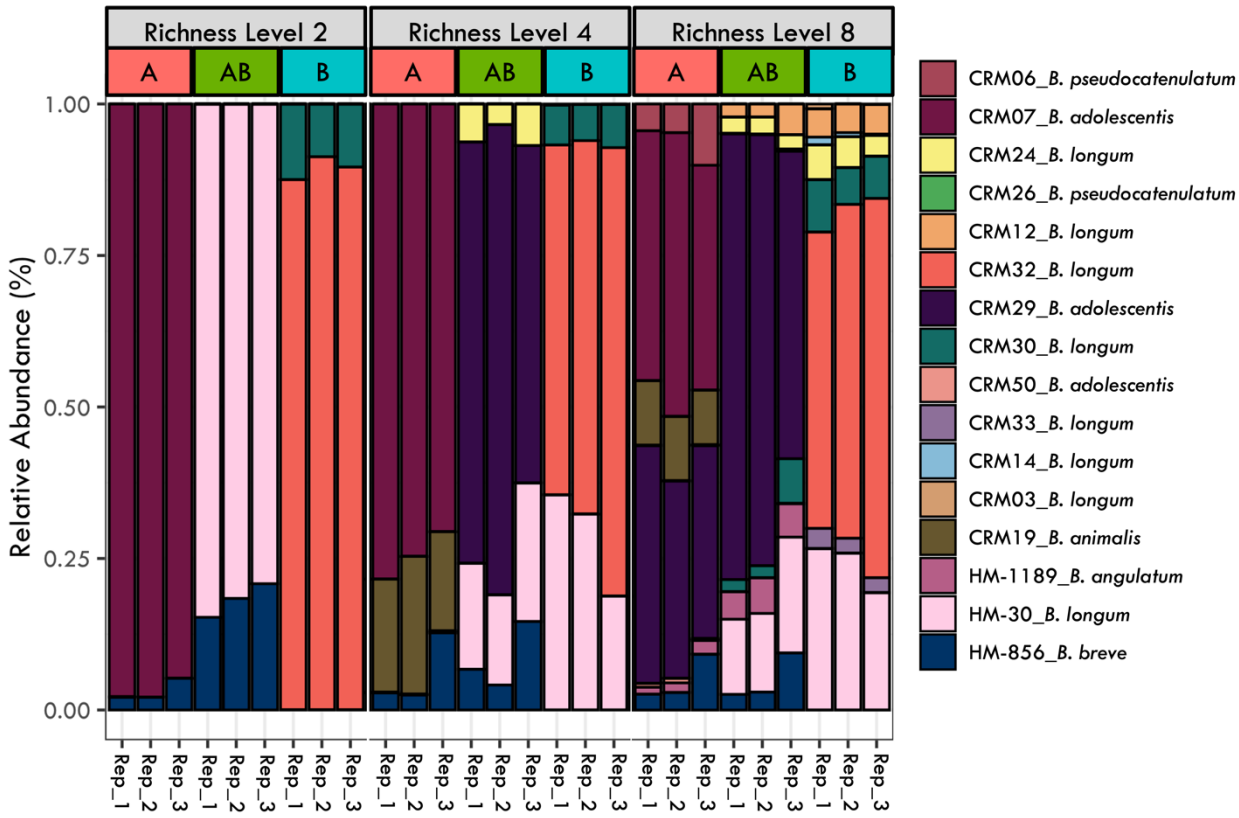


Fig. 3.6: Taxa bar plots of co-cultures at 72h. Bacterial community composition was obtained using MIDAS. Each column represents the community composition for each replicate per microcosm (Replicate 1, 2, and 3). The columns are sorted by richness levels starting with the lowest level (2). Functional groups are marked across richness levels by orange (A), green (AB), and blue (B).

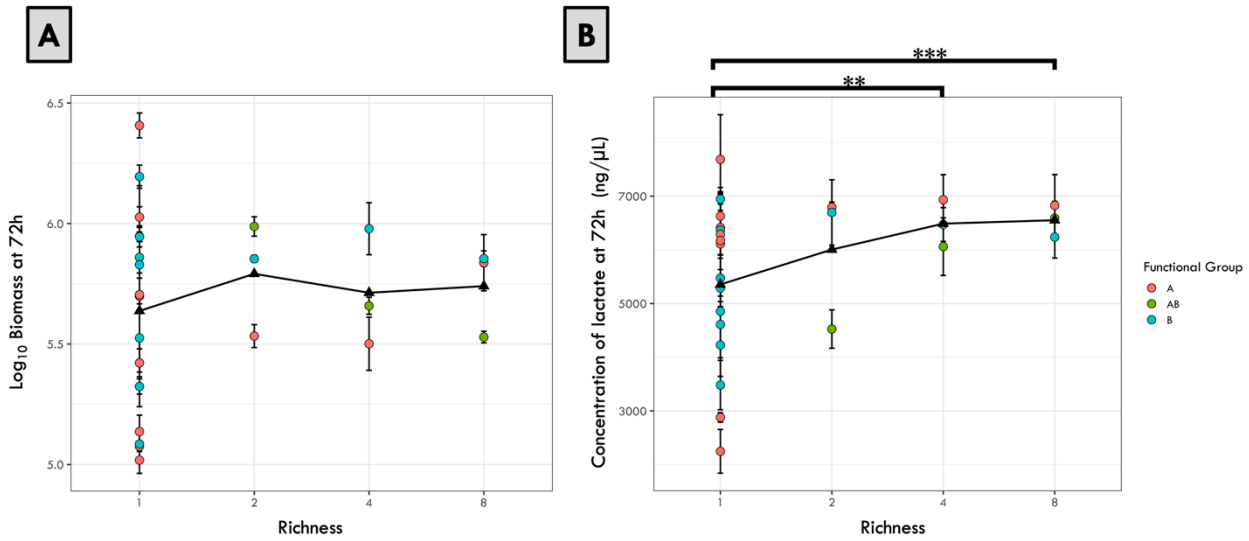


Fig. 3.7: Ecosystem functioning across bifidobacteria richness levels at 72h. Panel A depicts the $\log_{10}$ cell count data and panel B shows the lactate concentration at the last timepoint (72h) across richness levels. Each dot represents the averaging of replicates by functioning metric (biomass or lactate) colored by Functional groups per richness level; their standard error bars are displayed. The horizontal line across is connecting the means (black triangles) at each richness level. No significant differences were found for biomass data across richness groups. On the other hand we found a statistically significant difference across richness levels for lactate concentration (Welch ANOVA, $F(3,22.42124) = 6.74$, $p = 0.002$, $n = 75$). The bars and asterisks across the lactate plot show the Games Howell Post-hoc significant test. Two asterisk represents a p-value less than 0.05; three asterisks represent a p-value less than 0.001

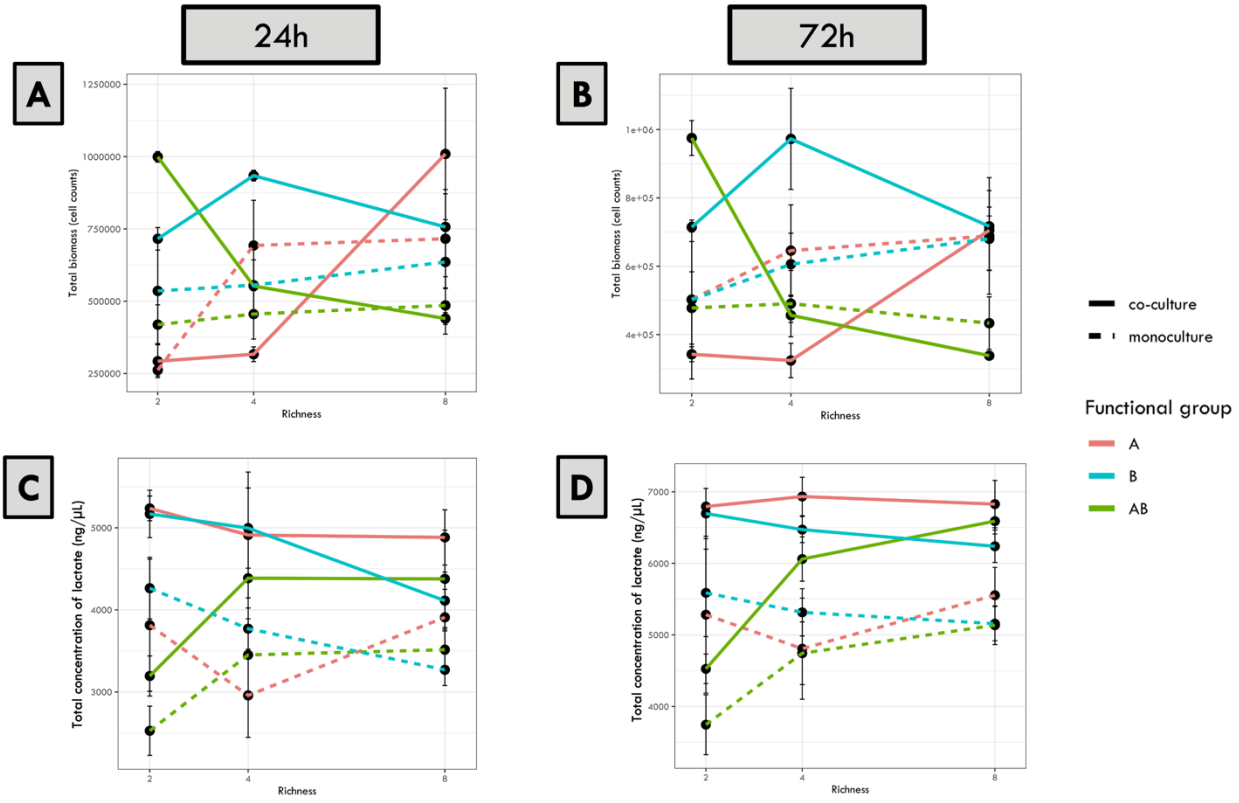


Figure 3.8: Total ecosystem functioning of co-cultures versus their corresponding monocultures. Panels A and B correspond to total biomass data at 24h and 72h, respectively. Panels C and D correspond to total lactate concentration at 24h and 72h, respectively. Each dot represents the average across replicates per functioning metric (biomass or lactate) and the mean standard error bars are displayed. Solid lines are connecting the co-culture average data points across richness levels and the dashed lines are connecting the average values of all monocultures that were included in that microcosm at that richness level. Lines are colored according to Functional group: orange (A), green (AB), and blue (B).

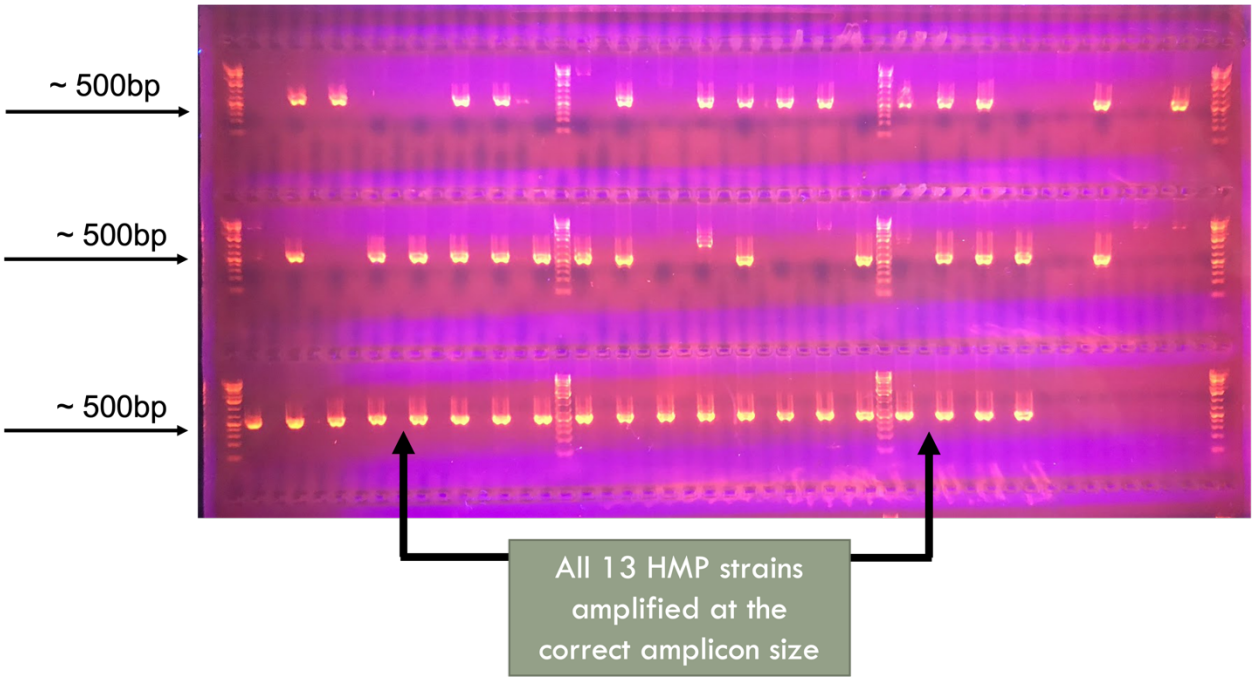


Figure S3.1: *groEL* amplification. A total of 52 potential bifidobacteria isolates were screened through PCR amplification of the *groEL* gene alongside the 13 bifidobacteria strains from the Human Microbiome Project.

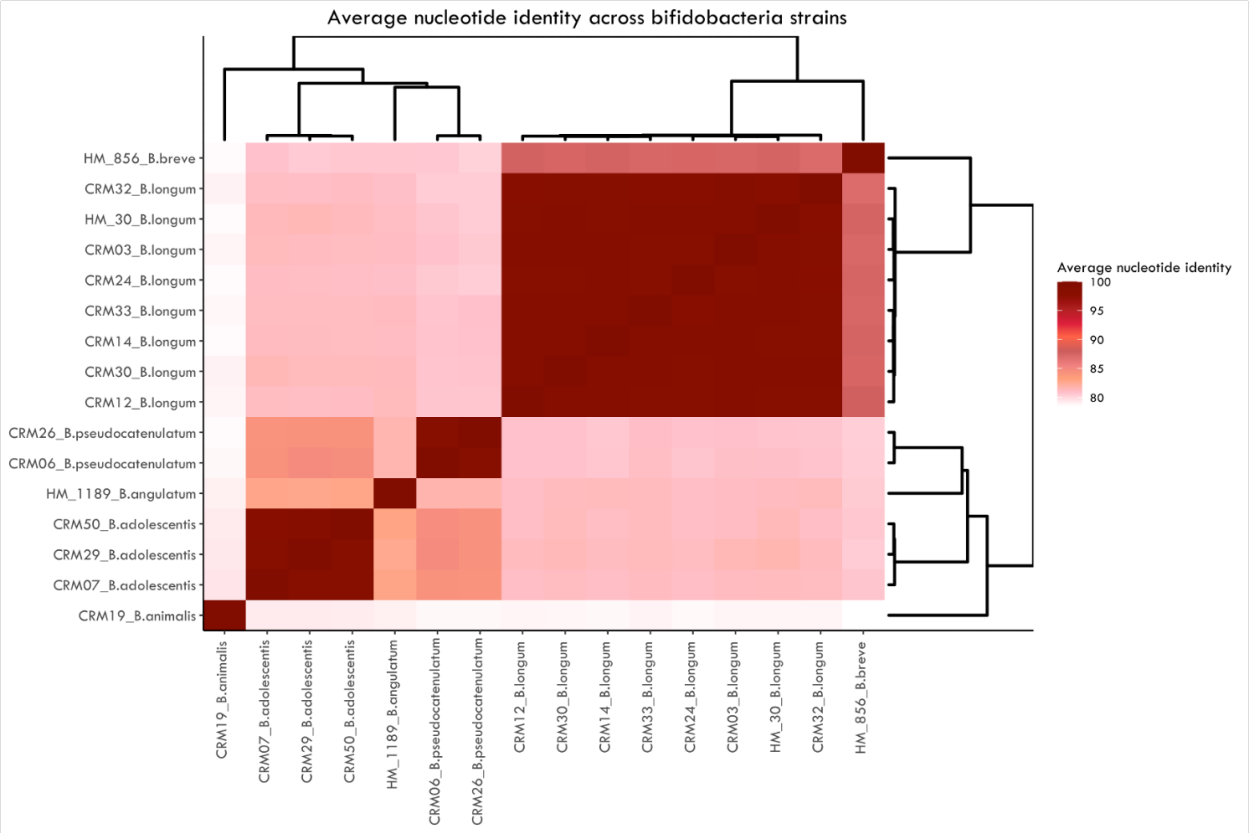


Figure S3.2: Average nucleotide identity across the 16 bifidobacteria strains selected for experiment. Average Nucleotide Identity (ANI) values obtained through FastANI are displayed in a heatmap format. The ANI values across strains range from 80-98.6%.

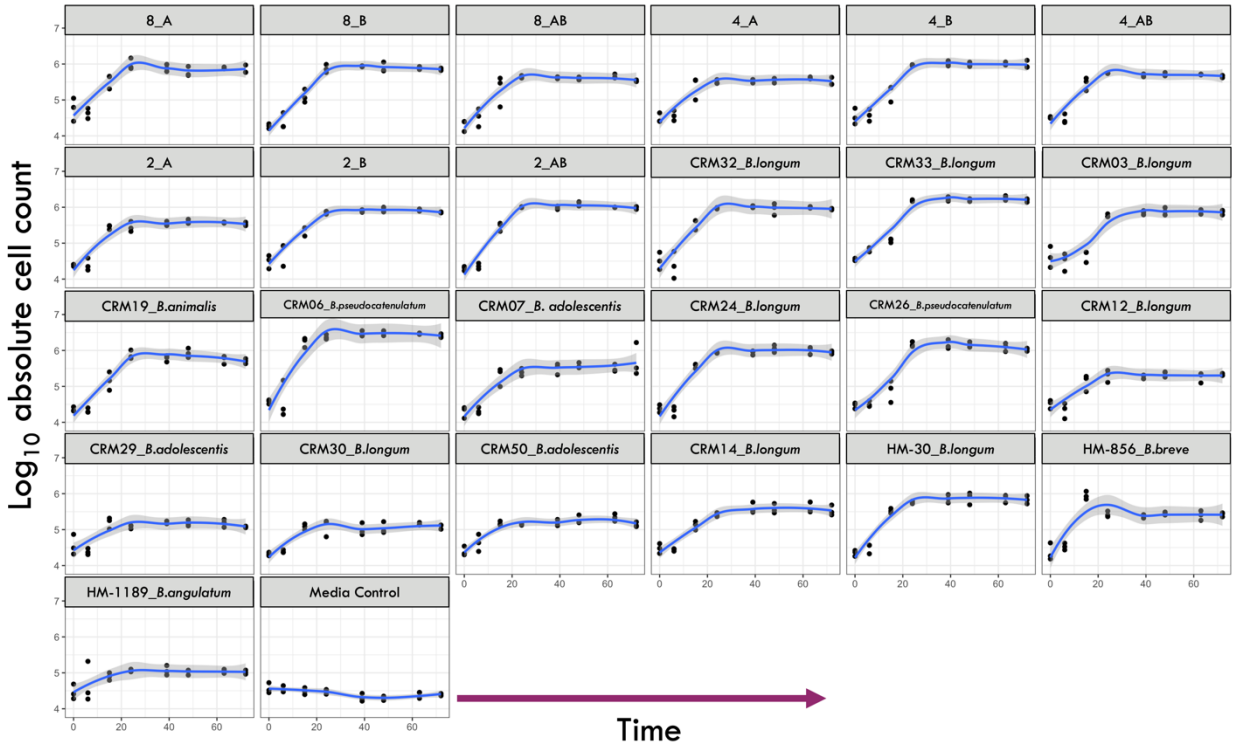


Figure S3.3: Growth curves for all microcosms. Cell counts were $\log_{10}$ transformed for better visualization.

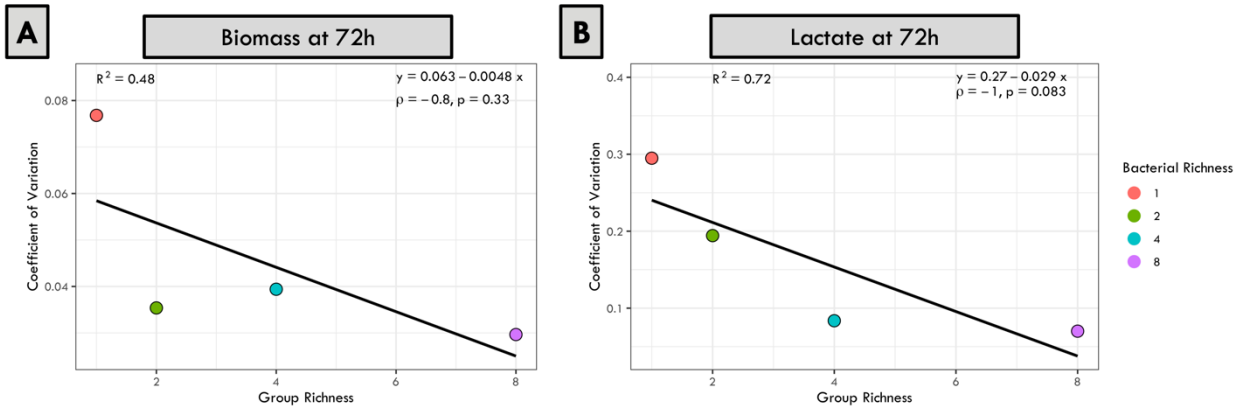


Figure S3.4: Coefficient of variation across richness levels for biomass and lactate functioning metrics. Each dot represents the Coefficient of Variation (CV) value per richness level for biomass (Panel A) and lactate (Panel B) calculated in base R version 4.0.2. A Spearman's correlation analysis was performed for both metrics to compare the relationship between richness and CV.

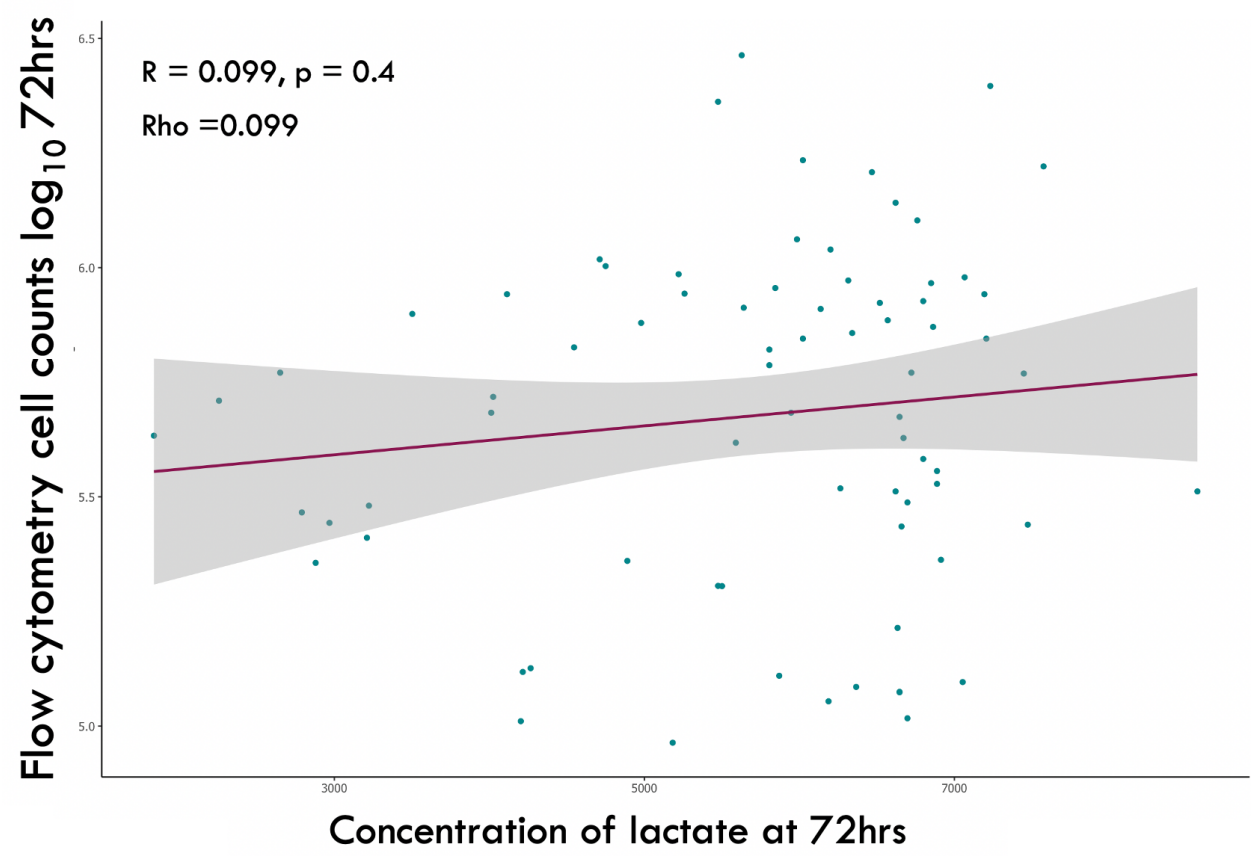


Figure S3.5: Spearman's correlation between lactate concentration and biomass at 72h. The $\log_{10}$ transformed data for cell counts is plotted against the concentration of lactic acid at 72h. The results from the Spearman's correlation test are displayed on the top right corner.

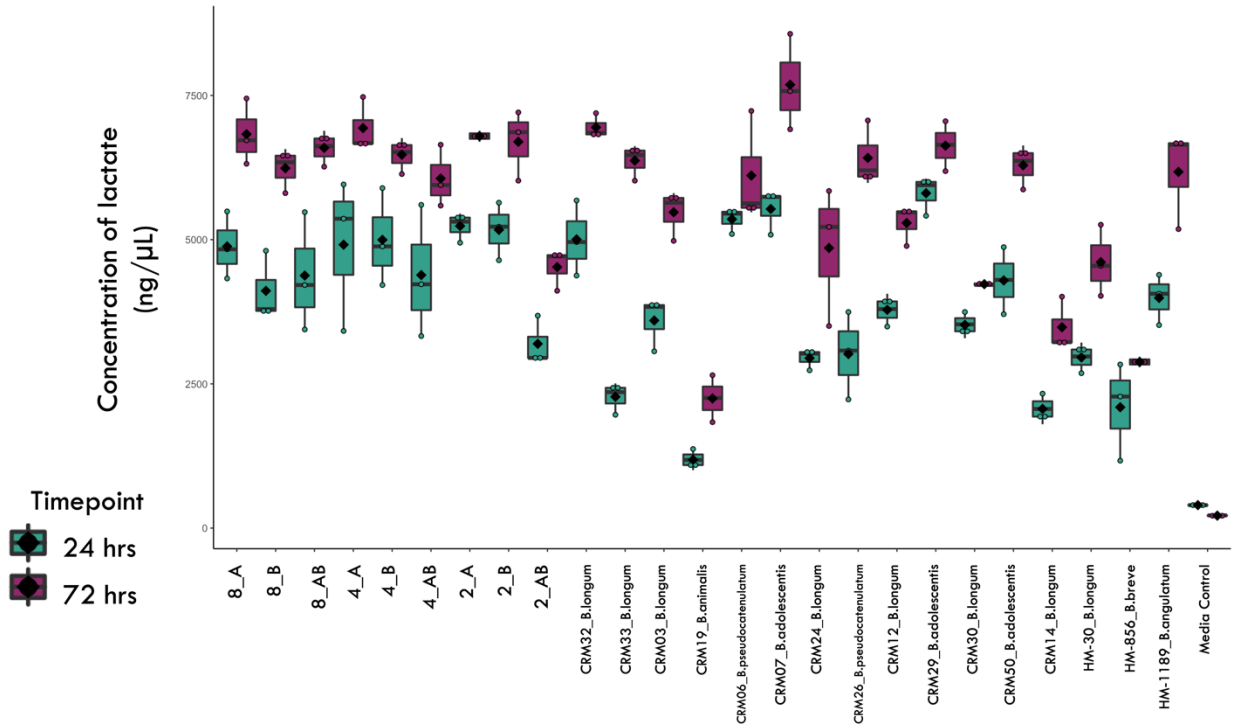


Figure S3.6: Concentration of lactate at the 24h and 72h timepoints. Boxplots displaying the concentration of lactate between two timepoints: 24h (green) and 72h (magenta) for all microcosms including co-cultures and monocultures.

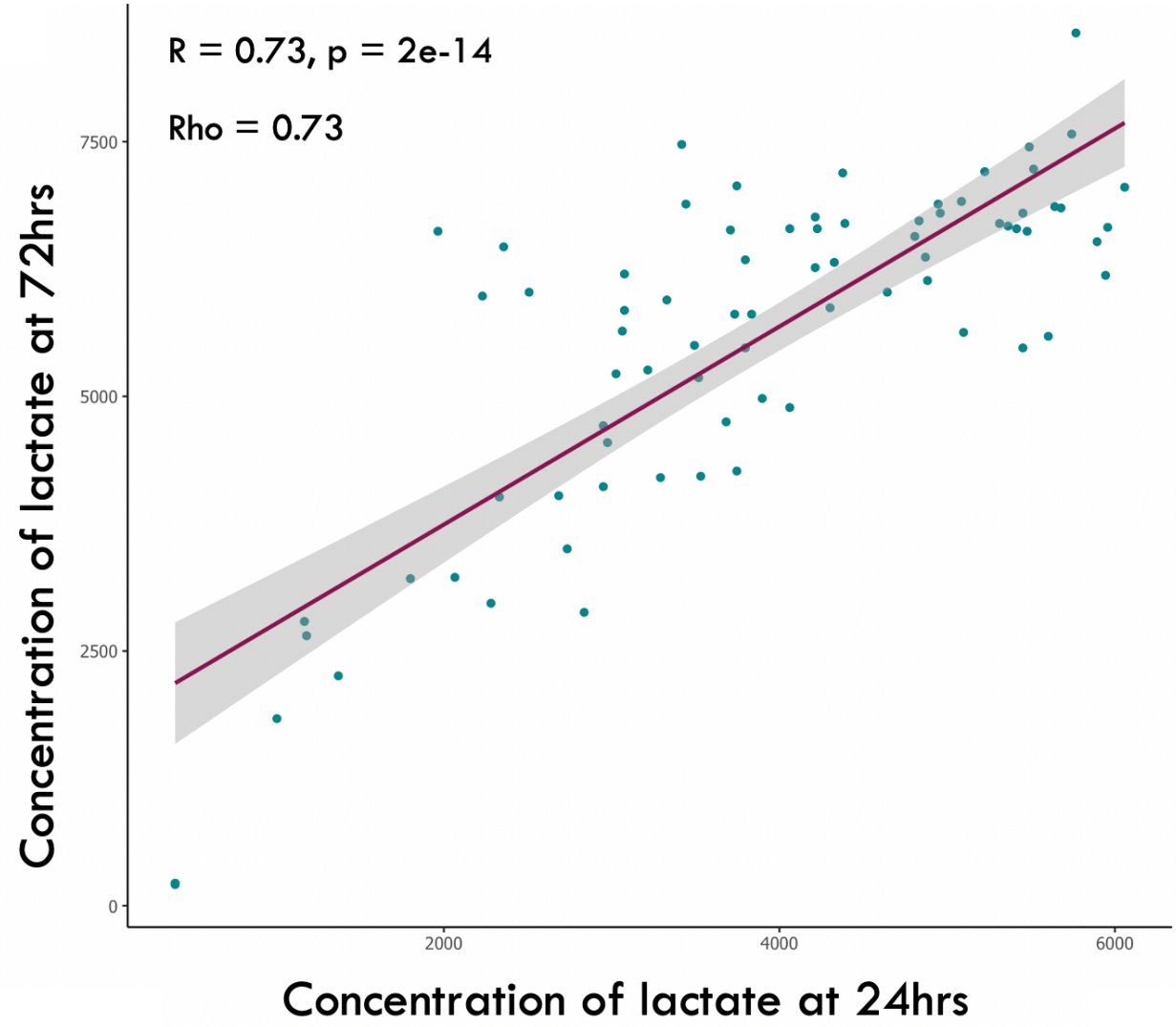


Figure S3.7: Spearman's correlation between lactate concentration at 24h and 72h. The concentration of lactic acid at 24h is plotted against the concentration of lactate at 72h. The results from the Spearman's correlation test are displayed on the top right corner.

Isolate Name	Species designation	Functional Group Designation	Isolation Source	No. contigs (>= 0 bp)	No. contigs (>= 1000 bp)	Largest contig (bp)	Total length (>= 0 bp)	Total length (>= 1000 bp)	N50	L50	GC (%)
CRM03	*B. longum*	B	fecal - adult UCI	86	47	310361	2448920	2438916	110024	7	60.26
CRM06	*B. pseudocatenulatum*	A	fecal - adult UCI	92	28	467374	2240016	2226421	137906	5	56.41
CRM07	*B. adolescentis*	A	fecal - adult UCI	15	10	808543	2126450	2125625	780879	2	59.43
CRM10	*B. longum*	B	fecal - adult UCI	106	57	195107	2485402	2477332	87506	9	60.26
CRM12	*B. longum*	B	fecal - adult UCI	133	41	361827	2368477	2349291	107714	7	60.33
CRM14	*B. longum*	B	fecal - adult UCI	99	43	256606	2376402	2364663	175862	6	60.05
CRM19	*B. animalis*	A	fecal - adult UCI	28	17	490360	1919137	1917772	201006	3	60.46
CRM24	*B. longum*	B	fecal - adult UCI	82	47	170378	2382856	2376163	102811	9	60.02
CRM26	*B. pseudocatenulatum*	A	fecal - adult UCI	53	14	445461	2225460	2220269	293705	3	56.42
CRM28	*B. longum*	B	fecal - adult UCI	164	49	327176	2387249	2364739	98678	8	60.17
CRM29	*B. adolescentis*	A	fecal - adult UCI	35	13	1001670	2266253	2262442	747494	2	59.52
CRM30	*B. longum*	B	fecal - adult UCI	124	54	339395	2443390	2426518	137731	6	60.23
CRM32	*B. longum*	B	fecal - adult UCI	127	60	267827	2539826	2527925	77040	10	60.07
CRM33	*B. longum*	B	fecal - adult UCI	118	52	367180	2441045	2430819	123918	6	60.28
CRM50	*B. adolescentis*	A	fecal - adult UCI	38	22	677149	2197691	2195458	474550	2	59.45
CRM51	*B. adolescentis*	A	fecal - adult UCI	66	20	760514	2286996	2276232	437354	2	59.49
CRM52	*B. longum*	B	fecal - adult UCI	113	44	273292	2399457	2384424	120530	6	60.03
HM-30	*B. longum*	B	sigmoid colon biopsy - adult HM	141	68	154415	2350006	2332726	59281	14	60.01
HM-411	*B. breve*	A	mid-vaginal wall HM	59	23	541677	2409695	2401717	222688	4	58.77
HM-412	*B. breve*	A	mid-vaginal wall HM	73	26	366161	2411051	2402405	191147	5	58.77
HM-633	*B. adolescentis*	A	fecal - infant HM	46	28	406175	2292230	2289138	374835	3	59.23
HM-845	*B. longum*	B	fecal - infant HM	105	44	272319	2566205	2552558	122337	8	59.7
HM-846	*B. longum*	B	fecal - infant HM	107	43	368987	2600368	2587841	117584	7	59.83
HM-847	*B. longum*	B	fecal - infant HM	140	37	261964	2479741	2456429	166087	6	60.12
HM-848	*B. longum*	B	fecal - infant HM	160	39	279338	2614126	2587324	145902	7	59.7
HM-856	*B. breve*	A	biopsy of ileo-anal pouch mucosa HM	48	23	596329	2474426	2469365	305067	3	59.04
HM-868	*B. sp.*	A	dental plaque HM	50	30	416888	2571887	2566490	181015	5	58.39
HM-1120	*B. breve*	A	vaginal swab-bacterial vaginosis HM	54	24	541327	2376390	2370292	366167	3	58.66
HM-1189	*B. angulatum*	A	fecal - NA HM	23	14	314423	2038367	2037598	248920	4	59.41
HM-1299	*B. longum*	B	vaginal swab from pregnant women HM	119	33	345522	2385999	2370117	197272	5	60.1
Average				**86.8**	**35**	**412858.17**	**2370183.93**	**2359633.63**	**230433.67**	**5.57**	**59.5**

Table S3.1 Pool of bifidobacteria isolates. Bacteria were isolated from fecal samples collected at UCI (CRM) and others were purchased to BEI from the Human Microbiome Project (HM).