Characterizing the Role of Sialylated Milk Glycans and the Infant Gut Microbiota in Growth and Metabolism

Mark Richard Charbonneau
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Recommended Citation
https://openscholarship.wustl.edu/art_sci_etds/642

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
Characterizing the Role of Sialylated Milk Glycans and the Infant Gut Microbiota in Growth and Metabolism

by

Mark Charbonneau

A dissertation presented to the Graduate School of Arts and Sciences of Washington University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

December 2015

St. Louis, Missouri
# Table of Contents

Acknowledgements .......................................................................................................................... viii  
ABSTRACT OF THE DISSERTATION ..................................................................................... xi  

## Chapter 1

### Introduction

Abstract ............................................................................................................................................. 3  
Introduction ...................................................................................................................................... 3  
Defining Human Postnatal Development from a Microbial Perspective ...................................... 4  
  Undernutrition and Gut Microbiota Immaturity ............................................................................. 6  
  Establishing Microbiota and the Maternal Influence .................................................................. 13  
The Impact of First Foods .............................................................................................................. 14  
  Breast Milk .................................................................................................................................. 14  
  Serial Introduction of Complementary Foods in Ways that Promote Maturation of the Gut Microbiota ......................................................................................................................... 17  
Additional Considerations Regarding the Developmental Biology of the Gut Microbiota........... 18  
  Obesity ........................................................................................................................................ 18  
  Impact of Antibiotics ................................................................................................................... 19  
Affordable Nutritious Foods: Societal Implications and Challenges ............................................. 20  
Opening the Public Discussion ....................................................................................................... 22  
  Science and Technology .............................................................................................................. 22  
  Ethics .......................................................................................................................................... 22  
  Policy and Governance .............................................................................................................. 23  
Closing Thoughts ............................................................................................................................ 23  
Acknowledgments ........................................................................................................................... 25  
References ...................................................................................................................................... 26  
Figure Legends ............................................................................................................................... 36  
Figures .......................................................................................................................................... 40
Chapter 2

Gut bacteria that rescue growth impairments transmitted by immature microbiota from undernourished children

Abstract ..........................................................................................................................................45
Main Text .......................................................................................................................................45
  A Malawian model of gut microbiota development .................................................................46
  Identification of age-discriminatory taxa that are also growth-discriminatory......................48
  The relationship between age- and growth-discriminatory taxa and femoral phenotypes 51
  Effects of gut microbiota immaturity on host metabolism .....................................................52
  Repairing impaired growth phenotypes with age- and growth-discriminatory strains present in a healthy donor microbiota........................................................................................55
  Culturing invasive growth- and age-discriminatory taxa and characterizing their effects on growth ........................................................................................................................................57
Prospectus ......................................................................................................................................59
References and Notes ....................................................................................................................62
Acknowledgements ......................................................................................................................68
Data Deposition ............................................................................................................................68
Declarations ...................................................................................................................................69
Author Contributions ..................................................................................................................69
Figure Legends ..............................................................................................................................70
Figures ..........................................................................................................................................73
Supplementary Materials .............................................................................................................78
Materials and Methods ................................................................................................................78
  Human Studies ..........................................................................................................................78
  Bacterial 16S rRNA sequencing ..............................................................................................78
  Random Forests-derived Models and Statistical Analysis .....................................................79
  Gnotobiotic Mouse Studies .......................................................................................................79
    Preparation of the Malawi-8 (M8) Diet .................................................................................80
    Generation of clarified stool for gavage into gnotobiotic mice ..............................................81
    Quantitative magnetic resonance (qMR) analysis of body composition ...............................81
    Sample collection ..................................................................................................................82
    Micro-computed tomography (micro-CT) .............................................................................82
Chapter 3

Sialylated milk glycans promote growth in gnotobiotic mice and piglets with a stunted Malawian infant gut microbiota

Summary ..............................................................................................................................................117
Introduction ........................................................................................................................................117
Results ..................................................................................................................................................119
  Human milk oligosaccharide content correlates with growth in Malawian infants........119
  Considerations for testing the physiological effects of HMOs in gnotobiotic models....121
  A clonally arrayed, genome-sequenced collection of bacterial strains from the fecal microbiota of a stunted Malawian infant .................................................................121
  S-BMO supplementation of a prototypic Malawian diet improves growth in young mice colonized with the stunted Malawian donor’s cultured bacterial community .................123
  Members of the distal gut microbiota respond transcriptionally to S-BMO in vivo .....125
  Cross-feeding between the B. fragilis and E. coli strains ......................................................127
  S-BMO supplementation modulates metabolism in host tissues .................................128
  S-BMO supplementation promotes growth in gnotobiotic piglets ............................131
Discussion ..........................................................................................................................................134
Experimental Procedures .............................................................................................................137
  Human studies ...............................................................................................................................137
    Generating a clonally arrayed culture collection from a 6-month-old Malawian infant .137
    Purification and compositional analysis of S-BMO ...............................................................137
    Design and preparation of the Malawi 8 (M8) diet ...............................................................137
  Studies involving gnotobiotic mice .........................................................................................138
    16S rRNA gene sequencing of fecal microbiota .................................................................138
    Microbial RNA-Seq ..................................................................................................................138
    Mass spectroscopy of metabolites contained in biospecimens obtained from gnotobiotic animals .....................................................................................................................138
Supplemental Information .............................................................................................................139
Author Contributions ......................................................................................................................139
Acknowledgements ..........................................................................................................................140
Accesssion Numbers .......................................................................................................................140
References .......................................................................................................................................141
Figure Legends ............................................................................................................................ 149
Figures .......................................................................................................................................... 153
Extended Experimental Procedures .............................................................................................. 159

**Human studies** .................................................................................................................. 159

- Mass spectrometry of breast milk samples ........................................................................... 159
- Generating a clonally arrayed sequenced collection of bacterial strains from a 6-month-old severely stunted Malawian infant ................................................................. 161
- Purification and compositional analysis of S-BMO ............................................................... 164
- Design and preparation of the Malawi 8 (M8) diet ............................................................... 165

**Studies involving gnotobiotic mice** ...................................................................................... 166

- Quantitative magnetic resonance (qMR) analysis of body composition ............................. 167
- Sample collection ................................................................................................................ 167
- Micro-computed tomography ............................................................................................ 167
- Histological analysis of distal femur morphology ............................................................... 168
- Multiplex sequencing of amplicons generated from bacterial 16S rRNA genes ............... 168
- Microbial RNA-Seq ............................................................................................................ 169
- Community profiling by sequencing (COPRO-Seq) .......................................................... 170

**Mass spectroscopy of metabolites contained in biospecimens obtained from gnotobiotic animals** ........................................................................................................... 171

- Targeted ultra high performance liquid chromatography-mass spectrometry (UPLC-MS) of liver, muscle, brain, serum and cecal contents ......................................................... 171
- Non-targeted gas chromatography/mass spectrometry (GC-MS) of serum ....................... 171
- Non-targeted analysis of brain metabolites using GC-MS .................................................. 173
- Non-targeted UPLC-MS of brain metabolites ................................................................... 173

**Experiments involving gnotobiotic piglets** ......................................................................... 174

- Derivation of germ free piglets ............................................................................................ 174
- Experimental design ............................................................................................................ 175
- Microbial RNA-Seq ............................................................................................................ 176
- Gut mucosa glycan analysis by triple quadrupole mass spectrometry ............................... 177

**Measurement of serum leptin, insulin and triglycerides** .................................................... 178

*In vitro* S-BMO consumption and growth assays ................................................................. 178

**Statistical analysis** ............................................................................................................. 179

**Supplemental References** .................................................................................................. 180
Chapter 4

Future Directions

Characterizing the effects of S-BMO on bone morphology and metabolic flexibility ..........192
Identifying milk oligosaccharide structures that promote growth in gnotobiotic animal models
......................................................................................................................................................194
Dependence of milk oligosaccharide-mediated growth on members of the infant gut microbiota
............................................................................................................................................................195
Challenges facing milk oligosaccharides as novel therapeutics for the treatments of childhood
undernutrition.........................................................................................................................................197
References........................................................................................................................................200

Appendices

Appendix A ........................................................................................................................................202
Acknowledgements

The time I’ve spent living in St. Louis and working toward my Ph.D. has been incredibly transformative for me, both as a scientist and as a person. My accomplishments, particularly the work presented in this dissertation, have been made possible by the dedication and endless support of many important people in my life. This section is dedicated to thanking them all, because I would not be the person I am today without them.

First, I would like to acknowledge my Ph.D. advisor, Jeffrey Gordon. I am immensely thankful to Jeff for giving me the opportunity to join his laboratory and pursue my Ph.D. research on a topic that I find fascinating, to access incredible leading-edge technology, and to gain such diverse experiences. In my mind, Jeff’s greatest talent is bringing together an amazing group of people whose expertise and personalities foster an environment of collaboration and discovery, and that environment has been the ideal place for me to grow as a thinker and as a scientist. Beyond that, Jeff has always been completely supportive of me as a person, particularly as I’ve gone through times of great change or loss in my personal life, for which I can not adequately express my gratitude.

I would next like to thank the members of my thesis committee, Dan Goldberg, Scott Hultgren, Clay Semenkovich, Jeremy Buhler, and Michael Talcott, for their extraordinary feedback, advice, and guidance through the process of developing my thesis work. I am grateful to Barak Cohen, director of the Computational and Systems Biology program, Melanie Relich, and Jeanne Silvestrini, for their helpful guidance and advice in navigating the requirements of my Ph.D. program. I also thank Stephanie Amen and Margie Andersohn for administrative assistance and Laura Kyro for extensive help with graphic design and preparing this dissertation.

Many members of the Gordon lab have been essential to my work, and I thank them for their unique contributions. Without David O’Donnell, Maria Karlsson, Sabrina Wagoner, and Justin Serugo, it would not have been possible to perform any of my work in gnotobiotic mouse models. Janaki Lelwala-Guruge has been my go-to source of knowledge and help with anaerobic
microbiology. Jiye Cheng has been a guru of mass spectrometry knowledge and guidance. Marty Meier’s hard work and ingenuity have made processing of vast numbers of experimental samples possible in the span of a single Ph.D. Jessica Hoisington-Lopez and Su Deng have made next generation DNA sequencing not only possible, but readily accessible, easy, and fast. Laura Blanton has been my partner in crime in the lab. She and I have worked closely together on designing and executing experiments that I consider to be pivotal to the work presented in this dissertation. I have been privileged to work alongside her on these projects. I am also grateful to Philip Ahern, Meng Wu, Mike Barratt, Nathan McNulty, Jeremiah Faith, Michelle Smith, Ansel Hsiao, Mark Gonzalez, Federico Rey, Alejandro Reyes, and Sid Venkatesh collectively for their guidance and mentorship and helping to shape my thinking about science as well as their endless help performing the experiments presented in this dissertation. Members of the Gordon lab, especially David, Maria, Sabrina, Janaki, Laura, Philip, Meng, and Ansel, have become my close friends, and this group of people has been unbelievably important in my life over the last several years.

I would like to specifically express my gratitude to the group of people that have made it possible to establish a gnotobiotic model in newborn piglets, including Jeff Gordon, Michael Talcott, David O’Donnell, Maria Karlsson, Sabrina Wagoner, Justin Serugo, Sid Venkatesh, Jeanette Gehrig, Angie Lewis, Julie Maurer, Charles Mitchell, and Alicia Sexauer. It has taken a tremendous amount of dedication, perseverance, coordination, and sleepless nights to make this model a reality.

My thesis work has benefited extensively from collaborations with brilliant and kind scientists at other institutions. I am thankful to David Mills, Carlito Lebrilla, Daniela Barile, Chao Wu, and Jasmine Davis for their immense knowledge of milk oligosaccharides and mass spectrometry as well as for allowing me access to mixtures of bovine milk oligosaccharides for use in my studies. I am also grateful to Christopher Newgard, Olga Ilkayeva, James Bain, and Mike Muehlbauer for their knowledge, helpfulness, and enthusiasm for metabolomics analyses.
Lastly, and most importantly, I would like to thank my family. It has been through the support, commitment to education, and love of my mother and father, Terri and Mark, that I have been able to pursue a Ph.D. and a career in the sciences. My brother Jason, his wife Stephanie, and their children, Carson and Claire, have brought me pride and happiness that helped to keep me level through the tumultuous process of graduate school. My grandparents, Ken and Ruth Charbonneau, and Ruth and Wilbur Schnell, were always sources of joy and support growing up, helping to make sure that I succeeded in my education. My extended family number too high to name individually for all the ways that they’ve helped shape me into the person I have become, but I am deeply grateful for what they have given me.

I want to conclude this section by properly acknowledging the person who has been the most influential in my life, and that is my mother, Terri. My mom has been, at every step of my life and education, the person driving me to succeed and fulfill my potential. It was thanks to her influence that I always had the resources I needed to push myself to excel in high school. My mom’s vision for my future is what helped me become the first college graduate in my immediate family, including working a second job to make sure that we could afford it. It was my mom’s encouragement and support at a time in my life when I was floundering about what to do with my career that I decided to apply for and persevere through graduate school. I owe a great deal of what I’ve accomplished singularly to her, and I am eternally grateful.
ABSTRACT OF THE DISSERTATION

Characterizing the Role of Sialylated Milk Glycans and the Infant Gut Microbiota in Growth and Metabolism

by

Mark Charbonneau

Doctor of Philosophy in Biology and Biomedical Sciences
Computational and Systems Biology
Washington University in St. Louis, 2015

Professor Jeffrey I. Gordon, Chair

Undernutrition is a pressing and pervasive global health problem. The pathogenesis of this disease remains unclear, but epidemiologic studies indicate that it is not due to food insecurity alone. Moreover, current therapeutic interventions have limited efficacy in preventing or ameliorating the long-term sequelae of undernutrition, including stunting and cognitive deficits. Recent culture-independent studies have demonstrated that the normal postnatal pattern of gut microbiota development is disrupted in children with undernutrition, leading to the proposal that perturbations in gut microbiota development impairs healthy growth of the host. Human breast milk contains a diverse repertoire of free and conjugated human milk oligosaccharides (HMOs) frequently decorated with fucose and/or sialic acid moieties. HMOs are not absorbed in the proximal gut and function as prebiotics that promote colonization of the infant gut with bacterial strains associated with numerous benefits (e.g., enhanced gut barrier function, protection from enteropathogen infection, and improved vaccine responses). To date, the relationships between the overall abundance of breast milk HMOs, development of a healthy gut microbiota, and the growth phenotypes of infants and children have not been well characterized. The central hypotheses of my thesis are that (i) the gut microbiota plays a key role in supporting healthy growth and metabolism in infants and children, (ii) perturbations in microbiota function are causally related to undernutrition in infants and chil-
dren, (iii) the representation of HMO species in mothers’ milk is significantly correlated with the nutritional status and growth outcomes of their offspring (as defined by anthropometry), and (iv) HMOs promote growth and modulate metabolism through microbiota-dependent mechanisms and may yield new therapeutic agents for treating and ultimately preventing undernutrition. My thesis is presented in four chapters.

The first chapter is presented as a perspective, describing the relationship between perturbations in postnatal development/maturation of the gut microbiota and childhood undernutrition. This chapter describes hypotheses and experimental models for performing proof-of-concept experiments that test the causal role of the gut microbiota, milk glycans, and interactions between these components in modulating host metabolism and growth.

Chapter two describes a set of experiments that involve transplanting fecal microbiota from Malawian infants and children (manifesting healthy growth or varying degrees of undernutrition) into young germ-free mice fed a representative Malawian diet. These studies revealed that immature microbiota from severely stunted/underweight donors transmit impaired growth phenotypes as compared to microbiota from healthy donors. The representation of several taxa in the microbiota of recipient animals correlated with lean body mass gain, bone morphology, and metabolic phenotypes in liver, muscle, and brain. Furthermore, co-housing these ‘humanized’ gnotobiotic mice shortly after colonization revealed that invasion of bacterial taxa from mice harboring a healthy infant’s microbiota to cagemates harboring an immature microbiota from a stunted/underweight donor ameliorated growth faltering. These results indicate that gut microbiota immaturity is causally related to undernutrition.

The third chapter begins by defining the relationship between infant growth outcomes and breast milk HMO content in two independent Malawian birth cohorts. Analysis of human milk oligosaccharides from 6-month postpartum Malawian mothers revealed that sialylated HMOs are significantly less abundant in milk from mothers with severely stunted infants. This chapter then describes the effects of sialylated milk oligosaccharides on growth and metabolism using young
gnotobiotic mice and newborn gnotobiotic piglets. In both cases, animals were colonized with a sequenced bacterial culture collection generated from a severely stunted Malawian infant, and fed a prototypic Malawian diet with and without supplementation using a purified preparation of sialylated bovine milk oligosaccharides (S-BMO). S-BMO produced a microbiota-dependent augmentation of body weight and lean body mass gain, changed bone morphology, and altered liver, muscle, and brain metabolism in ways indicative of a greater ability to utilize nutrients for anabolism. These two preclinical models establish a causal microbiota-dependent relationship between S-BMO and growth promotion.

Chapter four details future research directions, including experimental approaches for determining the microbial dependencies of S-BMO mediated growth and identifying bioactive structures present in S-BMO. Finally, this chapter describes the potential for and challenges facing the use of dietary milk oligosaccharide supplements to treat childhood undernutrition.
Chapter 1

Introduction
Chapter 1

Introduction

Cultivating Healthy Growth and Nutrition through the Gut Microbiota

Sathish Subramanian\textsuperscript{1,2}, Laura V. Blanton\textsuperscript{1,2}, Steven A. Frese\textsuperscript{3}, Mark Charbonneau\textsuperscript{1,2}, David A. Mills\textsuperscript{3}, Jeffrey I. Gordon\textsuperscript{1,2}

\textsuperscript{1}Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63108, USA

\textsuperscript{2}Center for Gut Microbiome and Nutrition Research, Washington University School of Medicine, St. Louis, MO 63108, USA

\textsuperscript{3}Departments of Food Science \& Technology and Viticulture \& Enology, University of California, Davis, Davis, CA 95616, USA

This chapter corresponds to the complete and accepted version of a manuscript published as:

Abstract

Microbiota assembly is perturbed in children with undernutrition, resulting in persistent microbiota immaturity that is not rescued by current nutritional interventions. Evidence is accumulating that this immaturity is causally related to the pathogenesis of undernutrition and its lingering sequelae. Preclinical models in which human gut communities are replicated in gnotobiotic mice have provided an opportunity to identify and predict the effects of different dietary ingredients on microbiota structure, expressed functions, and host biology. This capacity sets the stage for proof-of-concept tests designed to deliberately shape the developmental trajectory and configurations of microbiota in children representing different geographies, cultural traditions, and states of health. Developing these capabilities for microbial stewardship is timely given the global health burden of childhood undernutrition, the effects of changing eating practices brought about by globalization, and the realization that affordable nutritious foods need to be developed to enhance our capacity to cultivate healthier microbiota in populations at risk for poor nutrition.

Introduction

Understanding the determinants of the nutritional value of different foods has never been more important, with population stabilization being unlikely this century (Gerland et al., 2014) and growing challenges related to sustainable agriculture. An integral part of understanding how best to deliver nutritious food to a burgeoning population is understanding how the microbial community in our gut (the microbiota) is shaped by what we eat and how that community in turn shapes our development and health. Nowhere will this kind of insight be more crucial than in raising the world’s children.

Current obstacles to achieving healthy and productive lives and societies are reflected in the United Nations’ millennium development goals that include reductions in child mortality and hunger and improvements in maternal health (http://www.un.org/millenniumgoals/). The scope of the problem of childhood undernutrition is described by parameters such as the International Food Policy Research Institute’s Global Hunger Index (http://www.ifpri.org/publication/2014-global-
hunger-index), which is an aggregate measure of calorie intake plus the rates of children being underweight and childhood mortality within a given region or country.

Much has been said about how changing patterns of food preferences brought about by economic development, globalization, and changes in food technology and food distribution systems are producing dramatic changes in how, what, and when we eat. These changes, combined with rapid population expansion and issues related to sustainable agriculture, create the need and the opportunity to drive innovation in the area of identifying new, affordable, and nutritious foods. Here, we focus on the importance of understanding the postnatal developmental biology of our gut microbial community—a highly adaptable microbial “organ” that is critically involved in the biotransformation of foods to products that can shape many aspects of human biology. In our view, studies of human gut microbial communities will markedly revise current thinking about many aspects of human nutrition. The knowledge gained could and should catalyze efforts to integrate agricultural policies, food production, and nutritional recommendations for consumers representing different ages, cultural traditions, and geographies. Preclinical research platforms are now available to evaluate the effects of foods that we currently consume and those that we envision creating in the future on the gut microbial community and host biology in ways that can inform clinical studies. Furthermore, studies of children with undernutrition are highlighting the importance of postnatal development of the gut microbiota for achieving healthy growth and providing us with a new set of metrics to define the efficacy of nutritional recommendations and interventions directed at infants, the maternal-infant dyad, and children. Finally, we emphasize the importance of addressing ethical, social, and regulatory issues related to research in this area now rather than later.

**Defining Human Postnatal Development from a Microbial Perspective**

The human gut microbiota is composed of all three domains of life; Bacteria, which predominate, Archaea, and Eukarya, plus viruses. The gut microbiota is composed of relatively few bacterial phyla compared to communities in other body habitats and is notable for its strain-level diversity. Application of low-error sequencing methods to PCR amplicons generated from the bacterial phy-
logenetic marker gene encoding the principal RNA in the small subunit of ribosomes (16S rRNA) has indicated that, once acquired, the majority of bacterial strains in a healthy adult are retained for long periods of time (Faith et al., 2013). Thus, early colonizers, once established in the gut ecosystem, have the potential to exert their effects on our biological features and health status for most and perhaps all of our adult lives. This latter finding emphasizes the importance of understanding whether there is a definable program of community assembly in healthy infants/children and whether such a program is shared or varies considerably across populations with distinct dietary habits and culinary traditions residing in different geographic locations. If such a developmental program were definable and a significant contributor to healthy growth, fostering its proper and full execution could represent the basis of an arm of preventive medicine designed to ensure long-term health through informed microbial stewardship.

Food is a major factor that shapes the proportional representation of microorganisms present in the gut microbiota and the relative abundance of its genes (the microbiome). Reciprocally, the configuration of the microbiota/microbiome influences the nutritional value of food. One illustration of this interrelationship comes from a culture-independent metagenomic analysis of the gut microbiomes of infants, children, and adults belonging to 150 families living in three countries located on three different continents (metropolitan areas of the USA plus rural villages in southern Malawi and the Amazonas state of Venezuela). The results revealed that the relative abundances of genes in the microbiome that are related to vitamin biosynthesis (e.g., folate, cobalamin, thiamine, and biotin), amino acid metabolism, and processing of complex polysaccharides change in an identifiable sequence during the postnatal period (Yatsunenko et al., 2012). In addition, differences between Westernized (USA) and non-Westernized populations were evident, with breastfed Malawian and Amerindian babies having higher relative abundances of microbial genes encoding enzymes involved in carbohydrate metabolism, vitamin biosynthesis (e.g., components of the biosynthetic pathway for riboflavin, a component of breast milk, dairy products, and meat), and urease (Yatsunenko et al., 2012). Urea represents up to 15% of breast milk nitrogen; its degradation to ammonia can be used for microbial biosynthesis of essential amino acids, potentially benefiting
both the microbiota and host when diets are deficient in protein. Significant differences in microbiome configuration were also observed between breast-fed and formula-fed infants, with the latter showing increased representation of genes involved in various aspects of carbohydrate and amino acid metabolism and cobalamin (vitamin B12) biosynthesis (Yatsunenko et al., 2012). Cobalamin is not only important for the host; the ability to transport cobalamin and other substituted corrinoids is an important determinant of survival for members of the microbiota (Degnan et al., 2014).

Together, these findings suggested that the gut community should be considered when assessing the nutritional requirements at different stages of the human life cycle and in different geographic/cultural settings. They also raised the question of whether perturbations in the functional development of the microbiota/microbiome were related to childhood undernutrition, the major cause of childhood deaths worldwide and a manifestation of a complex set of still poorly understood intra- and intergenerational factors, rather than food insecurity alone (Lazzerini et al., 2013, Caulfield et al., 2014 and Richard et al., 2014).

**Undernutrition and Gut Microbiota Immaturity**

The World Health Organization’s (WHO) Multi-Center Growth Reference Study (http://www.who.int.beckerproxy.wustl.edu/childgrowth/mgrs/en/) defines three anthropometric (physical) parameters (weight-for-age, height-for-age, and weight-for-height Z scores) to describe normal early childhood growth and nutritional status from its evaluation of 8,440 infants and children living in six distinct sites around the world (USA, Oman, Norway, Brazil, Ghana, and India). A recent study provided another definition of healthy growth but from a microbial perspective (Subramanian et al., 2014). It did so by examining gut microbiota assembly in 50 children residing in Dhaka, Bangladesh whose anthropometry during their first 2 years of life indicated healthy growth. Fecal samples were collected monthly from birth through the end of the second postnatal year, and the relative abundances of bacterial strains were analyzed by 16S rRNA amplicon sequencing. The results revealed that interpersonal variation in the bacterial component of their gut communities was significantly smaller than the variation associated with age. Applying Random Forests, a machine-
learning method, to regress relative abundances of bacterial taxa across these children revealed age-discriminatory bacterial strains. Separating these 50 children into training and validation cohorts, the regression was optimized to include the most informative taxa for accurate prediction of microbiota “age.” The results were formally validated to prevent over-fitting and over-estimation of generalizability and produced a sparse model composed of 24 strains that could be used in aggregate as a microbial signature for describing a shared program of microbiota development in healthy individuals and two derived metrics for defining deviations from that normal program: “relative microbiota maturity” and “microbiota-for-age” Z (MAZ) score (Figure 1).

Severe acute malnutrition (SAM) is defined by weight-for-height Z (WHZ) scores more than 3 SDs below the median of children in the WHO reference cohort. Application of this sparse model to 64 Bangladeshi children with SAM (WHZ −4.2 ± 0.72 [SD]) revealed they had gut microbiota that appeared significantly “younger” than their chronological age (relative microbiota maturity of −6 ± 0.7 months and MAZ scores of −1.7 ± 0.2). Moreover, this immaturity was incompletely and only transiently rescued following a customary period of administration of either one of two types of ready-to-use therapeutic foods (RUTFs; typically given for 2 weeks until a 15% increase in weight gain is achieved; http://www.ClinicalTrials.gov, number NCT01331044). Bangladeshi children with moderate acute malnutrition (WHZ between −3 and −2) also exhibited significant microbiota immaturity, although less severe than children with SAM (Subramanian et al., 2014). These results indicate that children with SAM have a persistent developmental abnormality affecting their gut microbial “organ” that is not durably repaired with existing therapy.

These observations raise a critical question: is microbiota immaturity a cause or an effect of childhood undernutrition? Many studies have shown that, although current protocols for treating children with (acute) undernutrition reduce mortality, they do not rescue its long-term morbidities, including stunting, immune dysfunction, and neurodevelopmental abnormalities (Victora et al., 2008, Gaayeb et al., 2014, Kosek et al., 2013 and Galler et al., 2012). For example, given the remarkable metabolic requirements of the neonatal brain, alterations in the normal postnatal de-
velopment of the gut microbiota may trigger marked impairments in brain development and lead to persistent disorders of cognition.

Support for a causal role for the gut microbiota in SAM comes from studies of gnotobiotic mice. In recent years, methods have been developed for transplanting previously frozen fecal samples from human donors into groups of germ-free mice at a selected stage of their lives (e.g., young, rapidly-growing animals that have been recently weaned or older animals) and with a designated genetic background. If the human microbiota sample is frozen shortly after it is produced and maintained at −80°C, the bacterial strains represented in the donor’s community can be transmitted efficiently and reproducibly to recipient mice (e.g., Turnbaugh et al., 2009a, Smith et al., 2013, Ridaura et al., 2013, Palm et al., 2014 and Kau et al., 2015). The recipient mice can be fed diets that contain ingredients used in foods consumed by the microbiota donor. Moreover, the ingredients and methods for preparing (cooking) such diets can be varied systematically. This approach allows myriad types of models to be constructed for studying the interaction of foods and the human gut microbiota in vivo. For example, diets can be given that are representative of those consumed by populations other than those of the donor to anticipate the effects of changes in food consumption patterns associated with Westernization or composed of ingredients that represent new potential sources of affordable, nutritious foods such as landraces and waste streams from current food manufacturing processes. Critically, these preclinical gnotobiotic animal models allow proof-of-concept tests of whether a donor phenotype is transmissible via his/her gut microbiota, the extent to which phenotypic transmission generalizes across different donor microbiota, and the sensitivity or robustness of phenotypic transmission to diet type. These preclinical models also permit simulations of existing or anticipated therapeutic interventions, including the opportunity to “randomize” a given individual’s microbiota to not just one but multiple treatment arms in order to directly compare the effect (and effect size) of the treatments on both the microbiota and host, to characterize underlying mechanisms, and to identify surrogate- or mechanism-based biomarkers that could be translatable to the microbiota donor or donor population (Figure 2).

Transplanting fecal microbiota from same-gender Malawian twins discordant for kwashiorkor, a form of SAM, into separate groups of adult germ-free mice and feeding the recipient animals
a representative micro- and macronutrient-deficient Malawian diet disclosed that the healthy and kwashiorkor co-twins’ microbiota transmitted discordant weight loss and metabolic phenotypes (as well as an enteropathy characterized by disruption of the small intestinal and colonic epithelial barrier in animals harboring kwashiorkor but not healthy microbiota) (Smith et al., 2013 and Kau et al., 2015). Unlike the transplanted healthy co-twins’ microbiota, the kwashiorkor microbiota was structurally and metabolically labile, reconfiguring itself upon exposure to a peanut-based RUTF, but not in a sustained way when animals were returned to the Malawian diet. The combination of a nutrient-deficient Malawian diet and a kwashiorkor microbiota was required to produce pathology in the recipient “humanized” mice, including inhibition of steps within the tricarboxylic acid cycle in host cells (Smith et al., 2013). These findings not only provided evidence for a causal relationship between the gut microbiota and SAM but also highlighted the importance of diet-by-microbiota interactions in disease pathogenesis.

If we consider children with persistent microbiota immaturity from the perspective of developmental biology, we can pose a number of basic and applied scientific questions. One question is whether the developmental program defined in Bangladeshi infants and children is generalizable to other populations representing different geographic and cultural settings. If so, it would reveal a fundamental shared aspect of postnatal human development and raise mechanistic questions about the factors that specify a healthy microbial community “fate.” Initial support for generalizability comes from an analysis of concordant healthy Malawian twin pairs, which showed that a number of the age-discriminatory bacterial strains with the highest feature importance scores in the Bangladeshi Random Forests model are also represented in the Malawian population (Subramanian et al., 2014 and Yatsunenko et al., 2012). The designation “same strain” was based on the same 16S rRNA sequence; whole-genome sequencing of a given age-discriminatory strain identified by its 16S rRNA sequence will be needed to determine its degree of gene conservation across different Bangladeshi and Malawian hosts. Bacterial 16S rRNA analyses of fecal samples obtained at monthly intervals from infants and children with healthy growth phenotypes enrolled in birth cohorts living at multiple low-income countries allow country/community site-specific, Random-
Forests-based models of microbiota maturation to be constructed, as well as an aggregate model representing data pooled from all sites. “Generalizability” can be established through reciprocal tests of the accuracy of the site-specific models (and aggregate model) for healthy individuals living at the different sites and whether these models reveal similar relationships between anthropometry and relative microbiota maturity/MAZ scores for undernourished children living at each of these sites.

A second question has to do with the relationship between microbiota development, enteropathogen load, and environmental enteric dysfunction (EED, also known as environmental enteropathy), an enigmatic and as-yet incompletely defined disorder of gut barrier function (Keusch et al., 2014 and Kosek et al., 2014). Does a primary failure to execute normal maturation of the microbiota directly influence risk for enteropathogen invasion, perturbations in development of mucosal immune system, and abnormalities in nutrient processing and absorption that ultimately results in growth faltering? Alternatively, is a holistic view required that considers each of these features of enteric biology as intimately and integrally related to one another? Large birth cohort studies such as MAL-ED and GEMS have provided an opportunity to measure the contributions of enteropathogen load/carriage and diarrheal incidence to growth faltering (MAL-ED Network Investigators, 2014, Platts-Mills et al., 2014 and Kotloff et al., 2013). Evidence is emerging that some of the age-discriminatory taxa that define normal microbiota maturation also protect the host from enteropathogen infection. Intriguingly, studies of Bangladeshi adults with acute cholera have shown that recovery from the diarrheal phase involves recapitulation of the sequence of appearance of the same age-discriminatory bacterial strains that define the normal pattern of assembly of the microbiota in healthy Bangladeshi infants/children, suggesting that an essential set of rules governs this assembly (successional) process (Hsiao et al., 2014). For example, *Ruminococcus obeum*, a bacterium that directly correlates with recovery from *Vibrio cholerae* infection in adult Bangladeshi subjects and defines later stages of normal gut microbiota maturation in healthy Bangladeshi children, restricts *V. cholerae* colonization of gnotobiotic mice harboring a representative human gut microbiota. Its mechanism involves production of an autoinducer-2 (AI-2) that causes
quorum-sensing mediated repression of *V. cholerae* colonization and virulence factor expression (Hsiao et al., 2014).

A third related question is the manner in which the mucosal immune system and the microbiota co-develop. How do these complex organs talk to and educate each other? The answers could help identify factors that legislate a normal developmental trajectory for a gut community and how developmental arrest of the microbiota could become fixed and difficult to overcome/advance. Immaturity of the microbiota may be associated with relative immaturity of mucosal immunity in ways that impede responsiveness to vaccines or enteropathogens. If so, can we use members of the microbiota as next-generation adjuvants to prime the immune system in the context of a defined antigen (Yilmaz et al., 2014)? One way to characterize maturation of the mucosal immune system is to use fluorescence-activated cell sorting (FACS) to identify microbial taxa targeted by its IgA responses as a function of chronologic age in hosts with healthy growth phenotypes and in those with undernutrition (critically, IgA targeting is not simply a reflection of the abundances of organisms in the gut community; Kau et al., 2015). This method, named BugFACS, has identified bacterial targets of gut mucosal IgA responses using fecal samples from children with healthy growth phenotypes or those with varying degrees of undernutrition, as well as fecal samples harvested from gnotobiotic mice harboring transplanted microbiota from healthy and undernourished donors fed diets representative of those that these children consume. BugFACS-purified viable IgA-targeted bacterial taxa were subsequently introduced into germ-free animals fed nutrient-deficient or -sufficient diets to characterize their functional properties. The results disclosed that IgA responses to members of the microbiota can be used as biomarkers of growth faltering, that they are influenced by enteropathogen load, and that they mediate a diet-dependent enteropathy characterized by small intestinal and colonic epithelial barrier disruption. Moreover, treatment with IgA-targeted bacterial strains purified from healthy donor microbiota can prevent development of the enteropathy (Kau et al., 2015), indicating that this approach may have utility that extends beyond diagnostics to therapeu tic lead discovery and defining mechanisms underlying EED pathogenesis.

A fourth and critical question is whether age-discriminatory taxa are not only just bio-
markers but also effectors of growth. If so, they become potential therapeutic agents and targets for manipulation, including food-based manipulations that allow for their establishment in an individual or population at the time of presentation with manifest disease or prior to that time. One way we are currently determining whether age-indicative taxa are also growth indicative is by transplanting microbial communities from children exhibiting varying degrees of growth faltering (defined by anthropometry), representing a particular geographic region, into young, actively growing germ-free animals fed diets representative of the donor population and then defining the effects of the different transplanted communities on the growth, metabolic and immunologic phenotypes of recipient gnotobiotic mice (Figure 2). 16S rRNA data sets generated from the animals’ fecal samples can be used to correlate strain abundances to these phenotypes. These strains can then be cultured from the microbiota of different donor populations. Determining the effects of subsequently introducing these strains—singly or as components of defined consortia—into young gnotobiotic mice harboring microbiota from different undernourished donors represents a way to address several challenges that would be faced when designing and interpreting a clinical study. For example, these preclinical studies could help to (1) define criteria used to select strains beyond their feature importance scores in the Random Forests models and cultivability (e.g., the extent of representation of virulence determinants in their genomes); (2) assess how to encapsulate these organisms, including anaerobes, in ways that permit their long-term storage and viability; (3) determine the extent to which consortia can invade and establish themselves in different microbiota representing individuals from a given population or different populations; (4) assess the nature of their effects on growth (e.g., gain of lean body mass), metabolism, and gut barrier function as a function of the degree of donor undernutrition and microbiota immaturity; and (5) ascertain the degree to which invasion and establishment of these strains in the targeted microbiota and their host effects are impacted by diet. Determining whether these strains are interchangeable between countries will influence the generalizability of microbial interventions or whether there would have to be local sourcing of these biological resources by or for the communities who are themselves afflicted by undernutrition.
Establishing Microbiota and the Maternal Influence

The origins of the microbes that colonize an infant’s gastrointestinal tract are complex, given that infants are exposed to different environmental sources. A major source is the mother and includes microbes from her vagina, skin, gut, and as some have reported, breast milk and possibly the placenta (Dominguez-Bello et al., 2010, Hunt et al., 2011, Grönlund et al., 2011, Cabrera-Rubio et al., 2012 and Aagaard et al., 2014).

A key knowledge gap relates to the “anthropology of microbes”: knowing how practices associated with pregnancy, including micronutrient supplementation, as well as traditional (and changing) societal “prescriptions” for dietary practices, impact a mother’s microbial ecology prior to and following parturition and how this may impact transmission of her microbes to her infant. A study of 91 pregnant Finnish women showed that the maternal microbiota changes between the first and third trimester (Koren et al., 2012) (Figure 3). Another analysis of Bangladeshi mothers revealed marked changes in their gut microbiota in the first month post-partum, followed by less substantial changes in the ensuing nine months (Subramanian et al., 2014). One testable hypothesis is that the maternal microbiota, much like the infant microbiota, undergoes stereotypical alterations during normal pregnancy designed to enhance maternal health and to promote transfer of strains to the infant. Testing this hypothesis will require detailed time series sampling of maternal microbiota throughout pregnancy and of the maternal-infant dyad, plus other environmental sources, including other family members and caregivers. If a program of pregnancy-associated changes in the maternal gut microbiota can be identified using approaches analogous to those described above to characterize maturation of the infant microbiota, it could provide an opportunity to use the most indicative or transmissible taxa as biomarkers of nutritional status and as reporters of the effects of different dietary practices or the efficacy of prescribed prenatal nutritional interventions.

Pregnancy is also a time of increased susceptibility to infection. Rowe et al. (2011) demonstrated that pregnant mice show increased bacterial burden in models of *Listeria monocytogenes* and *Salmonella typhimurium* infection, mediated via active immune suppression by a population
of FoxP3+ regulatory T cells (Tregs). Moreover, ablation of the Treg compartment resulted in near-complete resorption of fetuses, indicating a delicate balance between immunological tolerance of the fetus and defense against enteropathogens (Rowe et al., 2011). It is not known how this period of deliberate immune suppression impacts the maternal microbiota and, in turn, transfer of pathogens (and other microbial community members) to the infant.

**The Impact of First Foods**

**Breast Milk**

The association between healthy postnatal growth and exclusive breastfeeding has led to the WHO’s recommendation for a minimum of six months of exclusive breastfeeding (Kramer and Kakuma, 2002). Human milk is composed of lipids (tri-, di-, and monoglycerides, phospholipids, glycolipids, and free fatty acids), protein components (including immunoglobulins, lactoferrin, lysozyme, and cytokines), and a large repertoire of human milk oligosaccharides (HMOs). Over time, this composition changes from colostrum, which is HMO rich, to mature milk, which contains fewer HMOs and protein while the fat content remains relatively stable (Coppa et al., 1993 and Lemons et al., 1982).

HMOs and other milk glycoconjugates pass undigested through the proximal gut (Engfer et al., 2000) and serve as nutrient substrates for saccharolytic microbiota in the colon. The microbiota of healthy exclusively breastfed infants is dominated by members of the genus *Bifidobacterium* (Figure 1; Yatsunenko et al., 2012 and Subramanian et al., 2014). These infant-associated bifidobacteria, notably *Bifidobacterium longum* subsp. *infantis*, possess a suite of genes involved in importing complex fucosylated and sialylated milk glycans, their further degradation, and subsequent utilization (Sela et al., 2008). The functions encoded by this suite of genes allow them to outcompete other saccharolytic taxa (Marcobal et al., 2010). Bifidobacteria also actively reshape milk composition. For example, they release N-linked glycans conjugated to milk glycoproteins
for use as a growth substrate. However, the effect of deglycosylation on milk protein digestibility and function is as-yet unknown (Garrido et al., 2012 and Garrido et al., 2013).

Colonization by *Bifidobacterium* species during nursing is associated with a range of benefits, including improved vaccine responses (Huda et al., 2014) and enhanced gut barrier function (Ewaschuk et al., 2008 and Weng et al., 2014), including stabilized epithelial tight junctions noted in both animal models (Bergmann et al., 2013) and human cell lines (Chichlowski et al., 2012). Recent work has shown that infants with high *Bifidobacterium* population densities exhibit a corresponding decrease in fecal milk glycans (De Leoz et al., 2015 and Wang et al., 2015), a relationship that could serve as the basis for developing inexpensive diagnostics to monitor development of a healthy gut microbiota in nursing infants.

Development of a healthy infant gut microbiota can be threatened by maternal undernutrition and premature birth. Maternal undernutrition during pregnancy increases risk for underweight and preterm births (Kramer et al., 1992). Children of undernourished mothers receive substantially less than the recommended intake of priority micronutrients during lactation (Allen, 2005). Fortified milk obtained from donors who have had a full-term pregnancy likely does not provide sufficient protein to preterm infants (Arslanoglu et al., 2009). Even when mothers of preterm infants can produce sufficient milk, alterations in milk fat, protein, oligosaccharide content (Weber et al., 2001 and De Leoz et al., 2012), and the repertoire of immunoactive components (Castellote et al., 2011) are observed, leading to a call for identifying additional elements for nutritional support of these infants (Gabrielli et al., 2011 and De Leoz et al., 2012).

A vicious cycle of maternal undernutrition and poor infant nutritional status can reflect alterations in the immune, HMO, and/or other components of mother’s milk. This has critical implications for infant health. Poor maternal health is associated with variations in breast milk immunoglobulins and glycoprotein structures during lactation (Smilowitz et al., 2013) and with decreased lactoferrin, a protein with antimicrobial activities (Hennart et al., 1991). Parasite-specific breast milk IgA titers to *Entamoeba histolytica* and *Cryptosporidium* spp. correlate with nutritional
status in a Bangladeshi infant population in which the burden of infection with these enteropathogens is very high (Korpe et al., 2013). Preterm delivery is associated with atypical variations in milk glycan structures (De Leoz et al., 2012), which poses additional risks. As HMOs have structural similarities to epithelial cell surface and mucus glycans, they can have anti-adhesive effects on enteropathogens. Sialic acid or fucose moieties are key determinants of this activity. Thus, variations in fucosylated HMOs associated with preterm birth may reduce the efficacy of milk oligosaccharides as anti-adhesive decoy molecules for pathogens (Ruiz-Palacios et al., 2003 and Jantscher-Krenn et al., 2012).

Understanding how breast milk glycan repertoires correlate with normal microbiota assembly and with impaired microbiota maturation and undernutrition provides an opportunity to identify new glycan streams that could be used to treat undernourished infants. Commercial prebiotics are commonly added to infant formula, where they increase Bifidobacteria titers in infant feces (Haarman and Knol, 2005, Knol et al., 2005 and Boehm et al., 2002) and lower the incidence of pathogens (Knol et al., 2005). However, current prebiotics, namely fructooligosaccharides and galactooligosaccharides, do not represent the constellation of complex glycan structures delivered in human milk. Moreover, their consumption is not restricted to the population of microbes that define normal gut microbiota maturation (Everard et al., 2014 and Dewulf et al., 2013). Numerous efforts to recreate the glycan landscape present in human milk are underway. The technology for chemical and chemoenzymatic construction of complex “milk” oligosaccharides has advanced tremendously, enabling wholesale construction of a limited number of HMO-like structures present in milk (Muthana et al., 2009). Alternatively, purification from animal milks presents another opportunity for rapid and large-scale acquisition of milk oligosaccharides and glycoconjugates. At present, a number of enriched or purified bovine milk glycoproteins, including immunoglobins, lactoferrin, glycomacropeptide, and glycolipids are commercially available or could be readily produced at scale for use in preclinical and clinical studies. Bovine milk contains a relatively low concentration of free oligosaccharides, but the distribution of structures observed roughly matches the most abundant species present in HMOs (Aldredge et al., 2013). Importantly, bovine milk
oligosaccharides (BMOs) can be sourced from numerous points in dairy processing, including cheese whey, suggesting an opportunity for large-scale production of fractions enriched for given (or similar) structures (Zivkovic and Barile, 2011).

**Serial Introduction of Complementary Foods in Ways that Promote Maturation of the Gut Microbiota**

A recent study compared the microbiota and immune system in bottle-fed versus breastfed macaques. The results showed that breastfed infant macaques develop more robust $T_{H17}$ cells in the memory pool, suggesting that the timing and trajectory of dietary exposures during early life may have lasting functional consequences beyond that period (Ardeshir et al., 2014). In breastfed humans, the transition to formula feeding and family foods (complementary feeding practices) varies considerably in terms of which food types are consumed, the order of their presentation, and the duration of their consumption. Documenting which foods growing infants consume and in what quantities has required innovative approaches, particularly in low-income countries where undernutrition is prevalent (Caulfield et al., 2014) (Figure 3). For example, data collection protocols across eight different countries have been harmonized to enable quantification of variations in child feeding practices in the MAL-ED consortium (Caulfield et al., 2014).

The co-linearity between the introduction of various types of solid foods, reduction in breast milk consumption, and maturation of the gut microbiota makes it challenging to identify causal relationships between specific ingredients and the representation of specific microbes through human studies. However, studies in gnotobiotic mice colonized with defined collections of cultured (and sequenced) human gut-derived bacteria have been successful in interrogating specific food-microbe associations (Faith et al., 2011). These relationships were identified using an experimental design in which a given gnotobiotic animal harboring a defined microbial consortium received a sequence of diets, composed of several different combinations of foods, whose concentrations are intentionally varied between diets. The order of presentation of the different diets was also varied between different mice in order to limit confounding from hysteresis.
effects. This approach has identified associations between various commercially available foods given in the USA during the complementary feeding period and specific microbes independent of their order of presentation, which would be virtually impossible to identify in clinical studies of developing human infants (Faith et al., 2011). This approach can be applied to young mice colonized with the age- and healthy growth-associated bacterial strains identified using the methods described above to determine which complementary foods promote their representation and expressed functional features. The results could lead to a recommended sequence of complementary feeding that reflects local food availability, affordability, and cultural practices and that sponsors healthy microbiota maturation. This information would advance current recommendations, which are not microbiota based and quite general (Kleinman, 2000).

Additional Considerations Regarding the Developmental Biology of the Gut Microbiota

Obesity

Although we have emphasized the global challenge of undernutrition in children, another vexing global health problem is the growing burden of obesity and associated metabolic dysfunction in children. Increasing attention is being paid to delineating differences in the gut microbiota of children who become obese in the hopes that early recognition of perturbed microbiota development may permit early interventions in at risk populations. For example, a recent culture-independent study of a Singaporean birth cohort disclosed that precocious maturation of the microbiota during the first 6 months of postnatal life was associated with significantly increased adiposity at 18 months (Dogra et al., 2015). Specifically, an unsupervised clustering approach based on bacterial 16S rRNA sequence data sets revealed three clusters of fecal microbiota configurations. The number of samples that binned into one of these clusters (cluster 3), which is characterized by high levels of Bifidobacteria and Collinsella and low levels of Streptococcus and Enterobacteriaceae, increased with age. A faster time to achieving a cluster 3 configuration was associated with significantly greater adiposity measured at age 18 months. Given the rapid rate of change in eating practices and incidence of childhood obesity, longitudinal studies of this type are timely. They should
be strategically applied to populations representing different manifestations of these economic, anthropologic, and epidemiologic transitions and accompanied by comprehensive, quantitative assessments of food consumption during the pre-weaning, weaning, and postweaning periods.

Obesity is associated with reduced organismal and genetic diversity in the gut microbiota/microbiome of adults (Turnbaugh et al., 2009b and Le Chatelier et al., 2013). Transplantation of intact fecal microbiota samples, or derived culture collections, from adult twins stably discordant for obesity into germ-free mice transmitted the donors’ discordant adiposity phenotypes, as well as obesity-associated metabolic dysfunction (Ridaura et al., 2013). Co-housing mice just after they received the obese donor’s (Ob) microbiota with mice just after they received the lean cotwin’s (Ln) microbiota, before their discordant adiposity/metabolic phenotypes became evident, prevented development of obesity and metabolic abnormalities in the Ob cagemate. This prevention was associated with unidirectional invasion of bacteria from the Ln cagemate’s gut community to the Ob cagemate’s microbiota. Invasion was diet dependent, occurring in mice fed a human diet formulated to reflect the lower third of saturated fat and upper third of fruit and vegetable consumption in the USA, but not when animals received an unhealthy diet representing the upper third of saturated fat and lower third of fruit and vegetable consumption (Ridaura et al., 2013). These results illustrate how niches can be filled in the Ob microbiota by Ln-derived bacterial taxa to prevent disease and how important diet is to the installation of these health-promoting strains. The results raise important questions about the origins of the reduced bacterial diversity observed in Ob microbiota.

**Impact of Antibiotics**

One active area of investigation is the role of frequent consumption of broad-spectrum antibiotics in determining the diversity and functional features of the developing microbiota. Studies in conventionally raised mice treated with low-dose penicillin from birth to 4, 8, or 28 weeks of age revealed that early and brief exposure was sufficient to produce durable changes in body composition (Cox et al., 2014). Practical issues (in many parts of the world, antibiotic consumption
in children is pervasive and poorly documented), ethical considerations, and the identification of suitable controls all confound the design of human studies that would seek to determine the effects of antibiotic administration on the developmental biology of the human infant gut microbiota and growth. In principle, pre-clinical tests that administer various classes of antibiotics in varying doses—together with representative human diets to gnotobiotic mice harboring transplanted microbiota from infants and children living in various parts of the world—followed by transplantation of their antibiotic-treated microbiota to a next generation of (antibiotic-free) gnotobiotic recipients, would provide one way to explore these questions.

**Affordable Nutritious Foods: Societal Implications and Challenges**

An imbalance of carbohydrate, fat, and protein consumption, food insecurity, and changing diets in low-income countries brought about by globalization, increases in food prices at the point of retail, and a global protein supply that needs to double by 2050 are some of the drivers for developing new types of affordable nutritious foods that are culturally acceptable, suitable for storage, and distributable given current and envisioned future infrastructure. A sustainable economic model in which local economies benefit from producing and/or distributing foods is also required to ensure long-term supplies. Moreover, there is a paucity of generally accepted metrics for defining foods that provide optimal nutrition at affordable cost (e.g., see the “nutrient-rich foods index” developed based on FDA recommendations; Drewnowski, 2010).

We propose that the gut microbiota provides a parameter that needs to be considered when developing nutrition options and that the type of preclinical gnotobiotic models described above will be useful for testing and defining dietary parameters. Studies with mice and other species provide means for characterizing interactions between food ingredients (at different levels of ingredient resolution and including culturally relevant spices and sweeteners), their methods of preparation and preservation, the gut microbiota of various consumer populations, and human metabolic, immunologic, and other physiologic features. These research platforms offer the promise of yielding next-generation foods designed to be satiating, delicious, nutritious, and able to manipulate
microbiota and host properties in ways that promote healthy growth and wellness. However, fulfilling this promise demands a holistic view of the nexus of human gut microbial ecology research, agricultural practices, food production, evolving consumer tastes in an era of rapid globalization, envisioned commercialization strategies, current regulatory structures/practices, ethical issues, and public education. For example, there is a need to more thoroughly and rapidly characterize, through readily searchable, accessible, well-annotated databases, emerging food consumption patterns in countries representing different cultural traditions, stages of economic development, and land/water resources. At the commercial level, there is an opportunity to define and differentiate foods based on their effects on different consumer populations with distinct biological phenotypes and with different gut microbial community configurations. There is an accompanying need to frame intellectual property laws in ways that provide appropriate incentives for private investment while protecting the public good.

To effectively and responsibly apply this knowledge in ways that benefit society, there is a need to work with government agencies to provide efficient and sensible regulatory schemes. These regulatory frameworks vary between nations and are evolving. Currently, the US Food and Drug Administration (FDA) defines “medical foods” as foods that make medical claims. A “dietary supplement” is a product intended for ingestion that contains a dietary ingredient designed to add further nutritional value to a diet. Dietary supplements can only contain ingredients that are “generally regarded as safe” (GRAS) or approved as food additives by the FDA after filing a “new dietary ingredient” (NDI) notification with full description of the ingredient and product in which it will be marketed, the basis for the manufacturer’s conclusion that it is an NDI, recommended use and proposed labeling, plus a history of its use and evidence of its safety to support the proposed use. Probiotics have been defined in various ways, including “live microorganisms that, when administered in adequate amounts, confer a health benefit on the host” (Joint FAO/WHO Expert Consultation on Evaluation of Health and Nutritional Properties of Probiotics, 2001), whereas prebiotics have been considered to be “a selectively fermented ingredient that allows specific changes both in the composition and/or activity of the gastrointestinal microbiota that confer benefits upon
host well-being and health” (Roberfroid, 2007). Synbiotics are combinations of prebiotics and probiotics. Regulation of prebiotics, probiotics, and synbiotics remains a work in progress, although any health claims they make will likely require a clinical development pathway that is the same as that employed for biologics.

**Opening the Public Discussion**

For public acceptance and societal benefit, a thoughtful proactive, science-based, educational outreach is needed with an understandable vocabulary tailored to targeted consumer populations and respectful of their cultural traditions. The goal would be to objectively describe the extent to which the nutritional value of food is related to a consumer’s microbiota and how food ingredients, food choices, and the microbiota are connected to health benefits.

We suggest that one way of framing a public discussion regarding the impact of human gut microbiome research on the nexus of food, agriculture, and nutrition is to divide it into three “sectors”: science and technology, ethics, and policy and governance.

**Science and Technology**

Ongoing and new studies will help to define (1) methods for selection and production of new food sources, (2) design of new foods/diets, (3) definitions of nutritional value and benefit and metrics for differentiation of foods, and (4) the role of the gut microbiota in determining nutritional status in pregnant women, infants and children, and adults throughout the course of their lives.

**Ethics**

The impact of gut microbiota research extends beyond conceptions of health to human rights. Key issues include (1) concepts of self and ownership of microbes and the shaping of these views by cultural, religious, socio-economic, educational, and political factors; (2) use of a person’s microbes to improve nutritional status within and beyond family, community, and country; (3) strategies for responsible stewardship of our (human) microbial resources; and (4) personal, familial,
and societal impact (and shared benefit) of methods envisioned to promote intergenerational trans-
mersion of beneficial microbes and to effect durable repair of defective gut microbial community
development early in life or functional restoration later in life.

Policy and Governance

Advances in gut microbiota research will have long-term impact on regulatory and other govern-
mental policies and agencies as they relate to agriculture, food, and nutritional health. These ef-
effects include (1) definitions of food safety, including the products of microbial biotransformation
of food ingredients; (2) definitions of nutritional benefit within and outside of the context of specific
human health claims; (3) laws concerning ownership of microbial strains and their distribution
within and across national borders (for example, in October 2014, the Convention on Biological
Diversity/Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of
Benefits from their Utilization entered into international force “stringent requirements for prior
informed consent and benefit sharing for research and commercial activities involving genetic
resources from plants, animals, and microorganisms” [http://www.cbd.int/abs/]); (4) laws concern-
ing intellectual property related to microbes, microbial consortia, and the products of microbial
interactions with food ingredients, including diagnostics and therapeutics; (5) policies related to
standards of manufacture, purity, and composition of probiotics and synbiotics; and (6) incentives
for linking plans for food production and distribution with gut microbiota health. A key challenge
is how to construe (1)–(6) in the context of a reference set of “representative” countries.

Closing Thoughts

Given the intricate links between first foods and long-term human health, ensuring availability of
appropriate food sources is of high priority. Because undernutrition is such a widespread afflict-
tion, it is critical to consider how to categorize the targeted populations, the cost and economic
sustainability, the efficacy (effect size and durability), and the cultural acceptability of various
therapeutic or preventative approaches, as well as the generalizability of both food-based and
microbial interventions to large populations within and across national/societal boundaries. One way of conceptualizing this complex set of challenges for treatment and prevention is to place, on one end of the spectrum of undernutrition, children with already manifest SAM and significant microbiota immaturity who could be treated with locally produced, readily and reproducibly manufactured, affordable and safe, culturally acceptable next-generation RUTFs, with or without microbial interventions of the type described above. Moving along this continuum, another group would consist of individuals who manifest growth faltering (stunting) in the first 1,000 days after conception, where the envisioned targets for interventions are pregnant and lactating women and their infants. At the other end of the continuum is a third group that are the targets of locally produced, consumer-focused, affordable nutrition products designed to improve dietary quality and increase the diversity of food choices.

Looking back over 800 million years of metazoan evolution, we appreciate more now than ever before the splendid innovation of having a gut that assembles microbial resources that enable efficient utilization of available nutrients (McFall-Ngai et al., 2013). We, humans, are now in a position to not only understand but to deliberately influence this process of microbial community acquisition in order to ensure its optimal execution. The challenges we face in designing and improving food systems and nutritional health are great and pressing. Hopefully, our gut instinct will be to honor and harness the intimate interrelationship between foods and “our” microbes in an attempt to address this challenge now and throughout the course of this defining century for our species and planet.
Acknowledgments

Work cited from the authors’ labs was supported in part by grants from the Bill & Melinda Gates Foundation (BMGF) and the NIH (DK078669, DK30292, and DK70977). We thank our colleagues in the BMGF-sponsored Breast Milk, Gut Microbiome, and Immunity Project, Tahmeed Ahmed and other members of our collaboration with the International Center for Diarrheal Disease Research, Bangladesh (ICDDR,B), Michael Barratt, Jeanette Gehrig, Siddarth Venkatesh and other members of our laboratories, plus William Petri and Andrew Serazin for their insights and inspiration. J.I.G. is a co-founder of Matatu, Inc., a company characterizing the role of diet-by-microbiota interactions in animal health. D.A.M. is a co-founder and President of Evolve Biosystems, Inc., a company focused on diet-based manipulation of the gut microbiota.
References


MAL-ED Network Investigators (2014). The MAL-ED study: a multinational and multidisciplinary approach to understand the relationship between enteric pathogens, malnutrition, gut


34


**Figure Legends**

**Figure 1. Developing metrics for describing gut microbial community development.** (a) Bacterial taxa that discriminate different stages of development were identified by a machine learning-based (Random Forests) regression of 16S rRNA datasets produced from monthly fecal samples collected from anthropometrically healthy infants and children living in an urban slum in Dhaka, Bangladesh during their first two years of postnatal life to their respective chronologic ages at the time of sample collection (Subramanian et. al, 2014). Shown are depictions of the typical distributions of these age-discriminatory taxa across the population. Taxa were selected based on their relative importance to the accuracy of the Random Forests model using a permutation-based ‘feature importance’. (b) The most discriminatory taxa, as defined by their feature importance, were used as inputs into a sparse 24-taxon model whose output (‘microbiota age’) is a microbiota-based prediction of the chronologic age of a healthy child. The plot on the left of the panel shows microbiota age against chronologic age of healthy children used as a training set to fit the regression (each dot is a fecal sample from an individual child). The plot on the right of the panel shows application of the sparse model to a validation set composed of a different group of children living in the same location that were not used to train the model. Applying the model to a separate validation set controls for over-fitting of the model to the training set, and ensures its wider usability. (c) Two metrics of microbiota maturation based on application of the model to two separate validation sets of singletons and a separate study of Bangladeshi twins/triplets. ‘Relative microbiota maturity’ is the deviation, in months, from a smooth-spline fit of microbiota age values with respect to chronologic age, fitted using the validation datasets (see black dashed curve). The red dot represents a fecal sample collected from a focus child that is 11 months below the spline fit, indicating negative relative microbiota maturity (i.e., an immature microbiota). A microbiota-for-age Z score (MAZ) is computed by dividing the difference between the focal child’s microbiota age and the median microbiota age of healthy controls in the same monthly chronologic age bin over the standard deviation within the same age bin. The median and standard deviation of each bin are computed using the validation datasets. The distribution of microbiota maturity and MAZ scores in birth-co-
hort studies have been studied using linear mixed models that take into account random variation specific to each serially-sampled child and family while estimating the fixed variation attributable to a factor observed across different children (e.g., diarrheal episodes) (Subramanian et al., 2014). Note that using Random Forests to study microbiota maturation is advantageous because of its non-parametric assumptions and utility in the context of high dimensional datasets (large numbers of predictors). Nonetheless, it is one of several methods that can be useful. For example, the rank-order Spearman correlation metric has been applied to infant microbiome datasets to detect monotonic relationships between microbiome-encoded functions/bacterial taxa and postnatal age (Yatsunenko et al, 2012).

**Figure 2. Integration of existing clinical observational and interventional studies into gnotobiotic mouse models to identify interactions between the gut microbiota, food, and host biology.** The discovery process depicted by the left circle illustrates how gnotobiotic animal models colonized with human donor microbiota and fed human diets can lead to a greater understanding of how diet-by-microbiota interactions are causally related to healthy growth and to phenotypes associated with undernutrition: e.g., immune system development, brain development and host and microbial community metabolism. New surrogate- or mechanism-based biomarkers of nutritional state emanating from these gnotobiotic models can be validated using biospecimens collected from the donors used to construct these gnotobiotic models, as well as from other members of the study population. The discovery/development process depicted on the right illustrates how dietary and microbial ‘leads’ can be tested in the context of humanized gnotobiotic animals to assess how they modulate biological processes already known, discovered or postulated to be involved in healthy growth and/or the pathogenesis of undernutrition. The downward pointing arrow in the middle of the figure points to next steps in clinical translation. See the main text for a discussion of the regulatory, ethical, societal and commercial implications of these efforts. Abbreviation: IND, Investigational New Drug.
Figure 3. Co-variation in gut microbiota assembly/maturation, dietary patterns and other facets of human postnatal development. (a) Illustration of the rate of change occurring in gut microbiota structure of both mother and child. Note that infant variation curves are known from both longitudinal and cross-sectional study designs (Yatsunenko et al., 2012, Subramanian et al., 2014). In the case of mothers, the curve is interpolated based on studies of pregnant Finnish mothers prior to delivery (Koren et al., 2012) and Bangladeshi mothers following parturition (Subramanian et al., 2014). (b) The food consumption pattern shown is at a population level and does not depict the great deal of temporal variation observed in food consumption patterns within a given child. Depicting the fractional contribution of each food to the consumption patterns of children in Bangladesh underscores how dietary changes occur simultaneously (lowering of breast-milk and increase in legumes and cow’s milk) and not in an orderly fashion (small fluctuations from month-to-month; re-entry and dropout of certain foods). It also underscores the challenge encountered in ascertaining how food and the microbiota interact to effect maturation of the community. (c) Major processes related to growth and how they vary in rate and magnitude over time. Curves are adapted from Bogin (1999). Note that the newborn brain represents 12% of body weight (a value six times greater than in adults). By the end of the first decade, the brain represents 6% of body weight and consumes twice the amount of glucose and 1.5 times the amount of oxygen as the adult brain. Approximately 30% of the glucose consumed by the infant brain is accounted for by aerobic glycolysis (versus 12% in adults) (Goyal et al., 2014). The dramatic changes in brain metabolism that occur over the first two decades of life coincide with the initial proliferation and then pruning of synapses to adult levels. Central questions that need to be addressed in this area include the biological effects of the gut microbial community on neurogenesis, synaptic connectivity, gliogenesis and glial-neuron interactions, neural circuit function and higher cognitive processes in the context of healthy growth versus undernutrition, and whether/how the gut-brain axis operates to influence/regulate other aspects of host physiology, metabolism and immunity in the infant/child. Moreover, if persistent immaturity of the gut microbiota is causally related to undernutrition and its long-term sequelae, including neurodevelopmental abnormalities, does durable repair of this im-
maturity require that nutritional interventions be administered earlier before disease becomes fully manifest (and the microbial ecosystem is so perturbed that restoration becomes very difficult)? Do nutritional interventions need to be applied for more sustained periods of time? Do new types of therapeutic foods need to be developed or is a microbial intervention also needed?
Figures

Figure 1.
Figure 2.
Figure 3.
Chapter 2

Gut bacteria that rescue growth impairments transmitted by immature microbiota from undernourished children
Chapter 2

Gut bacteria that rescue growth impairments transmitted by immature microbiota from undernourished children

Laura V. Blanton1,2, Mark R. Charbonneau1,2, Tarek Salih1,2, Michael J. Barratt1,2, Siddarth Venkatesh1,2, Olga Ilkaveya3,4, Sathish Subramanian1,2, Mark J. Manary7,8, Indi Trehan7,9, Josh M. Jorgensen10, Yuemei Fan11, Bernard Henrissat12,13, Semen A. Leyn14, Dmitry A. Rodionov14,15, Andrei L. Osterman15, Kenneth M. Maleta8, Christopher B. Newgard3,4,5,6, Per Ashorn11,16, Kathryn G. Dewey10, and Jeffrey Gordon1,2

1Center for Genome Sciences and Systems Biology, and 2Center for Gut Microbiome and Nutrition Research, Washington University School of Medicine, St. Louis, MO 63108 USA. 3Sarah W. Stedman Nutrition and Metabolism Center, 4Duke Molecular Physiology Institute, 5Department of Pharmacology and Cancer Biology, and 6Department of Medicine, Duke University Medical Center, Durham, NC, USA. 7Department of Pediatrics, Washington University School of Medicine, St. Louis, MO 63110, USA. 8School of Public Health and Family Medicine, College of Medicine, University of Malawi, Chichiri, Blantyre 3, Malawi. 9Department of Paediatrics and Child Health, College of Medicine, University of Malawi, Chichiri, Blantyre 3, Malawi. 10Department of Nutrition, and Program in International and Community Nutrition, University of California, Davis, Davis, CA 95616, USA. 11Department for International Health, University of Tampere School of Medicine, Tampere 33014, Finland. 12Architecture et Fonction des Macromolécules Biologiques, Centre National de la Recherche Scientifique & Aix-Marseille Université 13288 Marseille cedex 9, France. 13Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia. 14A.A.Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 127994, Russia. 15Infectious and Inflammatory Disease Center, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, 92037 CA USA. 16Department of Pediatrics, Tampere University Hospital, Tampere 33521, Finland

Correspondence to: jgordon@wustl.edu


**Abstract**

Children with undernutrition exhibit impaired normal gut microbial community development. Transplanting fecal microbiota from 6- and 18-month-old Malawian donors manifesting healthy growth or varying degrees of undernutrition into recently weaned germ-free mice fed a representative Malawian diet revealed that immature microbiota transmit impaired growth phenotypes, with the representation of several age-discriminatory taxa in the microbiota of recipient animals correlating with lean body mass gain, metabolic phenotypes in liver, muscle, and brain, plus bone morphology. Co-housing mice shortly after they received microbiota from healthy (H) or severely stunted/underweight (Un) infants demonstrated that invasion of age-/growth-discriminatory taxa from H to Un cagemates’ microbiota ameliorates growth faltering. Adding two invasive species, *Ruminococcus gnavus* and *Clostridium symbiosum*, to the Un microbiota also produced amelioration of growth and metabolic abnormalities. These results indicate that gut microbiota immaturity is causally related to undernutrition, and reveal bacterial strains that are potential therapeutic targets and agents.

**One sentence summary:** Gnotobiotic mice provide evidence that delayed development of the human gut microbiota is not only associated with childhood undernutrition but causally related to it.

**Main Text**

Undernutrition is a leading cause of infant and childhood mortality worldwide (1-5). The mechanisms that underlie disease pathogenesis, including the persistent abnormalities in growth, immune function, and cognition, remain obscure. There is accumulating evidence that childhood undernutrition is not due to food insecurity alone, but also results from a combination of factors including diets with low nutrient density/bioavailability, pathogen burden and gut mucosal barrier dysfunction (6-10).

The World Health Organization (WHO) has developed anthropometric standards to define nutritional status [weight-for-height Z-Score (WHZ), weight-for-age Z-Score (WAZ), and height-
for-age Z-Score (HAZ)] (11, 12). A recent study of infants and children with healthy growth phenotypes living in Mirpur, an urban slum in Dhaka, Bangladesh, involved monthly fecal collection from birth through the end of the second year of life; 16S rRNA analysis of bacterial membership in their gut microbiota and application of a machine learning method [Random Forests (6)] yielded a model composed of 24 ‘age-discriminatory’ strains whose changes in relative abundance over time defined a program of normal ‘maturation’ of the microbiota across biologically unrelated individuals (6). This model served as the basis for computing two related metrics, ‘relative microbiota maturity’ and microbiota-for-age Z-score (MAZ), that significantly correlated with the chronological age of children with healthy growth phenotypes. Applying these metrics revealed that children living in Mirpur with moderate acute malnutrition (MAM) or severe acute malnutrition (SAM) have significant gut microbiota immaturity compared to age-matched children from the same locale, with the degree of immaturity being greater in SAM than MAM (13). Moreover, treatment of children with SAM with either one of two ready-to-use therapeutic foods (RUTFs) produced incomplete and only transient improvement in this immaturity, and no improvement in their HAZ scores (13).

These findings raise several key questions. To what extent are these age-discriminatory strains also indicative of the normal development of the gut microbiota of infants and children representing other populations living in geographically distinct areas of the world with different cultural traditions and diets? Are these age-discriminatory strains simply biomarkers of gut microbiota development or are they critical mediators of healthy growth? If they are mediators of healthy growth, can their introduction into immature microbiota prevent disease?

A Malawian model of gut microbiota development

To address these questions, we first developed a Random Forests-based model of gut microbial community development from a cohort of Malawian twins concordant for healthy growth and then applied the model to serially sampled Malawian infants/children in order to define the degree of
correlation between nutritional status and microbiota maturation and ascertain whether an MAZ score could predict future growth performance.

We previously followed a cohort of 317 twin pairs and three sets of triplets, living in five rural villages in southern Malawi from birth to 36 months of age, with periodic sampling of their fecal microbiota (14). To define healthy microbiota development in this population, we sequenced PCR amplicons generated from variable region 4 (V4) of bacterial 16S rRNA genes represented in 238 fecal samples collected from 27 twin pairs and two sets of triplets whose serial anthropometric measurements were indicative of consistently healthy growth [WHZ 0.09±0.93; fecal collection over an age range of 0.6 to 33.5 months; 3.96±1.6 samples (mean±SD)/individual; see Tables S1A and S2A for microbiota donor characteristics and a summary of sequencing datasets]. V4-16S rRNA reads with ≥97% nucleotide sequence identity were grouped into operational taxonomic units (97%ID OTUs) and taxonomic assignments were made as described in Materials and Methods.

We modeled microbiota maturation in these healthy Malawian infants/children by regressing OTUs against chronological age using Random Forests [training subset, 121 fecal samples from 14 twin pairs and 1 set of triplets; test subset, 117 samples from 13 twin pairs and 1 set of triplets; see Fig. S1 which shows that the age distribution of infants/children at the time of sample collection in each subset are not significantly different (p=0.51, Kolmogorov-Smirov test)]. Application of the training subset-based model to the test subset yielded a positive and significant correlation between host chronological age and predicted microbiota age ($r^2=0.78$, p<0.0001). Random Forests assigns a mean squared error (MSE), or feature importance score, to each OTU that indicates the extent to which each OTU contributes to the accuracy of the model. The 25 most age-discriminatory taxa, ranked by MSE, yielded a sparse model that predicted microbiota age and accounted for ~80% of the observed variance in the healthy cohort (Fig. 1A-C, see Table S3A for a complete list of OTUs and MSE values). Remarkably, multiple OTUs from the sparse Malawian model had ≥97% sequence identity to OTUs in the sparse Bangladeshi model (Fig. 1A, Table S3B).
We used the Malawian Random Forests-derived model to analyze the relationship between microbiota maturity and growth/nutritional status in children enrolled in a randomized nutritional interventional trial at 6, 12, and 18 months of age (iLiNS-DYAD-M) (15) that took place in the Mangochi district of southern Malawi (see Materials and Methods). There was a significant positive correlation between MAZ and WHZ (ρ=0.1664, p=0.0073, Spearman’s rank correlation) and MAZ and WAZ (ρ=0.1715, p=0.0056) but not HAZ (ρ=0.1022, p=0.1 for HAZ) at 18 months of age in the group of 259 iLiNS-DYAD-M enrollees who had fewer than three reported days of antibiotic consumption during the previous 12 months (Table S2B). Remarkably, there was a significant correlation between MAZ at 12 months and growth at 18 months [Spearman’s rank correlation ρ=0.1406 (p=0.02) for WHZ and ρ=0.1373 (p=0.02) for WAZ] [Note that ρ=0.1184 (p=0.05) for HAZ]. These results suggest that MAZ may be useful for predicting future (ponderal) growth. Further studies and analyses are required to discern the effects of a number of variables, such as duration of prior antibiotic use, enteropathogen load, number of diarrheal days, geography, and various nutritional interventions, on the relationship between MAZ measurements at various postnatal ages, anthropometry and other metrics of healthy growth (e.g., cognitive testing and immunization responses).

Identification of age-discriminatory taxa that are also growth-discriminatory

To establish a causal relationship between microbiota maturity and growth, we selected fecal samples from 19 Malawian infants representing either healthy or undernourished growth phenotypes for transplantation into young, actively growing germ-free C57Bl/6J mice. Of the 19 donor samples, nine were from 6-month-old infants [four classified as healthy with WHZ 1.45±0.57 (mean±SD), WAZ 1.75±0.56, and HAZ 1.27±0.45, and five moderately or severely underweight and stunted (WHZ -1.27±0.49, WAZ -3.90±1.82 and HAZ -4.36 ±2.07)]. Ten samples were from 18-month-old children, four with healthy weights (WHZ 1.44±0.08, WAZ -0.3±0.57 and HAZ -2.84±0.99; note that all 18-month-old donors were members of twin pairs where HAZ scores are typically lower than in singletons), and six who were moderately or severely underweight
and stunted (WHZ -1.75±0.17, WAZ -2.88±0.42 and HAZ -3.28±0.82) (see Table S4A for MAZ metrics). Fecal samples from the 18-month-old children were from the Malawi twin cohort while the 6-month-old microbiota donors were members of the iLiNS-DYAD-M study who had not yet received a nutritional supplement.

Each microbiota sample was transplanted into a separate group of 5-week old male germ-free mice (n=5 animals/sample). Three days before transplantation, all mice were switched onto a sterile (irradiated) Malawian diet formulated based on the results of a dietary survey of the complementary feeding practices of 9-month-old Malawian children enrolled in the iLiNS-DOSE study (#NCT00945698) that took place in the Mangochi district. We selected eight ingredients to produce a cooked, representative Malawian diet consumed by children (‘M8’; see Materials and Methods); its micro- and macronutrient content does not fulfill the needs of humans or mice (see Table S5A,B for a list of ingredients and the results of a direct nutritional analysis).

Following a single gavage of the donor microbiota, recipient mice were followed for 4-5 weeks (see Fig. 2A for experimental design). Fecal samples were collected weekly for bacterial V4-16S rRNA analysis. Growth was monitored by serial measurements of total body weight and body composition [lean mass and fat mass as defined by quantitative magnetic resonance (qMR)] while femurs were removed for micro-computed tomographic (micro-CT) characterization of bone morphology.

Mice harboring donor microbiota from healthy infants gained significantly more weight (p<0.0001; 2-way ANOVA) and lean body mass (p=0.0001; 2-way ANOVA) than mice colonized with microbiota from undernourished donors (Tables S6A,B), yet there was no significant difference in food consumption between groups (p>0.05, Student’s t-test, Table S7). 16S rRNA sequencing of fecal samples obtained from recipient mice revealed variable transplantation efficiency. Of the 19 donor samples, eight produced transplantation efficiencies of ≥50% [i.e., ≥50% of the 97%ID OTUs represented in the input community, weighted to relative abundance, were detectable in recipient mice throughout the course of the experiment; see Table S4B. Note that bacterial
viability, as quantified by the number of colony forming units 24 hours after plating onto rich medium, was significantly greater (p<3.2x10⁻⁶; Student’s t-test) in the case of fecal samples from the twins, which had been flash frozen in liquid nitrogen within 5-10 min after they were produced. In iLiNS-DYAD-M, the time to freezing was often several hours]. When we restricted our analysis to the eight donor samples producing >50% transplantation efficiency [four six-month-old singleton infants from iLiNS-DYAD-M (two healthy, two undernourished) and four from 18-month-old cotwins from the Malawian twin study (one healthy, three undernourished)] the discordant growth phenotypes between recipients of healthy versus undernourished microbiota became even more pronounced (weight gain, p<0.0001; lean mass gain, p=0.0005; 2-way ANOVA). Importantly, there was no significant difference in fat mass gain (p=0.78, 2-way ANOVA, Table S6B) or food consumption (p>0.05, Student’s t-test; Fig. 2B,C; Tables S6A,B, and S7) between the groups.

For both the 6-month and 18-month age bins, mice colonized with healthy donor microbiota gained significantly more total body weight and lean mass than those colonized with undernourished donor microbiota (weight for recipients of the 6- and 18-month donor microbiota, p=0.0003 and p=0.0043, respectively; lean body mass, p=0.03 and p=0.0013, respectively; 2-way ANOVA; Table S6A,B). Intriguingly, recipients of the healthy and undernourished 6-month-old donors’ fecal samples grew more than recipients of the corresponding 18-month-old healthy and undernourished donor microbiota (p<0.0001 for healthy; p<0.0001 for undernourished, based on lean mass gain; 2-way ANOVA).

Growth over the course of the 5-week experiment in the 19 different groups of recipient mice ranged from 105% to 152% of starting weight (averaged per group; Table S6A). There was no significant relationship between growth phenotypes and bacterial diversity in the fecal microbiota of recipient animals (Spearman’s rank correlation ρ=0.1, p=0.3275 for Shannon Index versus lean body mass gain at day 28 following gavage of the donors’ microbiota). We applied Random Forests to regress the growth phenotypes of recipient gnotobiotic mice against 97%ID OTUs identified in their fecal microbiota. Two models were generated, one based on weight gain, the other based on lean body mass gain (see Fig. 2D and Fig. S2, respectively, plus Table S8A,B). The
weight and lean body mass gain models explained ~66% and ~58% of the observed phenotypic variance, respectively (p<0.0001, permutation test; 999 permutations). Remarkably, two of the growth discriminatory species represented in the weight and lean mass gain Random Forests-based models, *Bifidobacterium longum* and *Faecalibacterium prausnitzii*, have high feature importance scores in the sparse 25 OTU Malawian and sparse 24 OTU Bangladeshi Random Forests-derived models of age-discriminatory taxa [ranked 1st and 8th, respectively, in the Malawian model shown in Fig. 1A, and 5th and 1st in the Bangladeshi model (13)]. *B. longum* is an early colonizer while *F. prausnitzii* becomes more prominent at later times in the healthy Malawian and Bangladeshi populations used to create these models (see Fig. 1B, (13)).

The relative abundances of 13 and seven of the OTUs with significant positive monotonic correlations to body weight and lean mass gain, respectively, also had significant correlations to chronological age in concordant healthy Malawian twins/triplets, including three OTUs for *F. prausnitzii* (p<0.0001; see Fig. 2D, Fig. S2, plus Table S9 for a complete list of OTUs with their Spearman’s rank correlation ρ-values). Moreover, 15 and 11 OTUs with significant positive correlations to weight and lean mass gain, respectively, were also significantly correlated with chronological age for the 259 infants/children from iLiNS-DYAD-M, including the same three OTUs for *F. prausnitzii* that were significantly correlated with age in the cohort of healthy Malawian twins and triplets (p<0.0001 for all correlations; Spearman’s rank correlation; see Table S9).

**The relationship between age- and growth-discriminatory taxa and femoral phenotypes**

We used micro-CT to assay morphological parameters in cancellous and cortical regions of femurs obtained at the time of sacrifice from mice harboring the eight donor microbiota that exhibited >50% transplantation efficiency. There were strong trends to higher femoral cortical BV/TV and volumetric bone mineral density (vBMD) in recipients of undernourished donor gut microbial communities (p=0.05, p=0.07, respectively, Mann-Whitney test). Differences between mice harboring microbiota from 6- and 18-month-old donors were evident in cancellous rather than cortical bone, regardless of donor nutritional status; mice colonized with 6-month-old donor communities...
had significantly higher BMD, BV/TV, trabecular connectivity and number, and significantly lower trabecular spacing irrespective of their nutritional status (Fig. S3, Mann-Whitney test, p<0.001 for all bone metrics).

We extended our Random Forests and ranked Spearman correlation analyses to identify OTUs that discriminate these femoral bone phenotypes (Fig. S4, Table S10). Six of the OTUs represented in both of the growth discriminatory Random Forests models, including one assigned to \textit{B. longum}, were also represented among the top 20 most discriminatory features in at least three of the five Random Forests-based models of bone metrics (Fig. S4). The \textit{B. longum} OTU (72820) is also a member of the sparse Malawian Random Forests-derived model of age-discriminatory taxa. Moreover, an \textit{F. prausnitzii} OTU (265871) was ranked within the top 40 taxa in all of the Random Forests models based on bone metrics as well as the lean mass gain model. Together, these results provide evidence for microbiota-dependent regulation of bone morphology (17) and that these effects are influenced by the age and nutritional status of the donor.

\textbf{Effects of gut microbiota immaturity on host metabolism}

We selected gnotobiotic recipients of two 6-month-old donor microbiota that transmitted the most discordant growth phenotypes in the initial screen of 19 microbiota to characterize the effects of microbiota immaturity on host metabolism (Fig. S5A). One was a healthy infant from Mangochi (HAZ 1.49, WAZ 1.43, WHZ 0.9) with age appropriate microbiota maturity (microbiota age of 6.7 months); the other was a severely stunted and underweight infant (HAZ -3.35, WAZ -3.08, WHZ -0.79) from Malindi (located 20km from Mangochi) with microbiota immaturity (microbiota age of 4.6 months). The configurations of these transplanted microbiota were quite distinct: mice colonized with the healthy donor community were dominated by \textit{F. prausnitzii} (33±19\% relative abundance) while mice colonized with the undernourished donor microbiota were dominated by \textit{Clostridium neonatale} (37±11\% relative abundance) (n=5 mice/donor microbiota). Of the 27 species (66 97\%ID OTUs) present in recipients of the healthy donor community and the 22 spe-
cies (33 OTUs) present in recipients of the undernourished immature donor microbiota, 13 OTUs, representing 13 different species were present in both (see Fig. S5B and Table S11).

In follow-up confirmatory experiments, mice received the same donor microbiota; serum, liver, gastrocnemius muscle, and brain were obtained at the time of euthanasia three weeks after gavage of these microbial communities, and analyzed by targeted mass spectrometry (n=5-6 mice/treatment group). The two donors’ microbiota conferred distinct metabolic phenotypes with the major differences involving fatty acid and amino acid metabolism; mice colonized with the healthy donor’s microbiota had significantly higher concentrations of long chain acylcarnitines in their serum and liver and significantly decreased concentrations of a number of amino acids in their gastrocnemius muscle (and cecal contents) (p<0.05; Student’s t-test; Fig. 3; Table S12). These data support the notion that the two microbial communities may regulate the manner in which amino acids are utilized, such that mice colonized with the healthy donor microbiota funnel amino acids into new protein synthesis and lean body mass, whereas those colonized with the undernourished donor microbiota preserve the amino acid pools for oxidation and energy production. Consistent with this idea, increased levels of the branched-chain amino acids Val and Leu/Ile in the gastrocnemius muscle of mice colonized with undernourished donor microbiota were accompanied by an increase in a metabolite produced by their oxidation, C3 (propionyl) acylcarnitine (Table S12).

These differences in metabolic profiles extended to the brain (Fig. S6; Table S12). Echoing results from serum and liver, levels of long chain acylcarnitines were significantly elevated in the cerebral cortex of mice harboring the healthy donor microbiota compared to the undernourished donor microbiota (p<0.05; Student’s t-test). Amino acid metabolism was also impacted with concentrations of histidine, proline, glycine, arginine, and citrulline being significantly higher in the brains of mice with the healthy donor’s microbiota. Additionally, there was evidence of perturbations in the TCA cycle, with citrate levels being significantly higher in animals harboring the undernourished infant’s gut community. The resource demands of the brain in children are prodigious: e.g., at the end of the first decade of postnatal life, a child’s brain consumes twice the amount of glucose and 1.5 times the amount of oxygen per gram of tissue as an adult’s brain, and
accounts for up to 50% of the body’s total basal metabolic rate (16-18). During this period there is an explosive growth of synapses. The differences in brain metabolism that are evident within three weeks after transplantation of undernourished versus healthy infant microbiota into young gnotobiotic mice fed the same prototypic, nutrient deficient Malawian diet provide preclinical evidence that brain development needs to be viewed in the context of the developmental biology of the microbiota and its capacity to metabolize components of the diets consumed by children.

There were strong correlations between the relative abundances of bacterial species in the fecal microbiota of gnotobiotic mice and serum C10, C12, and C14 acylcarnitine levels, as well as levels of amino acids in gastrocnemius muscle. Species enriched in the healthy community that significantly correlated to serum acylcarnitines and amino acids included taxa that were also weight- and/or lean mass gain-discriminatory: *R. gnavus* (C10:2, C12, C12-OH/C10-DC acylcarnitines; valine, leucine/iso-leucine, and methionine), *Dorea formicigenerans* (C10, C10:1, C12, C12-OH/C10-DC , C14, C14:2 acylcarnitines; methionine), *F. prausnitzii* (C10-OH/C10-DC, C14 acylcarnitines, leucine/iso-leucine), and *Clostridium symbiosum* (C10:1, C12, C12-OH/C10-DC, C14 acylcarnitines; and 11 amino acids including valine, leucine/iso-leucine, and methionine) (p<0.05, see Figs. 2, S2, and S7A,B; plus Table S13 for a complete list with Spearman’s ranked correlation ρ-values). *B. longum* and *F. prausnitzii*, the two age- and growth-discriminatory species, were also correlated with hepatic levels of the TCA cycle intermediate citrate (p<0.05; see Fig. S7C, Table S13).

While these two donors’ microbiota transmitted discordant weight, lean mass gain and metabolic phenotypes, they did not transmit discordant bone morphometric phenotypes (p>0.05 on all metrics; Student’s t-test). This implies, perhaps not surprisingly, that ‘anthropometrically undernourished’ and ‘anthropometrically healthy’ donor microbiota do not transmit discrete sets of effects, but may produce overlapping phenotypes, each component of which may vary in its degree of discordance from donor to donor. A corollary is that different host phenotypes need to be assayed to obtain a more robust functional classification of ‘healthy’ versus ‘undernourished’ microbiota.
Repairing impaired growth phenotypes with age- and growth-discriminatory strains present in a healthy donor microbiota

We next determined whether age- and growth-discriminatory bacterial species were capable of invading the stunted/underweight microbiota and repairing the growth abnormalities that this microbial community transmitted to mice. Taking advantage of the coprophagic behavior of mice, we gavaged 5-week-old male germ-free C57BL/6J animals with either the healthy or stunted/underweight donor’s microbiota. Four days later, before phenotypic differences were apparent, we dually housed mice with the healthy (H) donor microbiota (H-H controls) and dually housed mice with the undernourished (Un) donor community (Un-Un controls) ($n=3$ cages of each control/experiment). In addition, H and Un mice were co-housed with one another (H CH-Un CH) ($n=6$ cages/experiment; $n=2$ independent experiments) (see Fig. 4A). All animals were fed the M8 diet beginning three days before colonization and throughout the course of the 3-week experiment. Animals were weighed twice a week while body composition and fecal samples were assayed once a week.

$H^{CH}$ and $Un^{CH}$ cagemates both gained significantly more lean mass compared to Un-Un controls (Fig. 4B; $Un^{CH}$ p=0.0121; $H^{CH}$ p=0.0447, Mann-Whitney test), but had no significant difference to lean mass gain in H-H controls. We proceeded to characterize invasion using a previously described approach (19) that employs Microbial SourceTracker to estimate the posterior probability that every species-level taxon or 97%ID OTU is derived from each of a set of source communities (7). The fecal microbiota of H-H or Un-Un controls sampled 4 days after gavage (just prior to co-housing) were used as source communities. The fecal communities belonging to each $H^{CH}$ and $Un^{CH}$ mouse were then traced to these sources. We defined the direction of invasion by calculating the log odds ratio of the probability of a H origin ($PH_i$) or an Un origin ($PUn_i$) for each species-level taxon or 97%ID OTU, $i$:

$$\log_2[PH_i/PUn_i]$$
A positive log odds ratio indicates that a species or 97%ID OTU is derived from an H source, while a negative log odds ratio indicates a Un source. An invasion score is calculated to quantify the success of invasion of each species or 97%ID OTU, $i$, into each co-housing group, $j$:

$$
\text{Invasion Score}_{ij} = \log_2 \left[ \frac{A_{ij}}{B_{ij}} \right]
$$

where $A_{ij}$ is the average relative abundance of taxon $i$ in all fecal samples collected from group $j$ after cohousing, and $B_{ij}$ is its relative abundance in all samples from that group prior to initiation of cohousing.

The mean of the distribution of invasion scores for Un\textsuperscript{CH} animals was significantly higher than that for Un-Un controls ($p \leq 0.0085$, Welch’s two-sample t-test). This was not the case for H\textsuperscript{CH} animals compared to H-H controls ($p > 0.05$), indicating that there was significant invasion of members of the H\textsuperscript{CH} microbiota into the microbiota of Un\textsuperscript{CH} cagemates, but not vice versa.

To quantify invasion further, we used the mean and standard deviation of the null distribution of invasion scores (defined as the scores from recipients of the H or Un microbiota that had never been co-housed with each other) to calculate a z-value and a Benjamini-Hochberg adjusted $p$-value for the invasion score of each species in H\textsuperscript{CH} and Un\textsuperscript{CH} mice. We defined a taxon as a successful invader if it (i) had a Benjamini-Hochberg adjusted $p \leq 0.05$, and (ii) had a relative abundance of $\leq 0.05\%$ before cohousing and $\geq 0.5\%$ in the fecal microbiota at the time of sacrifice. Fig. 4C and Table S14 provide information about the direction and success of invasion. We identified nine OTUs from the H\textsuperscript{CH} cagemate microbiota that consistently invaded the microbiota of cagemates with the stunted/underweight donor microbiota (total of 12 co-housed mice). These nine OTUs together comprised 58.7±14.2% (mean±SD) of the fecal microbiota of Un\textsuperscript{CH} animals (data averaged throughout the 17 day cohousing period) and 67.7±6.7% on the day of sacrifice. Based on the rank order of their invasion scores, the two most successful invaders were $F. praunitzii$, an age- and growth-discriminatory taxon in the Random Forests models and the most abundant OTU in the fecal microbiota of Un\textsuperscript{CH} mice at the conclusion of the co-housing experiment, and $R. gnarus$, a growth-discriminatory taxon. In contrast, only two OTUs from the stunted/underweight
microbial community successfully invaded the gut community of HCH mice; these two OTUs, one assigned to *Enterococcus* and the other to *Eubacterium limosum* represented 2.7±1.7% of the community after co-housing [note that the *Enterococcus* OTU 4316566 had a significant negative correlation to lean mass gain in the initial screen of 19 donor microbiota (ρ=-0.14, p=0.0065; Spearman’s rank correlation)].

Culturing invasive growth- and age-discriminatory taxa and characterizing their effects on growth

Acquisition of HCH-derived bacterial taxa in the microbiota of UnCH cagemates was accompanied by reductions in the relative abundances of 19 OTUs, six of which were below the limits of detection (<0.01%) by the end of the co-housing experiment. These observations raised two key questions. Which member or members of the group of invasive OTUs mediated the observed effects on growth phenotypes? To what extent was the reduction in other OTUs that coincided with invasion of these taxa responsible for improved growth in UnCH cagemates? Therefore, we initiated an effort to culture *F. prausnitzii* and *R. gnavus* from the fecal microbiota of healthy Malawian infants from the iLiNS-DYAD-M cohort. This culturing effort, which encompassed two iLiNS-DYAD-M donor microbiota samples (8243C and 3092C), was successful and yielded three additional strains: *Clostridium nexile* (positively correlated to lean mass gain), *Clostridium symbiosum* (lean mass gain discriminatory), and *Dorea formicigenerans* (weight and lean mass gain discriminatory).

Analogous to the co-housing experiment, male germ-free C57Bl/6J mice were placed on the M8 diet at 4.5-weeks-of-age: three days later they were colonized with the intact microbiota from the Un donor, with or without an equivalent mixture of the 5-member consortium ([Fig. 5A](#)). This consortium was given as single gavage. Mice were group housed (n=5/cage/treatment group/gnotobiotic isolator) based on the notion that there would be a greater chance that consortium members would be evenly distributed among the coprophagic animals. The results revealed that administration of the 5-member consortium produced a significant increase in body weight and lean mass gain compared to the untreated group (p=0.03, p=0.01, respectively, at the time of...
sacrifice 21 days after gavage, Mann-Whitney test) (Fig. 5B). Remarkably, only two members of the consortium, *R. gnnavus* strain TS_8243C and *C. symbiosum* strain TS_8243C successfully colonized recipient mice [27±3% (mean±SD) and 2.6±0.3% relative abundance, respectively, in feces obtained at sacrifice] (Fig. 5C). None of the other members were detected in recipients sampled on days 4, 7 and 14 after gavage. Moreover, the efficiency of incorporation of members of the Un donor’s microbiota into recipient mice was indistinguishable between the two treatment arms: i.e., (i) *R. gnnavus* and *C. symbiosum* did not result in extirpation of any major constituents of the community (the only species that fell below 0.1% was *B. bifidum*, an organism whose relative abundance was only 0.6±0.1% (mean±SD) at time of sacrifice in the control group); and (ii) the proportional representation of OTUs with respect to one another was similar in the two treatment arms (Fig. 5C, Table S15). These results provide direct evidence that *R. gnnavus* and *C. symbiosum* ameliorate the impaired growth phenotype transmitted by an immature undernourished donor’s microbiota. A reconstruction of selected metabolic subsystems in *R. gnnavus* TS_8243C and *C. symbiosum* TS_8243C yielded predictions about factors (prototrophies, auxotrophies, capacity to salvage/exchange critical nutrients) that may critically affect their representation in a normal developing gut microbiota or that could be used to manipulate their representation in an immature gut microbiota (see Supplementary Results, Table S16, and Fig. S8).

The presence of these two organisms also affected metabolic phenotypes (Fig. 5D, Table S17). The metabolic features that most discriminated mice with *R. gnnavus* and *C. symbiosum* from the untreated group were acylcarnitines (C5-C16), which were significantly increased in cecal samples and decreased in the liver and serum in mice harboring the two taxa (p<0.05, Students t-test). Some of these acylcarnitines also distinguished mice harboring the H compared to Un donor microbiota (Fig. 3A,B). Among metabolites that were decreased in the livers of treated mice were C5, C5-DC and C6-DC acylcarnitines, all of which can be derived from branched-chain amino acid catabolism. As proposed for the comparison of mice harboring the H versus Un donor microbiota shown in Fig. 3, the decrease in these liver metabolites suggests an impact of the presence of two growth-promoting taxa on host metabolic machinery that drives amino acids away from
oxidation in favor of protein synthesis and lean mass formation. The mechanisms by which the gut microbiota communicates metabolically with other tissues remain to be defined.

**Prospectus**

Human postnatal development is typically viewed from the perspective of our ‘human’ organs. As we come to appreciate how our microbial communities assemble following birth, there is an opportunity to determine how this microbial facet of our developmental biology relates to healthy growth, as well as to the risk for and manifestations of disorders that produce abnormal growth. The preclinical studies reported here provide evidence that gut microbiota immaturity is not only associated with undernutrition but causally related to it.

There are a number of interrelated factors that could result in disruption of normal gut microbiota development in infants and children. They include (i) poor maternal nutritional status and its potential effects on peripartum changes in the maternal microbiota and breast milk composition, (ii) enteropathogen invasion, where pathogen susceptibility may be linked to maturity of the microbiota if age- and growth-discriminatory taxa are able to produce signals that impact colonization efficiency/virulence factor expression, (iii) the history of consumption of antibiotics that may affect not only enteropathogen burden but also the representation of age/growth-discriminatory members of the microbiota, (iv) disturbances in gut mucosal immune system development which may perpetuate microbiota immaturity and impede maturation in the face of therapeutic interventions (20), and/or (v) the history of complementary feeding and the relationship between these complementary foods and the nutrient requirements of age/growth-discriminatory taxa. Carefully phenotyped birth cohorts provide an opportunity to perform correlation analyses designed to test the significance of the relationship between microbiota maturity (and the representation of specific age-/growth-discriminatory taxa), anthropometry, and factors (i)-(v) as well as the enigmatic disorder currently described as environmental enteric dysfunction (21).

An intriguing observation from the current study is that microbiota from 6-month-old children produce greater effects on growth in recently weaned mice than microbiota from 18-month-
old donors, although in each donor microbiota age bin the effects produced by a healthy donor’s age-appropriate community were greater than those produced by an immature undernourished donor microbiota. We take these findings to suggest that (i) in healthy children microbiota development is optimized to satisfy the different growth needs of the host at different ages, (ii) an immature microbiota that is not configured in a manner appropriate for the chronological age of the host manifests a form of neoteny that is not conducive to healthy growth, at least in the context of a macro- and micronutrient deficient diet, and (iii) repair of microbiota immaturity needs to be complete enough so that it attains age-appropriateness in order to restore healthy growth. An associated question relates to how the host will adapt to therapeutic interventions that result in a rate of progression of the microbiota from an immature to age-appropriate state over a time frame that is shorter than the extent of immaturity (e.g., rescue occurring over 2 months when the child has a microbiota which, at time of diagnosis, is 8 months younger than what is appropriate for his/her chronological age). Are there adaptive mechanisms (including those involving the mucosal immune system) that feed back to regulate the rate of restoration of microbiota maturity? The answers have implications for designing therapeutic strategies that produce durable rescue of stunting, neurodevelopmental and/or immunologic abnormalities associated with undernutrition. Young gnotobiotic mice colonized with undernourished donor microbiota manifesting varying degrees of immaturity, and subjected to microbial interventions at various times after colonization, offer one way of exploring these hypotheses/questions pre-clinically.

Clinical trials with ready-to-use complementary foods (RUCFs) of varying composition administered for varying periods of time with and without antibiotics have generally produced only modest effects on linear or ponderal growth and the associated longer-term sequelae of undernutrition (22). These less than satisfactory clinical outcomes may reflect the need for earlier intervention, longer duration treatment, or development of new food-based interventions, including those specifically targeting taxa in the microbiota that are key mediators of healthy growth/metabolism/immune function. One testable hypothesis is that certain locally available complementary foods
that are provided after cessation of exclusive breastfeeding have the ability to promote colonization of growth-discriminatory gut taxa in proportions that are age-appropriate.

Gnotobiotic mice colonized with microbiota from chronologically age-matched healthy and undernourished donors, and fed diets representative of those consumed by microbiota donors, should permit highly controlled, direct tests of the effects of factors such as antibiotics, breast milk components (e.g., bovine mimics of human breast oligosaccharides), and complementary foods on the representation, expressed functions, and mechanisms by which growth-promoting bacterial strains influence growth, metabolic, bone, immune and neurologic phenotypes. Gnotobiotic mice offer an opportunity to ‘enroll’ the microbiota of a given child with undernutrition into all arms of an existing, or an envisioned, randomized clinical trial of different interventions. They can also be used to ‘rewind time’ for a given individual and re-enact a therapeutic trial that they have participated in with original as well as new interventions to compare relative effect sizes. Such studies could be used to test the ability of microbiota-directed complementary food (MDCF)-based interventions alone to durably repair the gut microbiota, or whether adjunctive microbial therapies, identified using approaches such as those in the present study, may be required for complete restoration of a healthy microbiota. Ultimately, an initial clinical study, involving the very population whose microbiota and diets were used to construct the preclinical gnotobiotic models will be needed to establish proof-of-concept that microbiota immaturity can be durably repaired. If proof-of-concept is established, then appropriately powered follow-on studies will be needed to ascertain the effect size of such repair on healthy growth.
References and Notes


34. B. Chevreux et al., Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res 14, 1147–1159 (2004).


50. B. B. Williams et al., Discovery and characterization of gut microbiota decarboxylases that can produce the neurotransmitter tryptamine. Cell Host Microbe 16, 495–503 (2014).


Acknowledgements

We are indebted to the parents and children from Malawi for their participation in this study. We thank Sabrina Wagoner, David O’Donnell, Maria Karlsson, and Justin Serugo for their assistance with gnotobiotic mouse husbandry, Daniel Leib and Matt Silva for their assistance with bone morphology assays, Janaki Guruge for help with anaerobic microbiology, Betsy Dankenbring for assistance with maintenance of the biospecimen repository, plus Marty Meier, Su Deng, and Jessica Hoisington-Lopez for their contribution to various facets of the DNA sequencing pipeline. This work was supported by the Bill & Melinda Gates Foundation and the NIH (DK30292,) with additional funding from the Office of Health, Infectious Diseases and Nutrition, Bureau for Global Health, USAID under terms of Cooperative Agreement No. AID-OAA-A-12-00005, through the Food and Nutrition Technical Assistance III Project (FANTA) managed by FHI 360. Data management and statistical analysis for the iLiNS-DYAD-M clinical study were also funded by the Academy of Finland (grant 252075), and the Medical Research Fund of Tampere University Hospital (grant 9M004). Micro-CT analysis was performed in the Washington University Musculoskeletal Research Center, which is supported by NIH P30 AR057235. L.V.B. received stipend support from NIH predoctoral training grants NIH T32 AI007172, T32 GM007067, and the Lucille P. Markey Special Emphasis Pathway in Human Pathobiology. D.A.R. and S.A.L. were supported by the Russian Science Foundation (Grant 14-14-00289).

Data Deposition

16S rRNA sequences, generated from fecal samples in raw format prior to post-processing and data analysis, plus shotgun sequencing datasets generated from the R. gnatus TS_8243C and C. symbiosum TS_8243C genomes have been deposited at the European Nucleotide Archive under accession number PRJEB9853.
Declarations

J.I.G. is cofounder of Matatu Inc., a company characterizing the role of diet-by-microbiota interactions in animal health. A.L.O. is an Adjunct Vice President for Research for Buffalo BioLabs, LLC.

Author Contributions

L.V.B., M.R.C. and J.I.G. designed the gnotobiotic mouse studies; L.V.B. and M.R.C. performed the experiments with gnotobiotic animals; I.T. and M.J.M. designed and implemented the clinical monitoring and sampling for the twin study and participated in patient recruitment, sample collection and preservation, and/or clinical evaluations; K.M.M., Y.F., J.M.J., K.G.D. and P.A. designed and oversaw the clinical studies, sample collection and processing, and/or clinical monitoring/evaluations in the iLiNS-DYAD-M study; L.V.B. generated the 16S rRNA data; L.V.B., M.R.C., S.V., and O.I. generated metabolomics data; T.S. and L.V.B. cultured bacterial isolates; B.H., D.A.R., S.A.L., and A.L.O. performed metabolic reconstructions of the R. gnavus and C. symbiosum genomes; L.V.B., M.R.C., M.J.B., S.S., C.B.N., and J.I.G. analyzed the data; L.V.B. and J.I.G. wrote the paper.
Figure Legends

Fig. 1. Sparse Random Forests-derived model of gut microbiota maturation obtained from concordant healthy Malawian twins/triplets. (A) Random-Forests regression of fecal bacterial 97%ID OTUs from a training set of healthy Malawian infants/children \( n=31 \) to chronological age yielded a rank order of age-discriminatory strains. The top 25 most discriminatory OTUs with their taxonomic assignments are shown ranked by feature importance (mean±SD of the Mean Squared Error, MSE). The insert shows the results of 10-fold cross-validation; as OTUs are added to the model in order of their feature importance rank, the model’s error decreases. Taxa highlighted in red indicate OTUs with >97% nucleotide sequence identity with an OTU present in a sparse Random Forests-based model of microbiota maturation in healthy Bangladeshi infants/children (13). (B) A heatmap of the changes over time in relative abundances of the 25 OTUs in the fecal microbiota of test set of healthy Malawian infants/children \( n=29 \). OTUs are hierarchically clustered according to pairwise distances by Pearson correlation. (C) Predictions of chronological age using the sparse 25 OTU model of microbiota age in healthy children comprising the test set cohort. \( r^2 \) calculated from Pearson correlation.

Fig. 2. Transplantation of microbiota from 6- and 18-month-old donors to young germ-free mice provides evidence of a causal relationship between gut microbiota maturity and growth phenotypes. (A) Experimental design of microbiota screen. Mice (4.5-weeks old) were switched to the M8 diet three days prior to gavage with the selected microbiota donor’s fecal sample \( n=5 \) mice per donor). Fecal samples, body weight and body composition were defined at the indicated time. Tissues for metabolic phenotyping were collected at sacrifice five weeks after gavage. (B,C) Gnotobiotic mice colonized with fecal samples from healthy children gain more total body weight (panel b) and lean mass (panel c) than mice colonized with microbiota from undernourished donors (mean±SEM shown; p-values shown for donor status effect based on 2-way ANOVA). All recipient mice harbor microbiota that represent >50% of OTU diversity present in the intact uncultured donor’s sample. (D) The 30 most weight-gain discriminatory OTUs and their taxonomic assignments, ranked by feature importance (mean MSE±SD values are plotted). Taxa in red indi-
cate OTUs that appear within the 30 most discriminatory OTUs for both the weight and lean mass gain Random Forests-based models. Taxa in purple indicate species that appear in the 25-member sparse Random Forests-derived model of Malawi gut microbiota maturation. Bars to the right of the OTU ID numbers represent Spearman’s rank correlation of the same OTU ID to chronological age within the healthy Malawian infant/child cohort (see Table S9).

**Fig. 3. Metabolic features that distinguish gnotobiotic mice colonized with a microbiota from a 6-month-old healthy infant or an immature microbiota from a 6-month-old severely stunted/underweight infant.** Metabolic profiles generated by targeted mass spectrometry of (A) acylcarnitines in serum, (B) acylcarnitines in liver, and (C) amino acids in gastrocnemiaus muscle. Each metabolite’s concentration differs significantly between the two recipient groups of mice (p < 0.05; Student’s t-test).

**Fig. 4. Co-housing results in invasion of species from the microbiota of cagemates colonized with the healthy donor’s community into the microbiota of cagemates containing the severely stunted/underweight donor’s community and prevention of growth faltering.** (A) Experimental design for co-housing experiments. Dually housed 4.5-week old mice were switched to the M8 diet and colonized 3 days later with either the intact uncultured healthy or stunted/underweight donor microbiota. Four days after gavage, subsets of the mice were co-housed (HCH and UnCH, respectively), while healthy and stunted/underweight control mice remained in their original isolators and were paired with a new cagemate from that isolator (H-H and Un-Un controls). Fecal samples were collected throughout the experiment; growth was assayed by changes in total body weight and body composition (the latter by qMR). Mice were sacrificed three weeks after colonization. (B) HCH and UnCH mice have increased lean mass gain relative to the Un-Un controls 15 days post colonization (Un-Un vs. HCH p = 0.0447, Un-Un vs. UnCH p = 0.0121, Mann-Whitney test; n=6 cages of co-housed mice, 3 cages of each dually housed control group/experiment; two independent experiments). (C) Heatmap showing results of the invasion assay. Each row represents a species-level taxon, while each column represents a mouse at a given day post colonization (dpc); rows of the heatmap were hierarchically clustered according to pair-wise distances using
Pearson correlation. Bars at the right side of each experimental arm represent the fold-change (fc) in that species’ relative abundance before and after cohousing (fold-change defined by the \( \log_2\left[\frac{\text{average relative abundance of species post cohousing (days 7 through 22)}}{\text{average relative abundance of species before cohousing (day 4)}}\right] \)). Species in red represent those identified as one of the top 30 growth-discriminatory taxa by the weight or lean mass gain Random Forests-based models shown in Fig. 2.

**Fig. 5. A consortium of five cultured age- and/or growth-discriminatory OTUs augments the growth of mice colonized with the 6-month severely stunted/underweight donor’s microbiota.** (A) Experimental design including the composition of the 5-member consortium of cultured bacterial strains. (B) Weight gain and lean body mass gain 21 days after gavage of the donor microbiota with or without the 5-member cultured consortium. (C) Comparison of the fecal microbiota of mice belonging to untreated control and treated experimental groups showing the establishment of OTUs from the consortium at 21 days post colonization. (D) The effects of treatment with the consortium on host metabolism (p<0.05 for all metabolites shown, Student’s t-test). Each row represents a metabolite from a given tissue, while each column represents an individual mouse. Tissues were collected 21 days after colonization.
Figures

Fig. 1.

A. Rank OTU ID Taxonomic annotation
1 72820 Bifidobacterium longum
2 4414476 Lactobacillus rogosae
3 197004 Clostridium
4 176269 Eubacterium rectale
5 4331360 Eubacterium eligens
6 199145 Faecalibacterium sp DJF VR20
7 681370 Clostridium
8 185763 Faecalibacterium prausnitzii
9 4318208 Prevotella copri
10 184464 Prevotella copri
11 4326870 Dialister succinatophilus
12 4451506 Clostridiales
13 1614788 Mitsuokella multacida
14 43000127 Clostridiales
15 188047 Prevotella copri
16 4410166 Prevotella copri
17 365385 Clostridium
18 4473509 Faecalibacterium
19 4410166 Streptococcus
20 193667 Clostridiales
21 306412 Roseburia faecis
22 177515 Bifidobacterium bifidum
23 4410166 Clostridiales
24 191361 Faecalibacterium sp DJF VR20
25 4436552 Prevotella copri

B. Age (months) OTU ID Taxonomy
4 72820 Lactobacillus rogosae
8 4414476 Clostridium
12 4331360 Prevotella copri
16 197004 Eubacterium rectale
20 176269 Eubacterium eligens
24 4331360 Faecalibacterium sp DJF VR20
28 199145 Clostridium
32 185763 Faecalibacterium prausnitzii
36 4318208 Prevotella copri
40 184464 Prevotella copri
44 4326870 Dialister succinatophilus
48 4451506 Clostridiales
52 1614788 Mitsuokella multacida
56 43000127 Clostridiales
60 188047 Prevotella copri
64 4410166 Prevotella copri
68 365385 Clostridium
72 4473509 Faecalibacterium
76 4410166 Streptococcus
80 193667 Clostridiales
84 306412 Roseburia faecis
88 177515 Bifidobacterium bifidum
92 4410166 Clostridiales
96 191361 Faecalibacterium sp DJF VR20
100 4436552 Prevotella copri
104 4441855 Faecalibacterium
108 184464 Prevotella copri
112 4326870 Dialister succinatophilus
116 4473509 Clostridium
120 4410166 Faecalibacterium
124 365385 Bifidobacterium longum

C. Microbiota age (months) Chronological age (months)
0 0
10 10
20 20
30 30
40 40
50 50 $r^2 = .80$
Fig. 2.

A) Timeline of experimental procedures:
- Fecal sample
- Body compositional analysis (qMR)
- Body weight Measurement
- Tissues for metabolic phenotyping
- Birth
- Begin M8 Colonization
- Sacrifice

Timeline:
- 4.5 weeks
- 3 days
- 5 weeks

B) Graph showing % Initial weight over Days post colonization for Healthy (3 donors) and Stunted/underweight (5 donors) groups.

C) Graph showing % Initial lean mass over Days post colonization for Healthy (3 donors) and Stunted/underweight (5 donors) groups.

D) OTU ID and weight indicative taxa:
- de novo 12617
- 2724175
- 2714942
- 4383953
- 72820
- 4374302
- 3709990
- 138389
- 2575651
- 170652
- 64384
- 342307
- 164546
- 681370
- 1033413
- 3203401
- 19466010
- 2148366
- 3665054
- 773251
- 4316929
- 4484492
- 199436
- 2700883
- 2148365
- 4348151
- 1860112
- 1142128
- 2424737

Weight indicative taxa:
- None
- Clostridium butyricum
- Ruminococcus gnavus
- Clostridium disporicum
- Bifidobacterium longum
- Faecalibacterium prausnitzii
- Enterococcus faecalis
- Bacteroides vulgatus
- Paenibacilluscola
do not hallucinate.
Fig. 3.

A. Serum acylcarnitines

B. Liver acylcarnitines

C. Gastrocnemius muscle amino acids

Healthy (H) vs. Stunted/underweight (Un)
Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.

A

B

C

Fig. 4.
Fig. 5.

A) Isolator 1 ("Control") and Isolator 2 ("Treated") with 5-member consortium:
- P. prausnitzii (OTU 265871)
- R. gravis (OTU 2724175)
- C. nexile (OTU 4436046)
- C. symbiosum (OTU 4378883)
- D. formicigenerans (OTU 4374302)

B) Analysis of weight and lean mass:
- Begin M8 Colonization Sacrifice
  - 3 days
  - 21 days
- Analysis:
  - Fecal sample
  - Body weight measurement
  - Body compositional analysis
  - Tissues for metabolic phenotyping

C) Percent relative abundance, 21 days post colonization:
- Clostridium symbiosum
- Ruminococcus gravis
- Weisella
- Pedicoccus
- Lactobacillus
- Eubacterium limosum
- Clostridium ramosum
- Eubacterium limosum
- Clostridium perfringens
- Clostridium difficil
- Clostridium botulim
- Enterococcus
- Clostridium neonatale

D) Metabolites in different tissues:
- Liver:
  - C18
  - C5-DC
  - C6-DC/C8-OH
  - C14:3
  - C5
  - C12-OH
  - C10-DC
- Serum:
  - C16:1-OH/C14:1-DC
  - C12-OH/C10-DC
  - C5
  - C10:2
- Gastrocnemius:
  - C21:3
  - C20:3
  - C19:3
  - C22:5
- Cecal:
  - Glx
  - His
  - Glt
  - C20:4
  - C18:1-OH/C16:1-DC
  - C16:1-OH/C14:1-DC
  - C20:1
  - C16:1
  - C16
Supplementary Materials

Materials and Methods

Human Studies

Human studies were conducted with approval of institutional review boards from the University of Malawi, Pirkanmaa Hospital, and Washington University School of Medicine. Details of the 317 infant twin pairs and three sets of triplets enrolled in the Malawi Twin Cohort study from birth to 32-months-of-age are described in an earlier publication (14).

Enrollment for the iLiNS-DYAD-M study [clinicaltrials.gov #NCT01239693] was open to consenting women over the age of 15 years with ultrasound confirmation of a pregnancy that was <20 weeks, who had come for prenatal care at one of three study clinics located in Mangochi District of southern Malawi (one public district hospital in Mangochi, one rural semi-private hospital in Malindi, and one rural public health center in Lungwena). This randomized, controlled clinical trial tested the effects of providing Small Quantity Lipid-based Nutrient Supplements (SQ-LNS) to pregnant and lactating women through 6 months postpartum and to their infants from 6-18-months of-age (15).

Bacterial 16S rRNA sequencing

Genomic DNA was extracted by suspending pulverized human fecal samples (~50mg), or in the case of mouse studies, resuspending fecal pellets in a solution containing 500μL of extraction buffer [200 mM Tris (pH 8.0), 200 mM NaCl, 20 mM EDTA], 210μL of 20% SDS, 500μL phenol:chloroform:isoamyl alcohol (pH 7.9, 25:24:1, Ambion), and 500μL of 0.1-mm diameter zirconia/silica beads. Cells were mechanically disrupted using a bead beater (BioSpec Products, Bartlesville, OK; maximum setting for 4 min at room temperature), followed by extraction with phenol:chloroform:isoamyl alcohol and precipitation with isopropanol.
PCR was used to generate amplicons (~365bp) of variable region 4 (V4) of the bacterial 16S rRNA gene using primers and cycling conditions described in a previous report (13). PCR primers incorporated sample-specific barcodes allowing samples to be subjected to multiplex sequencing using the Illumina MiSeq instrument (paired-end 250nt reads). Paired V4-16S rRNA sequences were trimmed to 200bp and merged into a single sequence with Flash software (23). Merged sequences were filtered for low quality reads and binned according to their sample-specific barcodes. Reads were clustered into 97%ID OTUs using UCLUST (24) and the Greengenes OTU reference database (version May 2013). Reads that failed to hit the reference dataset were clustered de novo using UCLUST. A representative OTU set was created using the most abundant OTU from each bin. Reads were aligned using PyNAST. A custom dataset of manually modified NCBI bacterial taxonomy (19) was used to train the Ribosomal Database Project (RDP) version 2.4 classifier (25) and to assign taxonomy to picked OTUs. Validation of this assignment strategy is described in a previous publication (19). For analysis of bacterial abundance and colonization efficiency, rare OTUs (<0.1% relative abundance within each sample) were removed.

**Random Forests-derived Models and Statistical Analysis**

Models were generated using the randomForests package in R. 1000 trees were generated per iteration, and 100 iterations were performed to generate average MSE values. The rfcv function was utilized for cross-validation.

Univariate analyses (t-tests and ANOVA) were performed using Microsoft Excel or Prism 6.0 (GraphPad Software, Inc.). Heat maps were generated using Gene-E software (http://www.broadinstitute.org/cancer/software/GENE-E/index.html). Spearman’s rank correlation analyses were performed in the R environment using the ltm package and rcor.test function.

**Gnotobiotic Mouse Studies**

All gnotobiotic mouse experiments were performed using protocols approved by the Washington University Animal Studies Committee. Male germ-free C57Bl/6J mice were maintained in sterile,
flexible, plastic gnotobiotic isolators (Class Biologically Clean Ltd., Madison, WI) under a strict 12-hour cycle (lights on at 0600, off at 1800h). Mice were fed an autoclaved low-fat, polysaccharide-rich chow (LF/HPP) diet (B&K University, East Yorkshire, U.K.; diet 7378000) from weaning until 3 days prior to the beginning of an experiment. At that time, 4.5-week old male mice were switched to the M8 diet for the remainder of the experiment. All animals were euthanized without prior fasting.

**Preparation of the Malawi-8 (M8) Diet**

This prototypic Malawian diet was formulated based on the results of a dietary survey of the complementary feeding practices of 9-month-old Malawian children enrolled in the iLiNS-DOSE study (#NCT00945698) that took place in the Mangochi district. The macro- and micronutrient content of components of the children’s diets was computed using the USDA Nutrient Database, and linear programming was employed to calculate combinations of ingredients that resembled the mean energy and nutrient values for foods identified from the dietary survey.

The M8 diet was prepared as follows. Maseca® corn flour was obtained from Restaurant Depot (College Point, NY). The remaining ingredients were purchased from Whole Foods Supermarkets. A relish containing 2kg mustard greens, 1.5kg onions, and 1.5kg tomatoes was first prepared by pureeing them in a food processor (Robot Coupe Model R23, Jackson, MS) and cooking the puree in 1L water for 60 minutes on a Corning stirrer/hot plot (high setting, until browned). After cooking, the relish was combined with a pureed mixture of 1kg ground peanuts, 700g soaked red kidney beans, 1kg canned pumpkin, and 2.5kg peeled bananas in an industrial mixer (Globe SP30P 30-quart pizza mixer; gear speed 1; Globe Food Equipment Company, Dayton, OH). Corn flour (5kg) and hot, freshly autoclaved water (5L) were then added slowly and mixed using the industrial mixer for 5 minutes to ensure uniformity of the food. Dry pellets of the M8 diet were generated (extruded diameter, 0.5-inch) (Dyets, Inc. Bethlehem, PA), placed in FDA/USDA-compliant poly-nylon vacuum pouches (#S-7556; Uline, Pleasant Prairie, WI; 500 g aliquots), double-bagged and sterilized by irradiation (20-50 kGy) within 24 hours of production.
(Steris Co; Chicago, IL). The nutritional content of all cooked and irradiated custom diets was defined by N.P. Analytical Laboratories (St Louis, MO) (Table S4). Irradiated food was stored at 4°C for up to six months. Sterility was determined by resuspending a small aliquot of each batch of food in pre-reduced Gut Microbiota Medium (GMM; (26)) under anaerobic conditions and incubating the suspension for 3 days at 37°C. Sterility was further verified by subculture on pre-reduced anaerobic GMM agar plates.

**Generation of clarified stool for gavage into gnotobiotic mice**

An aliquot of the frozen sample was pulverized in a Biosafety Class II hood with a ceramic mortar and pestle filled with liquid nitrogen. An aliquot (1g) of the pulverized material, sealed in a sterile screw-capped tube (Axygen SCT-200-C-S), was brought into an anaerobic Coy chamber (atmosphere: 20% CO₂, 5% H₂, and 75% N₂), immediately suspended in 15mL GMM and vortexed on maximum speed (four cycles of blending for 20 seconds followed by a 30 second pause). The sample was allowed to stand for 5 minutes so that particulate matter could settle by gravity; the resulting supernatant was passed through a 100µm pore diameter filter (BD systems, Inc. Franklin Lakes, NJ) to remove remaining particulate material, mixed with an equal volume of pre-reduced GMM containing 30% glycerol (final concentration 15% glycerol) and placed in Wheaton crimp top tubes for storage at -80°C. Mice received a single oral gavage (200µL) of a clarified uncultured human fecal sample 3 days after switching to the M8 diet.

**Quantitative magnetic resonance (qMR) analysis of body composition**

Body composition was defined using an EchoMRI-3in1 instrument (EchoMRI, Houston, TX). Each mouse was transported from a gnotobiotic isolator to the MR instrument in a 9.5” long, 2” diameter, HEPA filter-capped glass vessel. At each time point surveyed, each mouse was scanned 2-3 times in the instrument over a 5 minute period and the values from each scan were averaged.
**Sample collection**

Fecal samples were collected at defined times after gavage. At sacrifice, cecal contents, liver, brain, and skeletal muscle (gastrocnemius) were collected and immediately frozen in liquid nitrogen. Gastrocnemius muscle was freeze-clamped immediately after sacrifice using pre-cooled forceps.

**Micro-computed tomography (micro-CT)**

Femurs were harvested from mice at time of sacrifice, cleaned of soft tissue and stored at 4°C in 70% ethanol solution until scanned. Micro-CT was performed using a μCT 40 desktop, cone-beam instrument (ScanCO Medical, Brüttisellen, Switzerland). For cortical analyses, 200-300 slices were taken of each femur in the transverse plane with a 6μm voxel size (high resolution). For all cortical scans, slices began at the midpoint of the femur and extended toward the distal femur. Boundaries of and thresholds for bone were drawn manually using the μCT 40 software, and volumetric parameters (bone volume/tissue volume, bone mineral density, and cortical thickness) were calculated using custom scripts.

**Co-housing experiments**

Four days after gavage of donor microbiota, all mice were placed into HEPA filter-capped glass transport tubes for qMR measurements. Following qMR, control mice were returned to their original gnotobiotic isolator and housed with a cagemate of similar body weight colonized with the same donor microbiota (H-H controls; Un-Un controls). For pairing of UnCH and HCH mice, animals were transported to a new gnotobiotic isolator. All mice received fresh autoclaved bedding (Aspen wood shavings, NEPCO) upon initiation co-housing. Bedding was changed weekly in all cages in each gnotobiotic isolator.

**Targeted mass spectrometry**

Liver, brain, and muscle tissue samples were homogenized in 50% aqueous acetonitrile containing 0.3% formic acid (50mg wet weight tissue/mL solution) using a high-speed homogenizer (IKA...
Model #EW-04739-21) set at maximum speed for 30-45 seconds. Samples were maintained on ice or dry ice throughout homogenization.

Amino acids, acylcarnitines, organic acids, acyl CoAs and ceramides were analyzed using stable isotope dilution techniques. Amino acids and acylcarnitine measurements were made by flow injection tandem mass spectrometry using sample preparation methods described previously (27, 28). Data were acquired using a Waters Acquity™ UPLC system equipped with a TQ (triple quadrupole) detector and a data system controlled by MassLynx 4.1 operating system (Waters, Milford, MA). Organic acids were quantified according to a previously published protocol (29) using Trace Ultra GC coupled to ISQ MS operating under Xcalibur 2.2 (Thermo Fisher Scientific, Austin, TX). Acyl CoAs were extracted and purified as described previously (30-32), and analyzed by flow injection analysis using positive electrospray ionization on a Xevo TQ-S triple quadrupole mass spectrometer (Waters, Milford, MA). Heptadecanoyl CoA was employed as an internal standard. Ceramides were extracted (33) and analyzed by flow injection tandem mass spectrometry using a Xevo TQS spectrometer (Waters Milford, MA,) for precursors of m/z 264.

**Culturing and characterization of age- and/or growth- discriminatory bacterial strains from the microbiota of healthy Malawian children**

*Culturing* – Fecal samples were resuspended in 10ml of GMM under anaerobic conditions, diluted 1:1000, and then plated onto GMM and incubated at 37°C under anaerobic conditions overnight. Isolated colonies were picked into single wells of a deep 96-well plate, with 600μ l liquid GMM per well, and grown anaerobically at 37°C overnight. Cultures were then mixed with an equal volume of pre-reduced GMM containing 30% glycerol (final concentration 15% glycerol) for storage at -80°C. To assign taxonomy, DNA was extracted from each well and V4-16S rRNA sequenced. Taxonomies were confirmed by sequencing full-length 16S rRNA amplicons generated using primers 8F and 1391R.
Genome sequencing of two strains obtained from the healthy infant Malawian donor 8243C

DNA libraries for sequencing *R. gnavus* TS_8243C and *C. symbiosum* TS-8243C on Illumina MiSeq were prepared using a multistep protocol of end-repair, A-tailing, and ligation to customized sample-specific barcoded Illumina adapters. The resulting adapter-ligated DNA was size-selected (250-350bp) using agarose gel electrophoresis and then used as a template for enrichment PCR (cycling conditions: 98°C for 30 sec, followed by 18 cycles of 98°C for 10 sec, 65°C for 30 sec, 72°C for 45 sec, and then 72°C for 5 minutes). Multiplex sequencing was performed (paired-end 250nt reads). Genomes were assembled with the mira software package (34). For *R. gnavus* TS_8243C N50 contig length was 2263429bp with an aggregate genome size of 3.7Mbp. For *C. symbiosum* TS_8243C, N50 contig length was 328298bp with a 5.3Mbp genome size. Annotation was initially performed using PROKKA (see *Supplementary Results* for metabolic reconstructions) (35).

**Supplementary Results**

**Genomics-based reconstruction of selected metabolic subsystems in *R. gnavus* and *C. symbiosum* strains**

Our approach to the reconstruction of selected metabolic pathways (also termed as subsystems) was based on functional gene annotation and prediction using two comparative genomics techniques: (i) homology-based methods and (ii) genome context analysis. First, we scanned their predicted proteomes against the KEGG Orthology, Pfam, CAZY, TCDB, Uniprot and SEED databases to identify proteins potentially involved in the carbohydrate utilization, amino acid metabolism and vitamin and cofactor biosynthesis and salvage pathways. We then analyzed the genomic and functional contexts of loci encoding the proteins and reconstructed the respective metabolic pathways and transcriptional regulons. For the comparative genomics-enabled pathway and regulon inference, we used additional closely related bacterial genomes available in the Integrated Microbial Genomes (36) and SEED databases (37). Transcriptional regulons (sets of genes co-regu-
lated via shared transcription factors (TFs) and transcription factor binding sites (TFBS) or shared RNA elements, riboswitches, responding to respective pathway metabolites) were predicted and reconstructed using the comparative genomics approach\(^{38}\) implemented in the RegPrecise database\(^{39}\). This integrative subsystems-based approach was previously used for reconstruction of carbohydrate utilization pathways and regulons in the *Shewanella*, *Thermotoga*, and *Bacteroides* lineages\(^{40-42}\).

**Carbon and energy: central carbon metabolism, carbohydrate utilization, and short chain fatty acid production (Fig. S8A and Table S16).**

**Overview**

We applied the integrated subsystems-based approach described above to systematically map central carbon metabolism, peripheral carbohydrate utilization, and fermentation pathways in the sequenced genomes of the two TS_8243C strains.

**Central carbon metabolism**

Both the *R. gnavus* and *C. symbiosum* strains possess the complete glycolysis/gluconeogenesis and pentose phosphate (PP) pathways, as well as the key enzymes involved in conversion of pyruvate to acetyl-CoA or oxaloacetate, and reverse conversion of oxaloacetate to phosphoenolpyruvate (PEP). Genes encoding key glycolytic enzymes, as well as key fermentation pathways (see below), are co-regulated by the redox-responsive transcriptional regulator Rex\(^{43}\). The PP pathway in *C. symbiosum* involves multiple paralogs of genes encoding transketolase (TktAB) and transaldolase (Tal). In addition, *C. symbiosum* (but not *R. gnavus*) encodes the oxidative pathway for utilization of glucose-6 phosphate via the PP pathway (Zwf, Pgl, Gnd).

The tricarboxylic acid (TCA) cycle is incomplete in both strains. In *R. gnavus*, it involves two parts: (i) interconversion of pyruvate to malate and fumarate and (ii) fusion of oxaloacetate and acetyl-CoA to produce 2-oxoglutarate, a key precursor for glutamate biosynthesis. In contrast,
*C. symbiosum* possesses only the former pathway for conversion of pyruvate to malate and fumarate. The *C. symbiosum* strain also has the fumarate reductase FrdABC allowing synthesis of succinate, which is further fermented to produce butyrate and butanol via the committed pathway (see below). Interestingly, FrdABC is also present in the reference ATCC 29149 strain of *R. gnavus* but it is missing from the TS_8243C strain, suggesting the former but not the latter strain is capable of producing succinate, which is likely excreted as an additional fermentation product since the succinate fermentation pathway is absent in *R. gnavus.*

**Carbohydrate utilization (CU) pathways**

The reconstructed carbohydrate utilization (CU) potentials of the two Malawian TS_8243C strains differ substantially from each other and include numerous novel carbohydrate-specific regulators, transporters, and enzymes that are not orthologous to previously characterized proteins (marked by an asterisk in Table S16 and Fig. S8A). In *R. gnavus,* we predicted 28 novel regulators, most of which belong to the AraC family and 35 transporter genes encoding 16 transporter complexes, whereas the reconstructed CU pathways in *C. symbiosum* contain 5 novel regulators and 17 transporter genes (5 transporter complexes). Among the most notable novel forms of monosaccharide catabolizing enzymes identified in the reconstructed pathways are xylose isomerase (XylA), tagatose-6P isomerase (LacC), galactitol kinase (GatK), and bifunctional fuculose-P/rhamnulose-P aldolase (FucA) in *R. gnavus,* as well as arabinose isomerase (AraI) and mannitol dehydrogenase (MtlD) in *C. symbiosum.*

The conserved CU components between *R. gnavus* and *C. symbiosum* include 17 genes encoding the glucose phosphotransferase system (PTS) transporter and glucokinase (PtsG, Glk), galactose catabolic enzymes (GalK, GalM, GalT, GalE), the fructose PTS transporter and catabolic enzymes (FruA, FruK, ScrK), N-acetylglucosamine and sialic acid catabolic enzymes (NagA, NagB, NanK, NanA, NanE), galacto-N-biose phosphorylase (GahP), and two neopalulullanase iso-enzymes (NplT, NepU). These shared CU genes, combined with other strain-specific upstream pathway components such as sugar transporters and hydrolytic enzymes, provided the respective CU
phenotype predictions for each strain. *R. gnnavus* is predicted to utilize glucose, fructose and its oligosaccharides, lactose, alpha- and beta-galactosides, N-acetylglucosamine, N-acetylneuraminate and sialic acids, as well as maltodextrins. (However, we were unable to find a galactose transporter in *R. gnnavus* suggesting it can only utilize galactose oligosaccharides.) In contrast, *C. symbiosum* uses the above 17 shared genes and additional upstream components for utilization of glucose, fructose and its oligosaccharides, beta-galactosides, N-acetylneuraminate and chitobiose (although the respective transporter is missing). In both strains, we were unable to identify a galacto-N-biose transporter, despite the presence of galacto-N-biose phosphorylase and all downstream enzymes for catabolism of galactose and N-acetylglucosamine-P.

In addition to the CU pathways described above that involve shared genes and other non-conserved upstream pathway genes, *R. gnnavus TS_8243C* and *C. symbiosum TS_8243C* possess 126 and 26 strain-specific CU genes, respectively. *R. gnnavus* has predicted pathways for utilization of sucrose, beta-xylosides and rhamnogalacturonides (but not xylose and rhamnose since the respective monosaccharide transporters are missing), fucose and fucosides (with at least eight copies of alpha-L-fucosidases), glucuronides and polygalacturonate (but not glucuronate/galacturonate for which we found no transporters), beta-glucosides and cellobiose, meliobiose, fucosyllactose, hyaluronate as well as galactitol (although a galactitol transporter is missing from the reconstruction). Orthologs of genes comprising these *R. gnnavus*-specific pathways are absent from *C. symbiosum TS_8243C*, but many of them can be identified in other gut symbionts. *C. symbiosum TS_8243C* has only three additional unique pathways for utilization of monosaccharides, namely arabinose, mannitol, and gluconate. Thus, we predict that *R. gnnavus* has a greater potential to utilize di- and oligosaccharides, including sugars originating from milk (such as lactose, fucosyl-lactose, fucosylated oligosaccharides) as compared to *C. symbiosum*.

*R. gnnavus TS_8243C* and *C. symbiosum TS_8243C* have shared and unique pathways for utilization of other carbon sources, in addition to the carbohydrates described above. Both strains share the glycerol and D-lactate utilization pathway components (GlpK, GlpD, GlpD, UgpABCE, and LdhD). *C. symbiosum TS_8243C* has unique pathways for utilization of ethanolamine and
citrate that involve the B_{12}-dependent ethanolamine ammonia-lyase EutABC and the CoA-dependent citrate lyase complex CitFEDC. The latter enzyme is represented by multiple paralogs in the *C. symbiosum* genome and is accompanied by two copies of the citrate transporter CitN. *R. gnnavus* has a unique pathway for utilization of D-glycerate, which involves a predicted transporter (GrtP*), kinase (GarK), and repressor (SdaR). Pathways for utilization of certain amino acids as potential carbon sources are described in the following section below.

In the case of *C. symbiosum*, comparison of CU genes in the TS_8243C and the reference WAL-14163 type strains did not reveal any differences. This was not the case for *R. gnavus*. The TS_8243C strain of *R. gnavus* has 36 CU genes that are missing in the ATCC 29149 type strain. Most of the *R. gnavus* TS_8243C-specific genes are organized into five individual gene loci encoding the complete pathways for utilization of beta-glucosides, beta-xylosides, meliobiose, lactose and beta-galactosides that involve dedicated transporters, regulators and sugar hydrolytic enzymes. Additional *R. gnavus* TS_8243C-specific genes, including four copies of alpha-L-fucosidases, three beta-glucosidases, two beta-galactosidases, alpha-rhamnosidase (all predicted cytoplasmic glycoside hydrolases), as well as an extracellular beta-glucanase, LicB, are scattered throughout the chromosome (whether they are co-regulated with downstream CU pathway genes remains to be determined). A second copy of rhamnogalacturonyl hydrolase (RhiN2) is encoded within the polygalacturonate utilization gene cluster in *R. gnavus* TS_8243C but not in the ATCC 29149 strain. In contrast, the ATCC 29149 strain of *R. gnavus* has only eight strain-specific CU genes that are missing from the TS_8243C strain. These genes are potentially dispensable since they encode second copies of maltose/maltodextrin and sucrose utilization operons.

**Short-chain fatty acid production**

Short-chain fatty acids (SCFAs), the end products of fermentation of dietary carbohydrates and polysaccharides, exert multiple beneficial effects on mammalian energy metabolism and provide a primary energy source for colonic epithelial cells (44). Our metabolic reconstruction of central carbon metabolism in the *R. gnavus* TS_8243C and *C. symbiosum* TS_8243C strains revealed both
shared and strain-specific pathways for fermentation of various carbon sources and production of SCFAs. Moreover, gene sets comprising the SCFA production pathways of the two TS_8243C strains (23 genes in R. gnavus and 38 genes in C. symbiosum) are essentially identical to those of the corresponding reference strains.

Both the R. gnavus TS_8243C and C. symbiosum TS_8243C strains possess complete pathways for production of formate, lactate, acetate and ethanol from pyruvate. R. gnavus has a unique pathway for production of propionate/propanol by fermentation of fucose and rhamnose via the committed CU pathways (see above). This pathway involves lactaldehyde reductase (FucO) and B$_12$-dependent propanediol dehydratase (PduABCD), plus several other downstream enzymes for conversion of propionaldehyde to propionate and propanol. We postulate that the predicted transcriptional regulator PduR* from the DeoR family controls genes from this unique pathway. C. symbiosum TS_8243C has the complete pathway for conversion of acetyl-CoA to butyrate and butanol via crotonoyl-CoA and butyryl-CoA intermediates (as previously described in (45)). Most of the butyrate synthesis pathway genes are organized into a single gene locus controlled by the redox-responsive transcriptional regulator Rex (43). Two other pathways in C. symbiosum TS_8243C that are potentially used for the synthesis of the crotonoyl-CoA (which is further converted to butyrate and butanol) start from succinate and 2-oxoglutarate, respectively. The first pathway includes the predicted succinate-semialdehyde dehydrogenase AldD* and proceeds via 4-hydroxybutyrate and 4-hydroxybutyl-CoA intermediates. Succinate for this pathway is provided by the fumarate reductase FrdABC, which is also present only in C. symbiosum. Another pathway for butyrate synthesis has a missing enzymatic activity, 2-hydroxyglutarate dehydrogenase, while the conversion of 2-hydroxyglutarate to crotonoyl-CoA is catalyzed by three enzymatic complexes (GctAB, HgdABC and GcdABC). The potential source of 2-oxoglutarate for the latter pathway of butyrate production is degradation of exogenous L-glutamate via glutamate dehydrogenase Gdh, since isocitrate dehydrogenase Icd is absent from C. symbiosum TS_8243C (see the section on amino acid degradation below).
C. symbiosum TS_8243C also has the complete pathway for further conversion of lactate to propionate, which involves propionate CoA-transferase (Pct), lactoyl-CoA dehydratase (Lcd), acryloyl-CoA reductase (Acr) and is equipped with predicted propionate permease (PrpP*) for end product excretion. Thus, C. symbiosum TS_8243C and R. gnatus TS_8243C possess distinct pathways for production of propionate.

Amino acids (Fig. S8B and Table S16)

By using the metabolic subsystem approach implemented in the SEED genomic platform (46) and RAST server (37), we reconstructed metabolic pathways for 20 amino acids by identifying the committed enzymes, potential amino acid transporters and transcriptional regulators in R. gnatus TS_8243C and C. symbiosum TS_8243C.

We found that both R. gnatus TS_8243C and C. symbiosum TS_8243C are capable of synthesizing most amino acids de novo. The histidine biosynthesis pathway in R. gnatus TS_8243C is composed of 11 committed enzymes, most of which are organized into a single operon, which is controlled by a HisR repressor. The HisR regulon in R. gnatus also includes the predicted ABC-type His transporter HisXYZ* suggesting it can also salvage exogenous His. In contrast, the reconstructed HisR regulon in C. symbiosum includes the hutH-yuiF operon involved in His uptake and degradation; other His degradation genes were also identified in this genome. Thus, C. symbiosum is a predicted His auxotroph able to salvage exogenous His and use it both for protein synthesis and as a source of carbon, energy and nitrogen. Glutamate is synthesized from 2-oxoglutarate using the committed enzyme glutamate synthase GltBD, which is present in R. gnatus but not in C. symbiosum, suggesting the latter is a Glu auxotroph. In support of this prediction, we found two copies of the glutamate uptake transporter gene glutS in C. symbiosum TS_8243C but not in R. gnatus TS_8243C. Furthermore, C. symbiosum TS_8243C but not R. gnatus TS_8243C has a unique pathway for degradation of exogenous glutamate to produce butyrate, an important short-chain fatty acid that provides energy to colonocytes and has been reported to be capable of
inducing production of colonic Foxp3+ CD4+ regulatory T cells (Tregs; (47-49)) (see Fig. S8A, plus the section on carbon and energy above).

In addition to the His and Glu uptake transporters described above, specific amino acid transporters were predicted for arginine, tyrosine, tryptophan, glutamine, branched chain amino acids (isoleucine/leucine/valine), alanine and methionine. These amino acid uptake transporters were found both in R. gnavus and C. symbiosum suggesting that these organisms can salvage these amino acids from the environment, thus providing an alternative way to supply their biosynthetic needs. Uptake transporters for other amino acids are as yet unknown in analyzed Clostridiales species.

As noted in the main text, we identified a link between perturbations in microbiota maturation and perturbations in brain metabolism, raising the question of how normal development of the gut microbiota is linked to healthy brain development. Tryptophan is a precursor of the β-arylamine neurotransmitter tryptamine, which was recently found to be produced by tryptophan decarboxylase in several gut Clostridiales, including R. gnavus (50). Tryptamine has a range of biological activities; notably, it is found in low quantities in the brain and is a ligand for the trace-amine-associated receptors (TAARs) that potentiate neuronal inhibition to serotonin (51). The R. gnavus gene RUMGNA_01526 (named Trd, see Fig. S8B) encoding tryptophan decarboxylase has an ortholog in the R. gnavus TS_8243C strain, suggesting that is able to produce tryptamine. In contrast, both strains of C. symbiosum do not have a Trd ortholog.

B-vitamins (Fig. S8C and Table S16)

Vitamins B1 (thiamin), B2 (riboflavin), B3 (niacin), B5 (pantothenate) and B6 (pyridoxine) are precursors of thiamin pyrophosphate (TPP), flavin mono- and dinucleotide (FMN and FAD), nicotinamide adenine dinucleotide (NAD), coenzyme A (CoA) and pyridoxal phosphate (PLP), respectively. Taken together, these cofactors drive hundreds of indispensable biochemical transformations including redox reactions (FMN/FAD and NAD), central carbon metabolism (CoA and TPP) and transamidation (PLP). Importantly, none of these phosphorylated cofactors can be deliv-
ered into cells in their final form. Other B-vitamins, B7 (biotin), B9 (folate) and B12 (cobalamin) and their derivatives serve as cofactors of important enzymes involved in fatty acid biosynthesis, single carbon metabolism, methionine biosynthesis, and a few other pathways.

We have previously applied the subsystems-based approach described above to characterize the biosynthesis, uptake/salvage, recycling of B vitamin related co-factors across numerous diverse bacterial species (52, 53). This analysis is captured in a curated collection of subsystems in the SEED genomic platform (http://pubseed.theseed.org/) (46), which we recently expanded to include > 1,000 genomes from Human Microbiome Project (HMP, http://hmpdacc.org/). In general, only limited variations of the main biosynthetic steps and corresponding enzyme families are observed, even between very distant species. Components of B-vitamin transport and regulatory systems are far more diverse and remain to be identified in many bacterial lineages. The most common type of variation is between: (i) organisms capable of synthesizing a particular cofactor *de novo* (prototrophs) and (ii) those that are strictly dependent on an exogenous (dietary) supply of a respective B-vitamin or other precursor(s) (auxotrophs). Genomic signatures distinguishing prototrophs and auxotrophs for eight common B-vitamins (B1, B2, B3, B5, B6, B7, B9, B12) were revealed from a recent analysis of 256 representative HMP genomes (54). Widespread auxotrophy among species colonizing the distal gut points to syntrophic B-vitamin metabolism (“vitamin exchange”) as a likely prominent feature of the community. Thus, elucidating specific B-vitamin requirements for strains that define the different stages of gut microbial community assembly, and whose representation in the microbiota is linked to growth, metabolic, immunologic and/or neurologic phenotypes is not only important for understanding the ‘rules’ that govern normal microbiota development but could also yield interventions designed to treat microbiota immaturity in undernourished children.

Addition of the *R. gnavus* TS_8243C and *C. symbiosum* TS_8243C genomes to eight B-vitamin subsystems (using RAST server (37)) for automated projection followed by manual cura-
tion, analysis of genomic context, and gap-filling confirmed their multiple auxotrophies, a characteristic feature of most host-associated Clostridiales. This distinguishes them from their free-living
phylogenetic neighbors (such as *C. acetobutylicum*) or other prominent members of the distal human gut microbiota, such as the *Bacteroides* that are typically prototrophic for all B-vitamins (except B12, as previously described ([54]). The gene sets comprising the B-vitamin/cofactor reconstruction (a total of 54 genes in *R. gnavus* TS_8243C and 65 in *C. symbiosum* TS_8243C) are essentially identical to those in the corresponding HMP reference type strains. These reconstructions are displayed as pathway diagrams in **Fig. S8C** and are briefly described below.

**Thiamin (vitamin B1)**

A common precursor of thiamin pyrophosphate (TPP), an essential cofactor of several ubiquitous enzymes in central carbon metabolism, most notably the pyruvate dehydrogenase complex. Neither *R. gnavus*, nor *C. symbiosum* have a complete set of enzymes for de novo synthesis of TPP, and thus should be considered B1 auxotrophs. Both strains contain the key thiamin salvage enzyme ThiN (thiamin pyrophosphokinase, EC 2.7.6.2). Different types of B1-specific transporters, ThiT (of ECF family ([55])) and ThiXYZ (of ABC family) could be identified in *R. gnavus*, but and *C. symbiosum*, respectively. Alternatively, or in addition to thiamin salvage, both strains are capable of synthesizing TPP using exogenous thiazole via the dedicated salvage enzyme ThiM (hydroxyethylthiazole kinase, EC 2.7.1.50). However, in this case, the respective thiazole transporter, ThiW ([55]) was confidently identified only in *C. symbiosum*. Both strains have the complete machinery for synthesis of the second biosynthetic intermediate (HMP-PP), joining it with thiazole-P, and downstream phosphorylation to yield the mature TPP cofactor.

The challenge of accurate homology-based functional assignment of transporters can often be effectively addressed by analyzing genomic context ([57], most importantly, operons and predicted regulons (sets of co-regulated genes). In this example, all of the aforementioned transporters in both *R. gnavus* TS_8243C and *C. symbiosum* TS_8243C were implicated as members of predicted regulons controlled by the TPP-specific riboswitch. RNA-based regulatory elements (riboswitches) are among the most common types of regulators of vitamin/cofactor metabolism that can be confidently mapped in genomes by computational methods due to their highly con-
served secondary structures (58). In addition to transporters, TPP-responding riboswitches regulate expression of nearly all biosynthetic enzymes (except ThiN, which is predicted to contribute to both de novo B1 synthesis and salvage routes); the repertoire of these co-regulated enzymes is conserved between *R. gnatus* and *C. symbiosum*.

**Riboflavin (vitamin B2)**

A precursor and biosynthetic intermediate of the essential redox cofactors flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) (involved in the electron transfer chain and several other essential processes). The complete canonical de novo riboflavin biosynthetic pathway, from GTP and D-ribulose-5-phosphate, is present in *C. symbiosum* TS_8243C but all respective genes are absent in *R. gnatus* TS_8243C. This is one of the two primary distinctions between these two strains, which are otherwise quite similar in terms of their B-vitamin proto- and auxotrophies. While this difference cannot be easily rationalized, *R. gnatus* TS_8243C, unlike *C. symbiosum* TS_8243C, lacks fumarate reductase, one of the primary flavin-dependent enzymes in central carbon metabolism (see above). Both strains have a committed riboflavin transporter RibU (55), which is essential for *R. gnatus* but likely dispensable for *C. symbiosum*, as well as an essential bi-functional enzyme RibF for converting riboflavin to FMN and then, FAD. A FMN-sensitive riboswitch is present to regulate RibU expression in both strains and de novo synthesis in *C. symbiosum* TS_8243C.

**Niacin (vitamin B3)**

Niacin, or nicotinic acid, as well as nicotinamide are converted to the most abundant redox cofactors, NAD and NADP; these cofactors are collectively involved in ~10% of all cellular biochemical transformations. Reconstruction of this subsystem in both *R. gnatus* TS_8243C and *C. symbiosum* TS_8243C includes both de novo synthesis of NAD from aspartate (prototrophy), and potential (optional) salvage of B3 and its amidated precursor. However, the B3 transporter NiaY (55) is present only in *C. symbiosum* TS_8243C. In the absence of additional data, it is impossible
to distinguish between two possible interpretations: (i) *R. gnavus* has a B3 transporter of another, yet unknown, type or (ii) the enzymes PncA and PncB are predominantly involved in recycling of intracellular nicotinamide, a well-known product of NAD degradation. The first possibility seems more likely since in both strains, all three genes of the upstream *de novo* pathway (from aspartate to NaMN) are predicted to be under the control of the NiaR repressor (based on the RegPrecise database), which would suppress this pathway when B3 is present (59).

**Pantothenate (vitamin B5)**

An intermediate in the biosynthesis of CoA, the universal and indispensable cofactor in a variety of acyl transfer reactions, most importantly in fatty acid biosynthesis (after transfer of pantetheine to acyl-carrier protein), degradation and the TCA cycle. Both strains lack components for *de novo* synthesis of pantothenate and, thus, would be considered B5 auxotrophs. However, auxotrophy is fully applicable only to *R. gnavus* TS_8243C, which has a complete set of downstream enzymes for the 5-step conversion of pantothenate to CoA. An interesting distinction of *C. symbiosum* TS_8243C is the absence of two enzymes CoaB and CoaC (typically a fusion protein in bacteria). This “gapped” pattern, which occurs in many diverse bacteria, was predicted (52) and later verified as a signature of the alternative salvage of pantetheine. This metabolite, a product of CoA degradation, can be phosphorylated (apparently via the second activity of pantothenate kinase), by-passing CoaBC-mediated steps and yielding a substrate for the next step catalyzed by CoaD. Therefore, *C. symbiosum* TS_8243C should be considered a pantetheine rather than a pantothenate auxotroph. This genomic signature is rather common among the HMP reference genomes, leading to a testable hypothesis that pantetheine should be present in human guts and function as prebiotic for some of members of the microbiota, such as *C. symbiosum* TS_8243C and raising the question about its source. We hypothesize that a homolog of the pantothenate symporter PanF, which is present in *C. symbiosum* instead of the PanT transporter as in *R. gnavus* TS_8243C, is a likely candidate for pantetheine transporter, since its canonical function would be obsolete in this organism.
Transcriptional regulation of CoA biosynthesis has not been described in any microbial system. Not surprisingly, our reconstruction of this pathway does not include any regulatory predictions.

**Pyridoxine (B6)**

A precursor but not a biosynthetic intermediate of the pyridoxal phosphate (PLP) cofactor that drives numerous transamination, decarboxylation and epimerization reactions of various amino acids. This pathway reveals a second major phenotypic distinction between *R. gnavus* TS_8243C and *C. symbiosum* TS_8243C. In contrast to the pattern with riboflavin (B2), *C. symbiosum* is a B6 auxotroph, whereas *R. gnavus* is a B6 prototroph. Both *C. symbiosum* and *R. gnavus* contain a B6 transporter (PdxU) (55), which is regulated by the PdxR repressor. We were unable to find a strong candidate gene encoding the enzyme PdxH, which oxidizes pyridoxine or pyridoxine phosphate to the functionally active aldehyde. Based on the pathway logics, it represents a classical “missing gene” problem (60), and thus a subject for further bioinformatic and experimental analyses. As in case of B5, essentially nothing is known and could be predicted about transcriptional regulation of B6/PLP metabolism in these two species.

**Biotin (B7)**

Biotin is the critical cofactor in transcarboxylation reactions (including the initiation of fatty acid synthesis); it performs its function after covalent attachment to carrier protein(s) by the universally conserved and essential enzyme BirA (biotin ligase). At first glance, both strains could be deemed B7 auxotrophs, and this is how they were described in the survey of 256 reference HMP genomes (54). Indeed, neither *R. gnavus* TS_8243C nor *C. symbiosum* has recognizable enzymes involved in the upstream steps of B7 biosynthesis, including the critical enzyme BioF (8-amino-7-oxononanoate synthase). Both strains contain the biotin transporter BioY (55) predicted to be under control of the same transcriptional repressor (the second function of the bifunctional protein BirA (61)). At the same time, both species also contain BioB pointing to dethiobiotin as a potential
alternative “surrogate” vitamin. As in case of B5, this genomic signature is rather widespread in HMP genomes suggesting that dethiobiotin may be yet another candidate prebiotic as well as another target for metabolomic and mechanistic analyses. Moreover, the presence of two additional enzymes, BioA and BioD, upstream of BioB, in C. symbiosum TS_8243C (and at least 50 other divergent species in the HMP collection of reference strains but not in R. gnavus TS_8243C) points to two other testable possibilities; (i) salvage of 8-amino-7-oxononanoate as yet another undocumented “surrogate” vitamin, and (ii) an alternative upstream pathway merging with the canonical route at the stage of 8-amino-7-oxononanoate. The first possibility appears more likely, and it is indirectly supported by the presence of the second BioY paralog in C. symbiosum (but not in R. gnavus) in the chromosomal neighborhood of the bioA-bioB operon, which may be considered a candidate for the 8-amino-7-oxononanoate transporter. If confirmed experimentally, it would point to yet another candidate for community vitamin exchange (and, potentially, a selective prebiotic). The second possibility would make C. symbiosum TS_8243C a B7 prototroph.

**Folate (B9)**

Folate is a precursor of folate-derived cofactors DHF, THF and its polyglutamate derivatives that play a critical role in single carbon metabolism, including numerous methylation reactions. While folate auxotrophy is relatively rare, and in many bacteria such as E. coli folate de novo synthesis cannot be compensated by exogenous folate, both TS_8243C strains are obvious B9 auxotrophs containing only downstream enzymes and the folate transporter FolT (55). This transporter in both strains and the FolC enzyme in C. symbiosum are under regulation of the THF-responding riboswitch (62).

**Cobalamin (B12)**

Cobalamin is involved in a limited number of very important reactions (e.g., methionine biosynthesis). However, in some B12-independent organisms these reactions are either not needed or are catalyzed by an alternative B12-independent enzyme (as in case of two distinct methionine
synthases, MetH and MetE). It is also the pathway with the largest number of committed biosynthetic steps. Therefore, it could be perceived as the first candidate for exclusion from the B-vitamin biosynthetic repertoire. Indeed, it is the only B-vitamin pathway missing in ~50% of analyzed Bacteroides spp. even though the exogenous supply and uptake of B12 was shown to be important for successful colonization of the respective auxotroph (63). In this light, it appears remarkable that both analyzed strains (each with multiple auxotrophies) have a complete B12 biosynthesis pathway. In addition, C. symbiosum TS_8243C (and not R. gnavus TS_8243C) contains the B12-specific transporter BtuFCD under regulation of a B12 riboswitch.
Supplementary Figure legends

Fig. S1. Distribution of ages at the time of sample collection of healthy microbiota donors that were used to generate the Random Forests-derived model of age-discriminatory bacterial 97%ID OTUs. Histogram of donor ages in the training set cohort (n=121 samples; minimum age 0.6 months, maximum age 33.51 months, median age 13.7 months) and testing set cohort (n=117 samples; minimum age 0.98 months, maximum age 31.8 months, median age 12.35 months). ns, no significant difference based on Kolmogorov-Smirnov test.

Fig. S2. Random Forests-derived model of lean body mass gain discriminatory OTUs. The 30 most lean mass gain discriminatory OTUs with their taxonomic assignments, ranked by feature importance as determined by mean squared error. Models were iterated 100 times, and average mean squared error values were used to generate the final rankings. Red indicates OTUs that appear within the 30 most discriminatory OTUs for both the weight and lean mass gain models. Purple indicates species that also appear in the 25-taxa sparse model of Malawian microbiota maturation. Bars to the right of the OTU ID numbers represent Spearman’s rank correlation of the same OTU ID to chronological age in the healthy Malawian infant/child cohort (Table S9).

Fig. S3. Effects of microbiota donor health status and age on femoral bone morphology. Micro-CT was performed on femurs harvested from gnotobiotic mice colonized with microbiota from 6- and 18-month-old healthy or undernourished donors (in all cases, the efficiency of transplantation was >50%). Each circle represents a single mouse. (A) Bone morphologic phenotypes expressed as a function of donor nutritional status irrespective of donor age. Differences in the ratio of bone volume to tissue volume (BV/TV) and volumetric bone mineral density (vBMD) in the cortical region of femurs are shown. (B) Bone morphologic phenotypes expressed as a function of age, irrespective of donor nutritional status. p-values are based on a Mann-Whitney test.

Fig. S4. Random Forests-derived models of OTUs that discriminate femoral bone metrics defined by micro-CT. Top 20 most discriminatory OTUs, ranked by mean squared error values, for (A) cancellous bone mineral density, (B) cancellous bone volume per tissue volume, (C) corti-
cal volumetric bone mineral density, (D) cortical bone volume per tissue volume, and (E) cortical thickness. Inserts show cross-validation of each model; error decreases as OTUs are added to the model. Red indicates that the OTU is also among the top 30 ranked OTUs the Random Forests-derived models of both weight- and lean mass gain-discriminatory taxa. Boldface indicates that the OTU appears in all five bone metric models.

Fig. S5. Discordant growth phenotypes transmitted by two exemplary iLiNS-DYAD-M donor gut communities: one from a healthy 6-month-old donor with normal gut microbiota maturity and the other from a severely stunted/underweight 6-month-old donor with significant microbiota immaturity. (A) Gain of lean body mass (n=5 recipients/donor microbiota). Mean±SEM values are shown. p<0.0001 (2-way ANOVA). (B) Bacterial composition of the fecal microbiota of gnotobiotic mice in the co-housing experiments (relative abundances, averaged over all time points, of taxa from all animals in each treatment group). Abbreviations: ‘screen’, initial survey of mice harboring donor microbiota; ‘H’, cagemates colonized with the healthy donor’s microbiota in each co-housing experiment; ‘Un’, cagemates colonized with the severely stunted/underweight donor’s microbiota.

Fig. S6. Differences in brain metabolism in gnotobiotic recipients of 6-month-old healthy compared to severely stunted/underweight donor microbiota. (A) Metabolic profile of the cerebral cortex recovered at sacrifice. Columns represent individual animals. Rows represent metabolites whose concentrations are significantly different between the two treatment groups (p<0.05 for all metabolites shown, Student’s t-test). Rows are hierarchically clustered based on pairwise distances calculated from Pearson correlation values. (B) Concentrations of amino acids in the cerebral cortex of mice colonized with the healthy (H) or stunted/underweight (Un) donor microbiota (mean±SEM values plotted; p<0.05 for all metabolites shown; Student’s t-test).

Fig. S7. Spearman’s rank correlation rho values for bacterial species relative abundances in the fecal microbiota versus concentrations of metabolites in gnotobiotic mouse recipients of healthy and severely stunted/underweight infant donor gut communities. Spearman correla-
tions involving (A) acylcarnitines in serum, (B) amino acids in liver and gastrocnemius muscle, and (C) organic acids in liver and gastrocnemius muscle. Rows are hierarchically clustered based on pairwise distances calculated from Pearson correlation values. Bars to the right of each heatmap indicate significant species enrichment (p<0.05; Student’s t-test) in the fecal microbiota of mice colonized with the healthy (purple) or stunted/underweight (yellow) donor’s gut community.

**Fig. S8.** Reconstructed *Ruminococcus gnavus* TS_8243C and *Clostridium symbiosum* TS_8243C metabolic pathways involved in carbohydrate and central carbon metabolism, amino acid biosynthesis and degradation, and vitamin and cofactor biosynthesis. Enzymatic reactions and transporters are depicted by solid and dashed lines, respectively. Red and purple lines indicate the presence of the enzyme or transporter in the cultured strains of *R. gnavus* and *C. symbiosum*, respectively, while gray lines indicate absence of the enzyme or transporter in either strain. Common names of enzymes and transporters are shown in black and blue font, respectively. Enzymes are also identified by their enzyme commission (EC) numbers. Major families of carbohydrate transporters indicated in parentheses are ATP-Binding Cassette (ABC) transporters, Major Facilitator Superfamily (MFS) permeases, TRipartite ATP-independent Periplasmic (TRAP) transporters, and phosphoenolpyruvate-dependent PhosphoTransferase System (PTS) transporters. Cognate transcriptional regulators of sugar utilization pathways are highlighted in yellow ovals. Central carbon metabolism genes controlled by the global redox-responsive regulator Rex in both genomes are marked by yellow vertical stripes. Genes involved in amino acid and vitamin metabolism that are controlled by local regulators and metabolite-responsive riboswitches are denoted by colored circles. Gene locus tags and functional annotations are listed in Table S16. Genes with novel predicted functions (see Supplementary Results) are marked with an asterisk. Putative pathways and transporters for which no gene annotation could be identified are indicated with a question mark.
Supplementary Figures

Fig. S1.

![Bar chart showing the distribution of samples across different ages for training and testing sets. The x-axis represents age in months (0, 4, 8, 12, 16, 20, 24, 28, 32) and the y-axis represents the number of samples. The chart includes two bars for each age, one black and one gray, indicating the training and testing sets respectively. The legend indicates the number of samples in each set: Training (n = 121 samples) and Testing (n = 117 samples). The significance level is indicated as ns (not significant).]
Fig. S2.

Increasing importance in accuracy of model

Rank  OTU ID  Taxonomic annotation  Lean mass gain indicative taxa
1  de novo 12617  None
2  4374302  Dorea formicigenerans
3  3203801  Clostridium
4  3709990  Clostridium innocuum
5  72820  Biﬁdobacterium longum
6  238205  Clostridiurn butyricum
7  2424737  Clostridium
8  4448492  Biﬁdobacterium longum
9  2724175  Ruminococcus gnavus
10  2148366  Raeni bacillus
11  2575651  Ruminococcus gnavus
12  2714942  Ruminococcus gnavus
13  64384  Weissella cibaria
14  138389  Enterococcus
clostridiales
15  4473509  Clostridium
clostridiales
16  242298  Clostridium
clostridiales
17  4383953  Pediococcus pentosaceus
18  773251  Enterococcus faecalis
19  1033413  Pediococcus
20  4380976  Eggerthella
21  1141218  Enterococcus
clostridiales
22  4316928  Finegoldia magna
23  1096610  Clostridium
clostridiales
24  4378683  Clostridium
clostridiales
25  342397  Clostridium
clostridiales
26  681370  Clostridium
clostridiales
27  265871  Clostridium
clostridiales
28  195436  Clostridium
clostridiales
29  3600504  Clostridium
clostridiales
30  1880112  Clostridium
clostridiales

Spearman’s correlation of OTU to age in healthy Malawi cohort

minimum ρ (-0.5) maximum ρ (0.5)
p-value <0.0001 0.4964

Increase in Mean Squared Error
Fig. S3.

A

Cortical BV/TV

mg HA/ccm

Healthy Undernourished

0.25 0.30 0.35 0.40 0.45

p = 0.05

Cortical vBMD

mg HA/ccm

Healthy Undernourished

300 350 400 450 500

p = 0.07

B

Trabecular BMD

mg HA/ccm

6 months 18 months

250 200 150 100 50

p = 0.0005

Trabecular BV/TV

mg HA/ccm

6 months 18 months

0.4 0.3 0.2 0.1 0

p < 0.0001

Trabecular Connectivity

mg HA/ccm

6 months 18 months

600 400 200 0

p < 0.0001

Trabecular Number

mg HA/ccm

6 months 18 months

9 8 7 6 5

p < 0.0001

Trabecular Spacing

mg HA/ccm

6 months 18 months

0.20 0.18 0.16 0.14 0.12 0.10

p < 0.0001
Fig. S4.

A) Cancellous BMD

<table>
<thead>
<tr>
<th>Rank</th>
<th>OTU ID</th>
<th>Rank</th>
<th>OTU ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4664032</td>
<td>11</td>
<td>4360353</td>
</tr>
<tr>
<td>2</td>
<td>4395052</td>
<td>12</td>
<td>3799990</td>
</tr>
<tr>
<td>3</td>
<td>4338564</td>
<td>13</td>
<td>4307714</td>
</tr>
<tr>
<td>4</td>
<td>188947</td>
<td>14</td>
<td>4316143</td>
</tr>
<tr>
<td>5</td>
<td>4309282</td>
<td>15</td>
<td>4307714</td>
</tr>
<tr>
<td>6</td>
<td>4234360</td>
<td>16</td>
<td>4338564</td>
</tr>
<tr>
<td>7</td>
<td>3903651</td>
<td>17</td>
<td>4383953</td>
</tr>
<tr>
<td>8</td>
<td>2148366</td>
<td>18</td>
<td>4383953</td>
</tr>
<tr>
<td>9</td>
<td>3197419</td>
<td>19</td>
<td>4316143</td>
</tr>
<tr>
<td>10</td>
<td>3451202</td>
<td>20</td>
<td>4395052</td>
</tr>
</tbody>
</table>

B) Cancellous BV/TV

<table>
<thead>
<tr>
<th>Rank</th>
<th>OTU ID</th>
<th>Rank</th>
<th>OTU ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4363532</td>
<td>11</td>
<td>3197419</td>
</tr>
<tr>
<td>2</td>
<td>3197419</td>
<td>12</td>
<td>4383953</td>
</tr>
<tr>
<td>3</td>
<td>4307714</td>
<td>13</td>
<td>4383953</td>
</tr>
<tr>
<td>4</td>
<td>4383953</td>
<td>14</td>
<td>2148366</td>
</tr>
<tr>
<td>5</td>
<td>2148366</td>
<td>15</td>
<td>4316143</td>
</tr>
<tr>
<td>6</td>
<td>4307714</td>
<td>16</td>
<td>4316143</td>
</tr>
<tr>
<td>7</td>
<td>4395052</td>
<td>17</td>
<td>4395052</td>
</tr>
<tr>
<td>8</td>
<td>4316143</td>
<td>18</td>
<td>4395052</td>
</tr>
<tr>
<td>9</td>
<td>191153</td>
<td>19</td>
<td>4363532</td>
</tr>
<tr>
<td>10</td>
<td>4363532</td>
<td>20</td>
<td>2148366</td>
</tr>
</tbody>
</table>

C) Cortical vBMD

<table>
<thead>
<tr>
<th>Rank</th>
<th>OTU ID</th>
<th>Rank</th>
<th>OTU ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4664032</td>
<td>11</td>
<td>4395052</td>
</tr>
<tr>
<td>2</td>
<td>4360353</td>
<td>12</td>
<td>3799990</td>
</tr>
<tr>
<td>3</td>
<td>4338564</td>
<td>13</td>
<td>4307714</td>
</tr>
<tr>
<td>4</td>
<td>188947</td>
<td>14</td>
<td>4316143</td>
</tr>
<tr>
<td>5</td>
<td>4309282</td>
<td>15</td>
<td>4307714</td>
</tr>
<tr>
<td>6</td>
<td>4234360</td>
<td>16</td>
<td>4338564</td>
</tr>
<tr>
<td>7</td>
<td>3903651</td>
<td>17</td>
<td>4383953</td>
</tr>
<tr>
<td>8</td>
<td>2148366</td>
<td>18</td>
<td>4383953</td>
</tr>
<tr>
<td>9</td>
<td>3197419</td>
<td>19</td>
<td>4316143</td>
</tr>
<tr>
<td>10</td>
<td>3451202</td>
<td>20</td>
<td>4395052</td>
</tr>
</tbody>
</table>

D) Cortical BV/TV

<table>
<thead>
<tr>
<th>Rank</th>
<th>OTU ID</th>
<th>Rank</th>
<th>OTU ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4664032</td>
<td>11</td>
<td>3197419</td>
</tr>
<tr>
<td>2</td>
<td>3197419</td>
<td>12</td>
<td>4383953</td>
</tr>
<tr>
<td>3</td>
<td>4307714</td>
<td>13</td>
<td>4383953</td>
</tr>
<tr>
<td>4</td>
<td>4383953</td>
<td>14</td>
<td>2148366</td>
</tr>
<tr>
<td>5</td>
<td>2148366</td>
<td>15</td>
<td>4316143</td>
</tr>
<tr>
<td>6</td>
<td>4307714</td>
<td>16</td>
<td>4316143</td>
</tr>
<tr>
<td>7</td>
<td>4395052</td>
<td>17</td>
<td>4395052</td>
</tr>
<tr>
<td>8</td>
<td>4316143</td>
<td>18</td>
<td>4395052</td>
</tr>
<tr>
<td>9</td>
<td>191153</td>
<td>19</td>
<td>4363532</td>
</tr>
<tr>
<td>10</td>
<td>4363532</td>
<td>20</td>
<td>2148366</td>
</tr>
</tbody>
</table>

E) Cortical thickness

<table>
<thead>
<tr>
<th>Rank</th>
<th>OTU ID</th>
<th>Rank</th>
<th>OTU ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4664032</td>
<td>11</td>
<td>4395052</td>
</tr>
<tr>
<td>2</td>
<td>4360353</td>
<td>12</td>
<td>3799990</td>
</tr>
<tr>
<td>3</td>
<td>4338564</td>
<td>13</td>
<td>4307714</td>
</tr>
<tr>
<td>4</td>
<td>188947</td>
<td>14</td>
<td>4316143</td>
</tr>
<tr>
<td>5</td>
<td>4309282</td>
<td>15</td>
<td>4307714</td>
</tr>
<tr>
<td>6</td>
<td>4234360</td>
<td>16</td>
<td>4338564</td>
</tr>
<tr>
<td>7</td>
<td>3903651</td>
<td>17</td>
<td>4383953</td>
</tr>
<tr>
<td>8</td>
<td>2148366</td>
<td>18</td>
<td>4383953</td>
</tr>
<tr>
<td>9</td>
<td>3197419</td>
<td>19</td>
<td>4316143</td>
</tr>
<tr>
<td>10</td>
<td>3451202</td>
<td>20</td>
<td>4395052</td>
</tr>
</tbody>
</table>
Fig. S5.

**A**

Lean mass gain

<table>
<thead>
<tr>
<th>% Initial lean mass</th>
<th>Healthy</th>
<th>Stunted/underweight</th>
</tr>
</thead>
</table>

Days post colonization

**B**

Relative abundance (%)

- Actinomycetes radingae
- Bacteroides
- Bifidobacterium
- Bifidobacterium bifidum
- Bifidobacterium longum
- Blautia producta
- Clostridiales
- Clostridium
- Clostridium barbetti
- Clostridium botale
- Clostridium butyricum
- Clostridium colinum
- Clostridium difficile
- Clostridium disporicum
- Clostridium glycolicum
- Clostridium rutiliewayi
- Clostridium innocuum
- Clostridium neonatale
- Clostridium nuxile
- Clostridium paraputricum
- Clostridium perfringens
- Clostridium ramosum
- Clostridium symbiosum
- Clostridium xylanolyticum
- Coprobacillus cateniformis
- Dorea formigenerans
- Eggerthella lenta
- Enterobacteriaceae
- Enterococcaceae
- Enterococcus
- Enterococcus faecalis
- Escherichia coli
- Eubacterium hallii
- Eubacterium limosum
- Eubacterium rectale
- Faecalibacterium prausnitzii
- Gordonibacter pamelaeae
- Lactobacillus
- Lactobacillus ruminis
- Leuconostoc lactis
- None
- Paenibacillus
- Paenibacillus sp P15 9
- Paenibacillus taiwanensis
- Paenibacillus xylanilyticus
- Pediococcus
- Pediococcus pentosaceus
- Ruminococcus gravis
- Streptococcus
c- Streptococcus galloyticus
- Turicibacter sanguinis
- Vagococcus
- Weissella
- Weissella cibaria

* Sparse model age-discriminatory taxon
* Within top 30 of weight-discriminatory taxa
* Within top 30 of lean mass gain-discriminatory taxa
Fig. S6.
Fig. S7.

A Acylcarnitines

B Amino acids

C Organic acids
Fig. S8A.

Central carbohydrate metabolism and fermentation
Fig. S8C.

**Vitamin and Cofactor Biosynthesis**

**Aminimidazole ribotide (AIR)**
- CoA
- Cysteine
- Tysine
- GTP
- AMP
- PP
- ATP
- TPi
- Thiamine pyrophosphate (TPP)
- Thiamin
- Aspartate
- Niacinophosphate (NAD)
- Niacinophosphate (NADP)
- Malonyl-ACP

**Cobalamin**
- Pyridoxal phosphate
- Pyridoxal
- Pyridoxamine
- Pyridoxin

**THF-polyglutamate**
- Folate
- 7,8-Dihydrofolate
- 7,8-Dihydroneopterin

**B12 riboswitch**
- TPP riboswitch
- THF riboswitch
- FMN riboswitch
- BirA repressor
- NaiR repressor
- PbxR repressor

**Transcriptional regulation by:**
- THF riboswitch
- TPP riboswitch
- FMN riboswitch
- B12 riboswitch
- BirA repressor
- NaiR repressor
- PbxR repressor

**Genomes:**
- Ruminococcus gnavus TS_8243c
- Clostridium symbiosum TS_8243c

**Function/Localization:**
- Biosynthetic enzymes
- Degradation enzymes
- Transporters
Supplementary Data Tables

Table S1. Summary of the ages, anthropometry, and time of fecal collection from Malawian cohorts. (A) Concordant healthy Malawian twins pairs/triplet sets. (B) Members of the iLiNS-DYAD-M birth cohort.

Table S2. Summary of bacterial V4-16S rRNA datasets from human donors. (A) Malawi Twin cohort. (B) iLiNS-DYAD-M cohort.

Table S3. Random Forests-derived sparse model of microbiota maturation in healthy Malawian infants/children. (A) Rank order of OTU feature importance. (B) Comparison of OTUs comprising the sparse Malawian and sparse Bangladeshi models.

Table S4. V4-16S rRNA-based characterization of fecal microbiota sampled from transplant recipients. (A) Maturity indices for microbiota selected for transplantation into gnotobiotic mice. (B) Summary of 16S rRNA datasets generated from the fecal microbiota of recipient mice in the initial screen of 19 donor microbiota. (C) Summary of 16S rRNA datasets generated from the follow-up transplant studies using microbiota from a 6-month-old healthy and a 6-month-old severely stunted/underweight donor.

Table S5. Composition of the Malawi-8 (M8) diet. (A) Summary of ingredients. (B) Nutritional analysis.

Table S6. Summary of growth phenotypes in recipients of the 19 microbiota from 6- and 18-month-old healthy and undernourished Malawian donors. (A) Weight gain. (B) Lean body mass gain and fat body mass gain.

Table S7. Consumption of M8 chow in each donor group of gnotobiotic recipient mice from the screen of 19 donor microbiota.

Table S8. Rank order list of feature importance scores of OTUs in Random Forest models based on the growth phenotypes of gnotobiotic mice with 19 different transplanted microbiota from 6- and 18-month-old Malawian donors. (A) Weight gain. (B) Lean body mass gain.
Table S9. Spearman’s rank correlations to mouse growth phenotypes and to chronological age in 238 members of concordant healthy Malawian twin pairs/triplet sets and 259 singleton members of iLiNS-DYAD-M. (A) Weight gain. (B) Lean body mass gain.

Table S10. Rank order list of OTU feature importance scores in Random Forest models based on the femoral bone metrics of groups of recipient mice with 19 different transplanted microbiota from 6- and 18-month-old Malawian donors.

Table S11. List of OTUs and their relative abundances used to construct the bargraphs shown in Fig. S4B.

Table S12. Metabolic phenotypes of gnotobiotic mice colonized with a normally mature microbiota from a 6-month-old healthy (H) infant and a significantly immature microbiota from a 6-month-old severely stunted/underweight (Un) donor.

Table S13. Spearman’s rank correlations between levels of metabolites and the abundances of bacterial species in the fecal microbiota of mice colonized with the healthy versus the severely stunted/underweight infant microbiota. (A) Serum acylcarnitines. (B) Amino acids in liver and gastrocnemius muscle. (C) Organic acids assayed in liver and gastrocnemius muscle.

Table S14. Direction and success of invasion in co-housing experiments involving $H^{CH}$ and $Un^{CH}$ mice.

Table S15. List of OTUs and their relative abundances (%) used to construct the bargraphs shown in Fig. 5C.

Table S16. Genes included into the reconstructed metabolic pathways for *Ruminococcus gnavus* TS_8243C and *Clostridium symbiosum* TS_8243C based on subsystem-based comparative genomic reconstructions. (A) Overall number of genes in the reconstructed metabolic pathways of the two strains. The three major functional categories of studied genes are those involved in (i) carbohydrate and central carbon metabolism, (ii) amino acid metabolism and (iii) vitamin and cofactor biosynthesis. (B,C). Detailed lists of genes in *R. gnavus* TS_8243C and *C. symbio-
sum TS_8243C, and their orthologs in the type strains, R. gnarus ATCC 29149 and C. symbiosum WAL-14163. Enzymes are identified by their enzyme commission (EC) numbers. Functional descriptions for carbohydrate active enzymes (CAZymes) include their assigned glycoside hydrolase (GH) families and their predicted localization (intracellular versus secreted as defined by SignalP). Transcriptional regulators and riboswitches that are predicted to control expression of individual genes from the reconstructed pathways (according to conservative propagation in the RegPrecise database) are listed in the last column. Central carbon metabolism genes controlled by the global redox-responsive regulator Rex are also indicated. Genes with novel predicted functions (see Supplementary Results) are marked with an asterisk.

Table S17. Metabolic phenotypes, as defined by targeted mass spectrometry, of gnotobiotic mice colonized with the microbiota from the severely stunted/underweight infant donor or mice colonized with the donor’s microbiota together with R. gnarus TS_8243C and C. symbiosum TS_8243C.
Chapter 3

Sialylated milk glycans promote growth in gnotobiotic mice and piglets with a stunted Malawian infant gut microbiota
Chapter 3

Sialylated milk glycans promote growth in gnotobiotic mice and piglets with a stunted Malawian infant gut microbiota

Mark R. Charbonneau\textsuperscript{1,2}, David O’Donnell\textsuperscript{1,2}, Laura V. Blanton\textsuperscript{1,2}, Sarah M. Totten\textsuperscript{4,5}, Jasmine C. C. Davis\textsuperscript{4,5}, Michael J. Barratt\textsuperscript{1,2}, Jiye Cheng\textsuperscript{1,2}, Janaki Guruge\textsuperscript{1,2}, Michael Talcott\textsuperscript{6}, James R. Bain\textsuperscript{7,8,9}, Michael J. Muehlbauer\textsuperscript{7}, Olga Ilkayeva\textsuperscript{7}, Chao Wu\textsuperscript{10}, Tedd Struckmeyer\textsuperscript{10}, Daniela Barile\textsuperscript{3,5}, Charles Mangani\textsuperscript{11}, Josh Jorgensen\textsuperscript{12}, Yue-mei Fan\textsuperscript{13}, Kenneth Maleta\textsuperscript{11}, Kathryn G. Dewey\textsuperscript{12}, Per Ashorn\textsuperscript{13,14}, Christopher B. Newgard\textsuperscript{7,8,9}, Carlito Lebrilla\textsuperscript{4,5}, David A. Mills\textsuperscript{3,5}, Jeffrey I. Gordon\textsuperscript{1,2}

\textsuperscript{1}Center for Genome Sciences and Systems Biology, and \textsuperscript{2}Center for Gut Microbiome and Nutrition Research, Washington University School of Medicine, St. Louis, MO 63108, USA

\textsuperscript{3}Department of Food Science and Technology, \textsuperscript{4}Department of Chemistry, and \textsuperscript{5}Foods for Health Institute, University of California at Davis, Davis CA, 95616, USA

\textsuperscript{6}Division of Comparative Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA

\textsuperscript{7}Sarah W. Stedman Nutrition and Metabolism Center and Duke Molecular Physiology Institute,\textsuperscript{8}Department of Medicine and \textsuperscript{9}Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, NC 27710, USA

\textsuperscript{10}Hilmar Ingredients, Hilmar Cheese Company, Hilmar, CA 95324, USA

\textsuperscript{11}College of Medicine, University of Malawi, Blantyre, Malawi

\textsuperscript{12}Department of Nutrition, University of California, Davis, California, USA

\textsuperscript{13}Department of International Health, School of Medicine, University of Tampere, Tampere, Finland

\textsuperscript{14}Department of Paediatrics, Tampere University Hospital, Tampere, Finland

Correspondence to: jgordon@wustl.edu
Summary

Identifying interventions that more effectively promote healthy growth of children with undernutrition is a pressing global health goal. Analysis of human milk oligosaccharides (HMOs) from 6-month postpartum mothers in two Malawian birth-cohorts revealed that sialylated HMOs are significantly less abundant in mothers with severely stunted infants. To explore this association, we colonized young germ-free mice with a consortium of bacterial strains cultured from the fecal microbiota of a 6-month-old stunted Malawian infant and fed recipient animals a prototypic Malawian diet with or without purified sialylated bovine milk oligosaccharides (S-BMO). S-BMO produced a microbiota-dependent augmentation of lean body mass gain, changed bone morphology and altered liver, muscle and brain metabolism in ways indicative of a greater ability to utilize nutrients for anabolism. These effects were also documented in gnotobiotic piglets using the same consortium and Malawian diet. These preclinical models establish a causal microbiota-dependent relationship between S-BMO and growth promotion.

Introduction

The global burden of childhood undernutrition is great, causing 3.1 million deaths annually and accounting for 45% of all deaths under five years of age in 2011 (Black et al, 2013). Linear growth faltering, or stunting, defined by a height-for-age Z-score (HAZ) below two standard deviations from the mean for a reference international cohort of children with healthy growth phenotypes, is the most prevalent form of undernutrition, affecting ~165 million children under five years. Stunting has lifelong consequences beyond reduced height, including impaired intellectual development (Prendergast and Humphrey, 2012; Victora et al, 2008). The causes of stunting are not fully understood but are believed to include environmental factors together with genetic and epigenetic components (Martorell and Zongrone, 2012). Current interventions in the first 2 years of life have largely focused on education, improved hygiene and provision of micronutrient-fortified complementary foods during the weaning period. The effects of these approaches on growth, immune function and neurodevelopmental outcomes have been modest (Dewey and Adu-Afarwuah, 2008).
Recent culture-independent studies of the bacterial composition of fecal samples collected serially from infants and children living in Malawi and Bangladesh have demonstrated that the normal pattern of microbiota assembly is disrupted in children with undernutrition. The resulting microbiota immaturity is not repaired with current therapeutic food interventions leading to the proposal that disrupted microbiota development impairs healthy postnatal growth (Subramanian et al, 2014).

The World Health Organization (WHO) recommends exclusive breastfeeding for 6 months (Kramer and Kakuma, 2002). Human milk is composed of lactose, lipids, micronutrients, proteins and many other bioactive substances including a diverse repertoire of free and conjugated human milk oligosaccharides (HMOs) that may be decorated with fucose and/or sialic acid moieties (Coppa et al, 2011; Smilowitz et al, 2014). HMOs are not absorbed in the proximal gut (Engfer et al, 2000) and function as prebiotics that promote colonization of the infant gut with bacterial strains, including members of the genus *Bifidobacterium*, which are associated with numerous benefits [e.g., improved vaccine responses (Huda et al, 2014), enhanced gut barrier function (Ehaschuk et al, 2008; Weng et al, 2014) and protection from enteropathogen infection (Fukuda et al, 2011)].

To date, the relationship between the overall abundance of breast milk HMOs (and constituent HMO classes) and the growth phenotypes in infants, particularly in low-income countries with populations where undernutrition is pervasive, has not been well characterized. The current study addresses this issue by first characterizing breast milk HMO content in mothers, enrolled in two birth cohort studies conducted in rural southern Malawian villages, whose 6-month-old infants either exhibited healthy growth or were severely stunted. Differences identified in their breast milk HMO profiles were then translated to preclinical tests of whether one difference, involving sialylated HMOs, was causally related to the infants’ growth phenotypes. These tests were conducted in two gnotobiotic animal models; young recently weaned mice and newborn piglets, a mammalian species with greater physiologic similarity to humans. In both cases, animals were colonized with a collection of sequenced bacterial strains from a 6-month-old Malawian infant and fed a representative Malawian diet with or without a purified preparation of sialylated bovine milk oligosaccharides (S-BMO; structurally similar to sialylated HMO). The results not only establish
that this S-BMO preparation produces a microbiota-dependent promotion of growth and metabolic changes indicative of improved nutrient utilization in both species harboring the stunted human donor’s gut community members, but illustrate a generalizable preclinical pathway for characterizing the effects of new lead or existing compounds for diet-based interventions designed to treat and ultimately prevent undernutrition and its co-morbidities in infants.

**Results**

**Human milk oligosaccharide content correlates with growth in Malawian infants**

Breast milk samples were collected from Malawian mothers at 6 months postpartum as part of the Lungwena Child Nutrition Intervention Study #5 (LCNI-5; Mangani et al, 2013). Samples were selected for HMO analysis from mothers (*n*=88) whose children exhibited either healthy growth at the time of breast milk collection [Height-for-Age Z-score (HAZ) > 0; *n*=29 mothers] or severe stunting (HAZ < -3; *n*=59). Liquid chromatography time-of-flight mass spectrometry (LC-TOF MS) detected over 50 unique oligosaccharide structures among the samples (**Table S1A**). HMOs containing sialic acid and/or fucose residues represented 71.8±1.5% (mean ± SEM) of total HMOs (**Table S1A**). Mothers of healthy infants had significantly higher concentrations of total, sialylated and fucosylated HMOs than mothers of severely stunted infants (**Figure 1A**). The most growth discriminatory sialylated HMO for the entire sampled LCNI-5 cohort was sialyllacto-\(N\)-tetraose b (LSTb), while the most discriminatory fucosylated HMOs were 2’-fucosyllactose (2’FL) and lacto-\(N\)-fucopentaose I (LNFP I) (**Table S1A**).

Mothers with a functional *FUT2* gene produce α1-2 linked fucosylated HMOs (‘secretor phenotype’; Newburg et al, 2005). The representation of mothers who were non-secretors (owing to a non-functional *FUT2*) was 21.6%. As expected, the total and fucosylated HMO content of breast milk samples collected from secretor mothers in this cohort (*n*=69) was significantly higher than that of non-secretor mothers [*n*=19; *p*<0.05, two-way ANOVA (**Table S1A**)]. Among non-secretor mothers, those whose children were severely stunted at 6 months of age had HMO profiles
that were significantly deficient in fucosylated and sialylated HMOs compared to milk from mothers of healthy growing infants (Figure 1A). The most discriminatory HMOs were the sialylated glycan, LSTb, as well as the recently reported sialylated 4021b species [Neu5Ac(α2-6)Gal(β1-4) GlcNAc(β1-3)Gal(β1-4)GlcNAc(β1-3)Gal(β1-4)Glc (Table S1A); Wu et al, 2010]. Among secreter mothers, no significant differences in HMO abundances were observed between those with healthy infants and those with stunted infants (Table S1A). One interpretation of these data is that some non-secretor Malawian mothers are unable to compensate for deficiencies in fucosylated HMOs through increased output of other (e.g., sialylated) HMOs, resulting in breast milk that is less supportive of healthy infant growth.

We extended our analysis of breast milk HMOs to mothers enrolled in a second birth cohort study (iLiNS-DYAD-M) conducted in rural Malawi (Ashorn et al., 2013). The target population for enrollment included healthy pregnant women at less than 20 weeks of gestation living in the Mangochi District of Malawi. Breast milk samples (n=215) were selected for HMO analysis based on infant HAZ scores at 6 months postpartum (HAZ < -2 or HAZ > 0). Similar to LCNI-5 mothers, we found that total and sialylated HMO content was significantly elevated for mothers of healthy infants (HAZ > 0; n=70) compared to mothers of stunted infants [HAZ < -2; n=145 (Figure 1B, Table S1B)]. Fucosylated HMO content was not significantly enriched for mothers of healthy infants in the iLiNS-DYAD-M cohort.

For this second cohort, we also observed that breast milk from non-secretor mothers of healthy infants (n=20) exhibited significantly higher levels of total, fucosylated and sialylated HMOs than milk of non-secretor mothers of stunted infants (n=40; Figure 1B). While milk samples from secretor mothers displayed no statistically significant differences in total or fucosylated HMOs between infant anthropometry bins (Table S1B), as observed with LCNI-5, levels of sialylated HMOs were significantly elevated in milk from those with healthy infants (n=50) compared to those with stunted infants (n=105; p<0.01, Welch’s t-test). Taken together, these data from two independent clinical studies in Malawi raise the possibility that HMOs may play an important role in infant growth.
Considerations for testing the physiological effects of HMOs in gnotobiotic models

We reasoned that the hypothesis that HMOs are causally related to healthy growth in Malawian infants could be tested in gnotobiotic animals colonized with members of the gut microbiota of a stunted Malawian infant and fed a Malawian diet representative of foods consumed after weaning by this population. The diets could then be supplemented with purified bovine milk-derived oligosaccharides (BMOs) that are analogous to HMOs. Fucosylated oligosaccharides are poorly represented in bovine milk. However, whole bovine milk contains sialylated BMOs that could be purified as a source for gnotobiotic animal experiments (Aldredge et al, 2013; Sundekilde et al, 2012).

We also postulated that colonizing germ-free animals with a genome-sequenced bacterial culture collection, generated from the fecal microbiota of an individual infant donor representing a human population and physiologic state of interest would be valuable for a number of reasons. First, members of the collection have co-evolved in the same environment and represent a snapshot of the microbial exposures experienced by the host, including inheritance of microbes from the mother. Second, the strains represented in that individual’s microbiota reflect selection placed on the microbiota of humans from a given geographic region by their dietary practices. Third, the culture collection in its entirety, or subsets of the collection, can be introduced into groups of recipient gnotobiotic animals in order to define the contributions of bacterial strains to host biology; this is not possible when transplanting intact, uncultured communities.

A clonally arrayed, genome-sequenced collection of bacterial strains from the fecal microbiota of a stunted Malawian infant

With the above considerations in mind, we generated a clonally arrayed collection of anaerobically cultured bacterial strains from the fecal microbiota of a severely stunted 6-month-old Malawian infant (HAZ: -3.48, weight-for-height Z-score: 0.05). Using a threshold of ≥96% genome-wide nucleotide sequence similarity to distinguish isolates in the arrayed culture collection as representing a given strain, we determined that the collection was composed of 25 distinct strains [44.65 ±
29.32-fold (mean ± SD) genome sequencing coverage; see Table S2A for statistics related to these de novo assembled draft genomes]. The bacterial composition of the uncultured fecal microbiota from which the culture collection was derived was characterized by sequencing 16S rRNA gene amplicons (V4 region) from that sample. Examination of the OTUs present at >0.1% relative abundance in the uncultured fecal sample revealed that 77.8% of the bacterial diversity detected in the intact community (97% ID OTUs, weighted by relative abundance) was represented in the culture collection (Figure 2A).

The culture collection included four species of Bifidobacterium that are common members of fecal microbiota sampled from breast fed infants, including B. bifidum, B. breve, B. catenulatum and two strains of B. longum subspecies infantis. The presence of virulence factors in members of the collection was determined by mapping draft genomes to the Virulence Factor Database (Chen et al, 2012). This analysis indicated that 16 strains possessed at least one putative virulence factor, and eight had more than three (Table S2B). We noted that the Salmonella enterica strain possessed homologs to 170 of the 176 virulence factors (96.6%) found in S. enterica subsp. enterica serovar Typhimurium, a pathogen in both humans and mice. The sole E. coli strain contained two key virulence factors, aggR and aap, associated with enteroaggregative E. coli.

Bacterial strains isolated directly from an individual can vary substantially from laboratory strains both in terms of their genome content and biological properties. Therefore, we compared each strain in the culture collection to the NCBI type strain harboring the most similar full-length 16S rRNA sequence. The full-length 16S rRNA amplicon sequences generated from 22 of the 25 strains shared greater than 97%ID with their closest assignable reference type strain’s 16S rRNA gene sequence (Figure 2B). In contrast, whole genome alignment revealed that only six isolates shared greater than 90% overall genome nucleotide sequence similarity, none had greater than 93% and three shared less than 40% with their reference type strain (Figure 2B). These observations highlight the importance of using primary bacterial isolates from populations of interest in preclinical studies using gnotobiotic models.
S-BMO supplementation of a prototypic Malawian diet improves growth in young mice colonized with the stunted Malawian donor’s cultured bacterial community

Since large-scale purification of HMOs from human milk is not feasible, we purified a monosaccharide and lactose-free mixture of sialylated bovine milk oligosaccharides (S-BMO) from a commercial cheese whey stream. 3’- and 6’-sialyllactose accounted for 88% of the oligosaccharide content of S-BMO; the remaining oligosaccharides were predominantly neutral trihexoses of glucose, galactose and N-acetyl-glucosamine/-galactosamine (~10.5%) and longer neutral oligosaccharides (1.5%; see Table S3A for a complete analysis of the oligosaccharide composition of this S-BMO preparation).

Dietary information for Malawian infants during the transition from exclusive breastfeeding to weaning was compiled from food diaries (see Experimental Procedures) and allowed us to develop a representative Malawian diet consisting of eight principal ingredients (M8). This diet does not satisfy the recommended daily nutritional needs of humans or mice (see Tables S3B,C for ingredients and measured nutritional composition).

Groups of co-housed 5-week-old male germ-free C57BL/6J mice (n=4-5/cage) were colonized with a single oral gavage of the defined 25-strain community. Mice were fed either M8 alone or M8 supplemented with S-BMO starting three days prior to colonization (see Figure 2C for experimental design). A separate group of mice received M8 supplemented with the structurally distinct, indigestible fructan, inulin. S-BMO and inulin were separately added to M8 at a concentration of 5.4% (g/g dry mouse diet), corresponding to HMO abundance in mature mother’s milk when expressed as a percentage of its dry components (Coppa et al, 2011). Importantly, the supplemented diets were isocaloric to M8, as judged by bomb calorimetry (153-155 kcal/100g, Table S3C).

We applied short read length shotgun sequencing of fecal DNA (COmmunity PROfiling by sequencing; COPRO-Seq), to determine colonization efficiency at strain-level resolution (McNulty et al, 2011). Colonization was operationally defined as achieving >0.1% mean relative abun-
dance in the fecal microbiota of at least one diet treatment group 44 days following gavage of the culture collection. The results revealed that 19 of the 25 strains (76%) colonized recipient gnotobiotic mice (Table S4A): those that failed to colonize included both strains of *Bifidobacterium longum* subsp. *infantis*, *Klebsiella varicola*, *Peptoniphilus harei* and one strain each of *Enterococcus faecalis* and *Olsenella uli*. The strains that successfully colonized gnotobiotic mice represent 10/11 (91%) of the bacterial genera and all four bacterial phyla that constitute the 25-member defined community. Thus, the great majority of the community’s overall phylogenetic diversity is represented in the gut microbiota of these animals. Colonization was highly reproducible among replicate animals within and across treatment groups (Figure S1A,B, Table S4A).

Body weight and composition (defined by quantitative magnetic resonance) was monitored over a 5-week period following gavage. Weight gain was significantly increased by S-BMO supplementation but not by supplementing with inulin (Figure 3A). Lean body mass gain was significantly increased by S-BMO (Figure 3B). No significant differences in fat mass gain were observed [1.72±0.33 g, 1.40±0.30 g, and 0.97±0.08 g for unsupplemented, inulin-supplemented and S-BMO-supplemented mice, respectively (mean±SEM); *p*>0.05, Student’s t-tests]. The significant effect of S-BMO on body weight and composition was replicated in two independent gnotobiotic mouse experiments.

The observed differences in weight and lean body mass gain were not attributable to differences in food consumption, which was not significantly different between treatment groups [2.17±0.44 g/mouse/day (mean±SD) vs 2.92±0.33 g/mouse/day for control and S-BMO-treated animals, respectively (consumption monitored daily for a 4-day period); *p*=0.08, Students t-test]. S-BMO-associated growth promotion was also microbiota-dependent, as it was not observed in germ-free animals (*n*=5 mice/treatment group; Figure 3C).

A previous comparison of germ-free and conventionally-raised mice indicated that the microbiota affects bone mass (Sjogren et al, 2012). Microcomputed tomography (µCT) of femurs harvested at the time of euthanasia revealed that S-BMO supplementation was associated with
significant increases in cortical thickness, cortical volumetric bone mineral density (vBMD) and cortical bone volume normalized to tissue volume (BV/TV; TV = cortical area plus medullary area) compared to controls consuming the M8 diet alone (Figure 3D, Table S5). Histological analysis of femurs also disclosed significantly increased BV/TV in the trabecular region in mice consuming the S-BMO-supplemented diet (Figure 3D,E). No significant difference was observed in osteoclast number normalized to bone surface area [2.24±0.24 (mean ± SD) cells/mm versus or 2.33±0.91 cells/mm, S-BMO versus control group; \( p = 0.86 \), Student’s t-test]. These data indicate that increased BV/TV and vBMD with S-BMO supplementation are not due to fewer osteoclasts resorbing bone. It is formally possible that the S-BMO associated bone phenotype is due to increased mechanical loading that accompanies increased body weight. Mechanical loading is associated with reduced endosteal bone resorption and a reduced medullary area in the cortical region of femur (Kodama et al, 1999). However, we observed no significant difference in cortical medullary area in the femurs of S-BMO-supplemented mice [1.05±0.01 mm\(^2\) (mean±SD)] compared to controls (1.07±0.11 mm\(^2\)) fed the unsupplemented M8 diet (\( p = 0.85 \), Student’s t-test), suggesting that increased loading is not responsible for increased BV/TV and vBMD associated with S-BMO treatment.

**Members of the distal gut microbiota respond transcriptionally to S-BMO in vivo**

16S rRNA analysis of fecal samples disclosed that S-BMO treatment did not produce a robust change in the relative abundance of any community member over the course of the experiment other than *E. coli* (Table S4C). The effect on *E. coli* was seen during the first 16 days following introduction of the culture collection, during which time this strain exhibited a significantly greater relative abundance in the fecal microbiota of SBMO-treated animals (Figure 4A).

We subsequently identified bacterial species that responded transcriptionally to S-BMO supplementation by sequencing RNA isolated from the cecal contents of treated and control animals and then mapping reads to the strains’ genomes (10.8±1.5 million reads/sample). Remarkably, S-BMO treatment did not produce statistically significant changes in gene expression, as
judged by DESeq (see Experimental Procedures), in any of the three *Bifidobacterium* strains (*B. catenulatum, B. breve* and *B. bifidum*) that established themselves in the gut microbiota of recipient gnotobiotic mice. In addition, no genes in the *S. enterica* strain were differentially expressed with S-BMO supplementation after correction for multiple hypotheses (α=0.1; data not shown). To further rule out the possibility that the growth effects observed in S-BMO-treated mice were attributable to this enteropathogen, groups of five-week old male germ-free animals (n=5 mice/cage) were colonized with the culture collection, with or without this strain; its removal had no significant effect on body weight or lean body mass gain (p>0.05, repeated measures two-way ANOVA; data not shown).

The two organisms that exhibited the greatest transcriptional response to S-BMO were *E. coli* and *B. fragilis* (Figure 4B). A total of 111 *E. coli* genes were differentially expressed (p<0.1, negative binomial test after Benjamini-Hochberg correction), with 110 exhibiting greater representation in the cecal meta-transcriptome of S-BMO-treated compared to untreated animals (Table S6). These transcriptional responses represented increased gene expression, rather than increased abundance of *E. coli*; there was no statistically significant difference in this strain’s relative abundance between the two groups of animals at the time their cecal contents were collected 44 days following gavage as determined by 16S rRNA sequencing (Figure 4A). The similar abundance of *E. coli* between treatment groups at this time point was independently confirmed by COPRO-Seq (1.65 ± 0.8% versus 1.19 ± 0.2% in M8 and M8 + S-BMO treated mice, respectively; p=0.54, Student’s t-test; Table S4A). Although this strain contains homologs of two key virulence factors associated with enteroaggregative *E. coli* (Table S2B), neither gene (*aggR* and *aap*) had detectable levels of expression in either diet context. KEGG annotation of the set of differentially expressed *E. coli* genes revealed several pathways involved in central energy metabolism that were upregulated by S-BMO, including the TCA cycle, plus pathways for glycolysis/gluconeogenesis, pyruvate metabolism, galactose metabolism, purine and pyrimidine metabolism and aminoacyl-tRNA biosynthesis (Figure 4C).
*E. coli* is not known to be a direct consumer of HMOs. Therefore, we hypothesized that its response to S-BMO reflects its membership in a bacterial food web, where primary consumers catabolize S-BMO and *E. coli* benefits secondarily. Members of the genus *Bacteroides* harbor arsenals of polysaccharide utilization loci (PULs), each of which specifies proteins that function to detect, import and catabolize dietary and host glycans (Comstock 2009, Martens et al, 2009). PULs are minimally defined by an adjacent pair of genes homologous to two *B. thetaiotaomicron* starch utilization system (Sus) genes: *susC* and *susD*. Two species of *Bacteroides*, *B. fragilis* and *B. ovatus*, were represented in the Malawian infant community that colonized recipient gnotobiotic mice. The former strain’s genome does not encode *B. fragilis* enterotoxin, a zinc-dependent metalloprotease associated with diarrheal disease in humans (Sears, 2001; 2009). While no differences in PUL expression were documented in *B. ovatus*, several *B. fragilis* PULs exhibited differential expression with S-BMO supplementation. One upregulated PUL contains *susC* (*BACFRAMC1_2659*) and *susD* (*BACFRAMC1_2660*) homologs adjacent to genes encoding an *N*-acetylneuraminic acid lyase (EC4.1.3.3) and an *N*-acetylglucosamine epimerase (EC5.1.3.8) (*Figure 4D*), consistent with this PUL being involved in S-BMO metabolism (*N*-acetylneuraminic and *N*-acetylglucosamine are major components of S-BMO; *Table S3A*).

**Cross-feeding between the *B. fragilis* and *E. coli* strains**

These findings suggested that *B. fragilis* is a primary consumer of S-BMO and that it could be involved in a food web where *E. coli* acts as a secondary consumer. To test this notion, we incubated *B. fragilis* or *E. coli* in PBS containing 5% S-BMO and assayed the supernatants for sialyllactose and sialic acid using ultra high performance liquid chromatography-mass spectrometry (UPLC-MS; see *Experimental Procedures*). This analysis confirmed that *B. fragilis* degraded sialyllactose, producing a concomitant increase in free sialic acid. In contrast, *E. coli* did not significantly alter the concentration of either compound under these conditions (*Figure 4E*).

To determine whether the products of S-BMO degradation by *B. fragilis* could support the growth of *E. coli*, S-BMO was incubated with *B. fragilis* and then used as the sole carbon source
for *E. coli* in minimal medium. Growth was robust in medium containing *B. fragilis*-conditioned S-BMO but minimal with unconditioned S-BMO (Figure 4F). Growth in medium containing conditioned S-BMO was also significantly increased compared to controls with sialic acid added as the sole carbon source (Figure 4F), even though *E. coli* degraded sialic acid [28.0±4.2% decrease (mean±SD) in sialic acid levels in culture supernatants compared to uninoculated controls, *p*<0.01, Student’s t-test; data not shown]. These results suggested that lactose, or its constituent monomers, glucose and galactose, are likely the primary substrates used by *E. coli* under these conditions.

These observations raised the possibility that the microbiota-dependent growth promotion observed in vivo with S-BMO supplementation could be attributed either directly to primary consumers of S-BMO (e.g., *B. fragilis*) or indirectly to secondary consumers (e.g., *E. coli*). Therefore, we assessed the ability of S-BMO to promote growth of 5 week-old male gnotobiotic mice colonized with just the *B. fragilis* and *E. coli* strains. S-BMO did not produce a significant difference in growth of co-colonized mice over a 5-week period compared to M8 controls (Figure 4G). We concluded that (i) colonization of the gut per se is not sufficient to produce the S-BMO-enhancement of growth, and (ii) other members of the community and/or higher-order interactions between *B. fragilis*, *E. coli* and these other members are required to mediate this growth promotion.

**S-BMO supplementation modulates metabolism in host tissues**

To examine host metabolic changes that accompany S-BMO-sponsored growth promotion, we used targeted MS/MS and GC/MS to measure a panel of 176 metabolites (Ferrara et al, 2008; Newgard et al, 2009), including amino acids, acylcarnitines, organic acids, fatty acyl CoAs and ceramides in liver, serum and brain obtained from S-BMO-supplemented versus M8 control mice at the time of euthanasia 5 weeks after initial gavage of the 25-member bacterial culture collection (*n*=4-5 mice/treatment group). The major impact of S-BMO supplementation was greatly enhanced changes in fatty acid-derived metabolite concentrations in the fasted compared to non-fasted (fed) states. As such, we observed a significantly lower level of medium- and long-chain acylcarnitine species in the sera of non-fasted, S-BMO-treated versus control animals (Figure
this observation was mirrored by decreases in 11 medium- and long-chain acylcarnitines and five fatty acyl CoAs in their livers (Figure S2A, Table S7A). The decreased levels of fatty acid metabolites in the liver and serum of these non-fasted mice could reflect decreased oxidation of fat and enhanced lipid storage, consistent with a more normalized anabolic state. Accordingly, we documented that these animals had significantly elevated concentrations of serum triglycerides, insulin and leptin (Figure S2B-D). We also noted a trend of higher serum non-esterified fatty acids (NEFA) [0.75±0.14 mM (mean±SD) versus 0.59±0.06 mM for S-BMO-supplemented and control animals, respectively (p=0.1, Student’s t-test)]. Low serum leptin levels were reported to be a major predictor of childhood mortality in a cohort of undernourished Ugandan children (Bartz et al, 2014). Moreover, leptin levels correlate with bone mineral density in women (Thomas et al, 2001).

These observations suggest that S-BMO-supplemented mice can more effectively utilize dietary components for anabolism. To test the corollary that S-BMO-treated animals are better equipped to mobilize nutrients during periods of fasting, we measured the same 176 metabolites in the livers and sera of mice that had been fed M8 with or without S-BMO (or inulin) and euthanized following an 8-hour fast. The concentrations of medium- and long-chain acylcarnitines (notably C18, C18:1 and C18:2) were elevated in sera of fasted, S-BMO-treated animals compared to M8 controls (Figure 5A, Table S7B). In addition, concentrations of 15 acylcarnitine and 29 long-chain fatty acyl CoA species were significantly higher in the livers of S-BMO-treated animals (Figure 5B, Table S7B). Importantly, these observations were largely specific to S-BMO: similar changes were not seen in mice consuming the isocaloric inulin-supplemented M8 diet. In aggregate, our findings demonstrate a greater increment of lipid-derived metabolite concentrations in S-BMO-treated mice between the fed and fasted states than in mice fed an unsupplemented diet. This metabolite pattern is consistent with more efficient switching from anabolic storage of fat in the fed state to its oxidation in the fasted state, which can also be described as metabolic flexibility (Muoio 2014).

We used targeted MS/MS to measure metabolites in the cecal contents of fasted mice and observed no statistically significant differences in the concentrations of organic acids or amino
acids (including branched-chain amino acids) between S-BMO-treated and control animals (Table S7B). We also broadened our search for metabolic biomarkers associated with S-BMO-sponsored growth by performing non-targeted gas chromatography-mass spectrometry (GC/MS; Scholtens et al, 2014) on serum samples obtained from fasted S-BMO-supplemented mice at time of euthanasia. Orthogonal projection to latent structures discriminatory analysis (O-PLS-DA) revealed that the serum metabolic profile of S-BMO-treated mice clustered distinctly from that of controls fed the M8 diet (Figure 5C inset). We used this method to identify particular serum metabolites that discriminate S-BMO-supplemented animals from controls and found that a branched-chain amino acid (BCAA) metabolite, 2-ketovaline, and a product of fatty acid oxidation, β-hydroxybutyrate, were associated with and increased by S-BMO in fasted mice (Figure 5C, Table S7D). The increase in β-hydroxybutyrate in fasted S-BMO-supplemented mice supports our interpretation that increased fatty acid oxidation underlies the increased levels of fatty acid-derived acyl CoA and acylcarnitine species in liver and serum of these animals. The increase in a BCAA-derived metabolite also suggests that amino acids are being utilized as energy substrates in the fasted state, an expected and normal physiologic response. BCAA are also known to promote muscle protein synthesis by activating mammalian target of rapamycin (Norton and Layman, 2006). This is notable given the significant augmentation in gain of lean body mass observed with S-BMO treatment (Figure 3B).

In addition to stunted growth, childhood undernutrition results in persistent deficits in cognitive development (Victora et al, 2008). Sialic acids, a group of compounds derived from neuraminic acid, including the nine-carbon monosaccharide N-acetylneuraminic acid (NeuAc), are integral to postnatal brain development; NeuAc is a component of gangliosides and is covalently linked to neural cell adhesion molecules (NCAMs) that mediate cell-cell interactions involved in synaptogenesis and memory (Wang, 2009; Wang and Brand-Miller, 2003). Dietary supplementation with sialylated glycoproteins has been shown to increase polysialylation of NCAM and improve memory (Wang et al, 2007a). Since NeuAc is a core component of S-BMO (Table S3A), we used nontargeted UPLC-MS to characterize the representation of metabolites in the cerebrums
of gnotobiotic mice. O-PLS-DA analysis revealed that the cerebral metabolic profiles of S-BMO-treated mice were distinct from control animals fed the M8 diet (Figure S3A inset, Table S7E). Interestingly, we compared the m/z feature that was most discriminatory for the brain of S-BMO-treated animals (mass 346.0546 Da; Figure S3A) to the human metabolome database (Wishart et al, 2013) and found that the compound closest in mass was an adduct of NeuAc (M+K-2H; 346.054587 Da).

Applying targeted GC-MS to the same cerebral sample extracts also revealed distinct metabolic profiles. Free NeuAc, inosine and adenosine were among eight metabolites identified as significantly elevated in the brains of S-BMO-treated mice (Figure S3B,C and Table S7F). Inosine treatment promotes axonal rewiring and improves behavior in animal models of traumatic brain injury (Chen et al, 2002; Dachir et al, 2014; Smith et al, 2007), while adenosine has both neuro-modulatory and neuroprotective functions in the brain (Fredholm et al, 2005).

**S-BMO supplementation promotes growth in gnotobiotic piglets**

We next examined whether the observed effects of S-BMO supplementation on growth and metabolism in mice also occurred in a second mammalian species whose physiology was closer to humans. We selected gnotobiotic piglets because (i) the digestive physiology of swine is similar to that of humans (Kararli, 1995; Miller and Ullrey, 1987), (ii) following birth, piglets grow very rapidly (Miller and Ullrey, 1987), providing a convenient window to test the effects of interactions between a stunted human donor microbiota and diet and (iii) procedures have been described for re-deriving piglets as germ-free (Willing and Van Kessel, 2007, 2009).

Since newborn piglets exhibit impaired gut barrier function (Pluske et al, 1997; Wijtten et al, 2011), we excluded the eight members of the defined community generated from the stunted Malawian donor that possessed more than three known virulence factors in their genomes, including *E. coli* (Table S2). A group of six germ-free domestic piglets was colonized at 3 days of age with the remaining 17-strain consortium (see Figure 6A and Table S2A for a list) and weaned from a sterile sow’s milk replacement formula to the M8 diet over the course of the next seven
days. At day 13 of postnatal life, piglets were split into control and experimental groups (n=3 piglets/group). For a period of six days, piglets in the experimental group were exclusively fed M8 supplemented with 5.4% S-BMO (wt/wt), while control animals were fed the M8 diet alone (see Figure 6A for a schematic representation of the experimental design).

COPRO-Seq analysis of fecal samples collected at time of euthanasia revealed that colonization was efficient, with 14/17 (82%) of input strains detected in the piglets’ fecal microbiota at >0.1% relative abundance (Table S4B). The failure of three human-derived strains (Peptoniphilus harei and both strains of Bifidobacterium longum subsp. infantis) to colonize germ-free piglets in addition to germ-free mice could reflect a number of factors including their inability to thrive in the M8 diet context, the absence of syntropic partners in the culture collection used for colonization or restricted host specificity. As observed in mice, microbiota structure was similar between S-BMO-supplemented and control animals (Figure S1C,D, Table S4D). Despite the short period of treatment, piglets receiving the S-BMO-supplemented M8 diet displayed significantly greater body weight gain compared to controls (Figure 6B).

The processing of S-BMO within the gut was characterized by triple quadrupole mass spectrometry of mucosal scrapings obtained at time of euthanasia from the proximal and distal segments of the small intestine and spiral colon (corresponds to the ascending colon in humans), plus cecal contents and feces. As expected, the concentrations of NeuAc were significantly higher in both the cecal contents and feces of S-BMO-supplemented animals (Figure 6C). Importantly, we documented a significant increase in the concentration of mucosal NeuAc in the proximal small intestine (Figure 6D), suggesting that dietary NeuAc (a primary component of S-BMO) is incorporated in host glycans. This observation was site-specific, as no significant differences in mucosal NeuAc concentrations were observed in the distal small intestine or spiral colon.

To investigate the effects of S-BMO on the community meta-transcriptome, we performed microbial RNA-Seq on cecal samples harvested at the time of euthanasia [n=42.5±15.2 million reads per sample (mean±SD)]. We were unable to detect differences in gene expression for any
community members using $p<0.1$ as a threshold for statistical significance (negative binomial test corrected using the Benjamini-Hochberg procedure). Relaxing this cutoff to $p<0.2$ revealed 1,482 bacterial genes that exhibited differential expression with 1,481 (99.9%) showing significantly higher expression with S-BMO treatment (Table S8). These differences are not attributable to bacterial abundance; COPRO-Seq analysis established that S-BMO did not significantly impact the relative abundance of any bacterial strain colonizing piglets ($p>0.05$, Student’s t-test; Table S4B).

Expression levels of 587 $B.\ fragilis$ genes were significantly elevated with S-BMO-supplementation, more than any other community member. Among these 587 genes, six encode enzymes that catalyze several key steps in the KEGG pathway for biosynthesis of three BCAA (valine, leucine and isoleucine; Figure S4). Differentially expressed genes were not evenly distributed across the bacterial community; five of the community’s 14 strains ($Bacteroides\ fragilis\ MC1$, $Bacteroides\ ovatus\ MC1$, $Collinsella\ aerofaciens\ MC1$, $Collinsella\ aerofaciens\ MC2$ and $Bifidobacterium\ ca\ tenulatum\ MC1$) together accounted for 1,300/1,482 (88%) of the differentially expressed genes (Table S8). One interpretation of these data is that, in the absence of $E.\ coli$, primary consumption of S-BMO (by $B.\ fragilis$, for example) opens a niche for other strains to fill.

Consistent with our observations in non-fasted gnotobiotic mice, the liver metabolic profiles of non-fasted, S-BMO-treated piglets exhibited significantly reduced levels of five acylcarnitine and nine fatty acyl CoA metabolites compared to controls (Figure 6E, Table S7C). In addition, we documented a striking pattern of increased amino acid concentrations, including BCAAs, in serum as well as skeletal muscle (biceps femoris) of S-BMO-supplemented piglets (Figure 6F,G, Table S7C). Non-fasted, S-BMO-treated piglets also displayed trends for decreased medium- and long-chain serum acylcarnitines [0.14±0.02 μM (mean±SD) versus 0.17±0.02 μM in controls; $p=0.12$, Student’s t-test], increased serum triglycerides [30.33±13.05 mg/dL (mean±SD) versus 10.33±8.08 mg/dL; $p=0.08$, Student’s t-test] and increased serum NEFA [0.07±0.03 mM (mean±SD) versus 0.04±0.01 mM; $p=0.2$, Student’s t-test]. Despite the increased expression of BCAA biosynthesis genes by $B.\ fragilis$, S-BMO administration produced no significant differences in the concentration of BCAA in cecal contents compared to controls ($p>0.05$, Student’s
t-test; Table S7C). The mechanism underlying the strong increment in amino acids documented in S-BMO-treated piglets and its relationship to their augmented growth rate remains to be defined.

Together, these results led us to conclude that the effects of S-BMO supplementation on weight gain and metabolic phenotypes are a shared feature of gnotobiotic piglets and mice and that these changes reflect, at least in part, improved utilization of dietary components. Furthermore, the fact that S-BMO promoted growth in the absence of E. coli provided additional evidence that other members of the microbiota and/or their higher order interactions are necessary to mediate these effects.

**Discussion**

We describe two gnotobiotic animal models, each colonized with a collection of bacterial strains cultured from a single stunted Malawian infant and fed a prototypic Malawian diet, to show that a purified preparation of sialylated bovine milk oligosaccharides, resembling sialylated HMOs depleted in the breast milk of mothers with stunted infants, produces beneficial microbiota-dependent effects on growth and metabolism.

The data presented here from two separate studies of undernutrition in Malawi indicate that breast milk of mothers whose children exhibit growth faltering contains lower levels of HMOs, including sialylated glycans, compared to milk from mothers of healthy infants. Sialic acid-containing oligosaccharides are present at up to 20-fold higher levels in mature human milk compared to mature bovine milk (Bruggencate et al, 2014; Tao et al, 2009; Wang et al, 2001). As a result, existing bovine milk-based infant formulas and complementary/therapeutic foods used to treat undernutrition are deficient in these important human milk components.

The brains of breastfed infants possess higher levels of ganglioside- and glycoprotein-bound sialic acid than the brains of formula-fed infants (Wang and Brand-Miller, 2003), consistent with the higher sialic acid content of breast milk compared to bovine milk-based infant formulas (Wang et al, 2001). Although the liver can synthesize sialic acid *de novo* from glucose, the activity
of the limiting enzyme, UDP-N-acetylglucosamine-2-epimerase, is low during the neonatal period (Gal et al, 1997), suggesting that in the absence of human milk, sialic acid may be a limiting resource at the time of rapid postnatal brain development. Our finding that dietary S-BMO supplementation of gnotobiotic mice harboring microbiota from an undernourished Malawian infant leads to elevated levels of NeuAc in the brain suggests that S-BMO can function as a supplementary source of bioavailable sialic acid and may represent a novel treatment for children at risk for stunting and associated neurodevelopmental deficits whose dietary sialic acid intake is inadequate.

In addition to neural development, sialic acid may play a fundamental role in other facets of healthy growth, including those involving the musculoskeletal system. Studies of a mouse model of myopathy in which GNE, which encodes a bifunctional rate-limiting enzyme in sialic acid biosynthesis, is mutated, showed that addition of 6′-sialyllactose to animals’ drinking water led to increased muscle mass and improved contractility (Yonekawa et al, 2014). BCAA promote muscle protein synthesis by activating mTOR, and an increase in a key BCAA-derived metabolite also suggests restoration of normal amino acid metabolism in fasted S-BMO-supplemented mice. S-BMO also increased bone cortical thickness and trabecular bone volume in young gnotobiotic mice. The underlying mechanisms need to be investigated, but this observation has a clinical correlate: studies in Uganda reported reductions in fibular bone width in 3.5- to 7.5-year old children that exhibited low body weight during their first 2 years of postnatal life (McFie and Welbourn, 1962).

Our data also indicate that orally administered S-BMO restores normal substrate utilization patterns, most notably: (i) a strong activation of fatty acid oxidation in fasting, as reflected by increased levels of acylcarnitines, acyl CoAs and the fatty acid oxidation end product β-hydroxybutyrate, and (ii) a pattern consistent with enhanced anabolic metabolism of lipids in the fed state, including increases in triglycerides and insulin, and decreased levels of fatty acid oxidation intermediates.
Our findings demonstrate that specific bacterial members of the gut microbiota obtained from an undernourished Malawian infant are able to metabolize S-BMO-derived sialyllactose to its constituent monosaccharides, which are then available to the host and other members of the microbiota, both mutualistic and pathogenic. The latter includes pathogenic *E. coli* that are unable to harvest sialic acids themselves from host glycans or dietary sources (Almagro-Moreno and Boyd, 2009). Therefore, it is important to consider that the presence of primary consumers of sialylated oligosaccharides may create opportunities for enteropathogens to achieve a fitness benefit that could have deleterious effects on the host (Ng et al, 2013). These findings encourage mindfulness with respect to balancing potential benefits and risks when designing gut microbiota-directed clinical trials involving milk oligosaccharides.

In conclusion, our analysis of breast milk specimens from clinical trials of undernutrition in Malawi has disclosed an association between HMO abundance and composition and infant growth. To test for a causal relationship, we have used preclinical gnotobiotic models to uncover microbiota-dependent growth promoting activities of purified S-BMO in young gnotobiotic mice and piglets. Preclinical models of the type we describe can play critical roles in characterizing BMO species that are structurally similar to human HMOs and that can be purified at a scale compatible with clinical testing and potentially into a commercially viable product. These preclinical models provide a means to (i) elucidate mechanism of action, including underlying effects on host metabolism that would otherwise be difficult or impossible to glean from non-invasive clinical studies, especially those involving infants, (ii) define the generalizability of the effects of S-BMO in the context of microbiota from different donors representing different geographic locations and dietary practices, (iii) identify clinically translatable microbial and host biomarkers related to mechanism and safety and (iv) guide patient stratification and inform the design of proof of concept clinical studies that incorporate appropriate de-risking strategies and careful surveillance to establish the efficacy and safety of approaches such as S-BMO supplementation to the diets of undernourished children.
Experimental Procedures

Human studies

Human studies were conducted with the approval of institutional review boards from the University of Malawi, Pirkanmaa Hospital (Finland), and Washington University School of Medicine in St. Louis. The mothers and infants included in this study were enrolled in two randomized, controlled, single-blind, parallel group clinical trials, LCNI-5 and iLiNS-DYAD-M [ClinicalTrials.gov identifiers: NCT00524446 and NCT01239693]. See Extended Experimental Procedures for additional information.

Generating a clonally arrayed culture collection from a 6-month-old Malawian infant

The 25-member clonally arrayed culture collection was produced from a frozen fecal sample collected from a severely stunted 6-month-old Malawian infant [patient ID h264A in Yatsunenko et al. (2012)] using methods described in Goodman et al. (2011). For detailed information relating to bacterial isolation, plus genome sequencing, assembly and annotation see Extended Experimental Procedures.

Purification and compositional analysis of S-BMO

S-BMO was purified from commercially available whey permeate using a series of ultrafiltration and chromatographic steps described in Extended Experimental Procedures.

Design and preparation of the Malawi 8 (M8) diet

This diet was formulated based on a dietary survey of the complementary feeding practices of 43 nine-month-old Malawian infants and children enrolled in the iLiNS-DOSE clinical study (#NCT00945698) that took place in the Mangochi district of Malawi. See Extended Experimental Procedures for details.
**Studies involving gnotobiotic mice**

All gnotobiotic mouse experiments were performed using protocols approved by the Washington University Animal Studies Committee. Male germ-free C57Bl/6J mice were maintained in sterile, flexible plastic gnotobiotic isolators. Mice received a single oral gavage of the complete 25-member culture collection or subsets of the collection 3 days after initiation of one of the M8-based diets. For additional information, see *Extended Experimental Procedures*.

**16S rRNA gene sequencing of fecal microbiota**

Multiplex sequencing of barcoded bacterial 16S rRNA amplicons generated from fecal microbiota samples was performed using an Illumina MiSeq instrument. 97% ID OTUs were clustered using UCLUST (Edgar, 2010). Taxonomic assignments were made with the Ribosomal Database Project (RDP) version 2.4 classifier (Wang et al, 2007b). See *Extended Experimental Procedures* for detailed information about sample preparation, PCR conditions, multiplex sequencing and analysis of the resulting datasets.

**Microbial RNA-Seq**

Multiplex microbial cDNA sequencing was performed using Illumina Hi-Seq2000 and MiSeq instruments. Sequencing reads were split according to 8-bp barcodes used to pool samples. Reads were mapped to genes in a custom database of 25 draft genomes included in the defined Malawian infant bacterial community. Read counts were normalized and analyzed with the R statistical package, DESeq (Anders and Huber, 2010). See *Extended Experimental Procedures* for additional information.

**Mass spectroscopy of metabolites contained in biospecimens obtained from gnotobiotic animals**

Tissue samples were collected from gnotobiotic animals at time of euthanasia, frozen immediately in liquid nitrogen and stored at -80°C. In preparation for GC/MS and MS/MS analyses, samples
were homogenized in 50% aqueous acetonitrile containing 0.3% formic acid (50 mg wet weight tissue/mL solution). See Extended Experimental Procedures for details about analytic methods.

Experiments involving gnotobiotic piglets

All experiments involving pigs were performed under the close supervision of a veterinarian using protocols approved by the Washington University Animal Studies Committee. Pregnant domestic sows (mixture of Landrace and Yorkshire genetic backgrounds, mated to Duroc breed domestic boars) were obtained from Oak Hill Genetics (Ewing, IL). For detailed procedures relating to the derivation of germ-free piglets and experimental design, see Extended Experimental Procedures.

Supplemental Information

Supplemental information includes Extended Experimental Procedures, four figures, and eight tables.

Author Contributions

M.R.C., D.A.M. and J.I.G. designed the experiments; D.O., M.R.C. and M.T. generated gnotobiotic piglets; M.R.C. and J.G. produced and characterized the clonally arrayed bacterial culture collection; L.V.B. and K.G.D. designed the M8 diet; C.W. and D.B. purified and characterized the S-BMO preparation; K.M., C.M., Y.F., J.J., K.G.D. and P.A. designed and oversaw the clinical studies, sample collection and sample processing; S.T. and J.C.C.D. analyzed HMO content of breast milk samples; M.R.C. and D.O. performed gnotobiotic mouse and piglet experiments; M.R.C. performed in vitro experiments; J.C., J.R.B., M.J.M. and O.I. performed mass spectrometry-based metabolomics analyses of gut microbiota, tissue and serum samples obtained from gnotobiotic mice and pigs; M.R.C. generated and analyzed 16S rRNA sequence data; M.R.C. generated the COPRO-Seq and microbial RNA-Seq datasets; M.R.C., M.J.B., C.B.N., C.L., D.A.M. and J.I.G. analyzed the data; M.R.C., M.J.B. and J.I.G wrote the paper.
Acknowledgements

We are grateful to Sabrina Wagoner, Maria Karlsson, Justin Serugo, Siddarth Venkatesh, Jeanette Gehrig and Josh Sommovilla for assistance with gnotobiotic mouse and piglet husbandry, Jeremiah Faith for assistance generating the Malawian infant bacterial culture collection, plus Marty Meier, Su Deng, Jessica Hoisington-Lopez, Paras Vora and Daniel Leib for superb technical assistance. J.I.G. is a co-founder of Matatu, Inc., a company characterizing the role of diet-by-microbiota interactions in animal health. D.A.M., D.B. and C.B. are co-founders of Evolve Biosystems, Inc., a company focused on diet-based manipulation of the gut microbiota.

This work was supported in part by grants from the Bill and Melinda Gates Foundation and the National Institutes of Health (DK30292), with additional funding from the Office of Health, Infectious Diseases and Nutrition, Bureau for Global Health, USAID under terms of Cooperative Agreement No. AID-OAA-A-12-00005, through the Food and Nutrition Technical Assistance III Project (FANTA) managed by FHI 360. Data management and statistical analysis for the iILNS-DYAD-M clinical study was also funded through additional grant support from the Academy of Finland (grant 252075) and the Medical Research Fund of Tampere University Hospital (grant 9M004). Micro-CT analysis was performed in the Washington University Musculoskeletal Research Center, which is supported by NIH P30 AR057235.

Accession Numbers

Bacterial V4 16S rRNA datasets, whole genome shotgun sequencing datasets from cultured bacterial strains, COPRO-Seq datasets and microbial RNA-Seq datasets have been deposited in the European Nucleotide Archive (ENA).
References


DOI: 10.1111/mcn.12068.

by the human gut microbiota: the Bacteroidetes Sus-like paradigm. J Biol Chem J 284,
24673-24677.

Martorell, R., and Zongrone, A. (2012). Intergenerational influences on child growth and

McFie, J., and Welbourn, H. F. (1962). Effect of malnutrition in infancy on the development of

McNulty, N., Yatsunenko, T., Hsiao, A., Faith, J., Muegge, B., Goodman, A., Henrissat, B., Oozeer,
milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. Sci Transl
Med 3, 106ra106.

382.


metabolic signature that differentiates obese and lean humans and contributes to insulin


Figure Legends

Figure 1: HMOs are more abundant in breast milk of Malawian mothers with healthy infants. (A,B) Abundance of total, fucosylated and sialylated HMOs present in breast milk of Malawian mothers collected from the LCNI-5 cohort (n=88 mothers, panel A) and iLiNS-DYAD-M cohort (n=215 mothers, panel B) 6 months postpartum, binned by anthropometry of their infants (LCNI-5: healthy, HAZ>0; stunted, HAZ<-3; iLiNS-DYAD-M: healthy, HAZ>0; stunted: HAZ<-2). HMO abundance values correspond to spectral abundance as determined by LC-TOF MS, normalized to the mean abundance of samples assigned to the healthy bin for each respective comparison. *p<0.05, **p<0.01, ***p<0.001 (two-tailed, Welch’s t-test). See also Table S1.

Figure 2: Cultured bacterial strain collection generated from the fecal microbiota of a 6-month-old stunted Malawian infant. (A) Taxonomic representation of 97% ID OTUs in the intact, uncultured infant fecal sample from which the culture collection was generated. (B) Comparison of culture collection strains and representative type strains. Black dots indicate percent sequence identity between full-length 16S rRNA gene sequences from each strain and its most similar reference type strain. Bars indicate percent nucleotide sequence similarity between the strain’s de novo assembled genome and its most similar type strain’s genome. Strains are clustered by Euclidean distance between full-length 16S rRNA gene sequences. (C) Experimental design. Mice were colonized at 5 weeks of age. See also Table S2.

Figure 3: S-BMO promotes growth in gnotobiotic mice harboring the Malawian infant culture collection and fed a prototypic Malawian diet. (A) Weight gain over time normalized to body weight at the time of colonization. (B) Lean body mass gain, as measured by quantitative magnetic resonance (qMR), 44 days after colonization of animals fed either the M8 diet alone or M8 diet supplemented with inulin or S-BMO. (C) Weight gain over time of germ-free mice fed M8 or M8 supplemented with S-BMO for 4 weeks (normalized to weight at 5 weeks of age). (D) Cortical and trabecular bone volume/tissue volume (BV/TV; TV = cortical area + medullary area) and
representative histomorphometry images of femurs from mice colonized with the culture collection and fed either M8 or M8 supplemented with S-BMO. Bone phenotypes were characterized 37 days post colonization. Black boxes highlight the trabecular region of interest, and arrowheads point to examples of trabeculae. For (A), ***p<0.001, two-way, repeated measures ANOVA. For (B) and (D), *p<0.05, **p<0.01 two-tailed, unpaired Student’s t-test. All comparisons are made to M8 controls. See also Figure S1 and Table S5.

Figure 4: Members of the gut microbiota respond transcriptionally to S-BMO in vivo and degrade S-BMO in vitro. (A) Relative abundance of *E. coli* (mean±SEM) over time in feces of gnotobiotic mice colonized with the stunted Malawian infant’s culture collection and fed M8 or M8 supplemented with S-BMO. *p<0.05 two-tailed, unpaired Student’s t-test. (B) Volcano plot of bacterial gene expression in the cecal microbiota of gnotobiotic mice. Expression is plotted on the x-axis as the log$_2$ fold difference between S-BMO-supplemented mice and unsupplemented mice. Significantly up- or downregulated genes are labeled in black or by bacterial species of origin. -Log$_{10}$(*p*-values) are plotted on the y-axis (negative binomial test, corrected using the Benjamini-Hochberg procedure). (C) Heatmap of select differentially regulated *E. coli* genes, grouped by KEGG functional pathway. Red asterisks denote that differences in expression are statistically significant after Benjamini-Hochberg correction for multiple hypotheses (α=0.1). See Table S6 for a full list of differentially expressed *E. coli* genes. (D) Effects of S-BMO on expression of genes in a *B. fragilis* PUL. Expression is plotted on the y-axis as the log$_2$ fold difference between S-BMO-supplemented and control mice. *p<0.05, **p<0.01 (two-tailed, unpaired Student’s t-test). (E) Abundance (mean±SEM) of sialyllactose and sialic acid in *E. coli* or *B. fragilis* monoculture supernatants after a 24-hour incubation with 5% S-BMO (wt/vol). Control incubations contained uninoculated PBS buffer. ***p<0.001 two-tailed, unpaired Student’s t-test. (F) OD$_{600}$ (mean±SEM) of *E. coli* grown in minimal medium containing various single carbon sources. ‘Conditioned S-BMO’ refers to filter-sterilized supernatant harvested from a 24 hour monoculture of *B. fragilis* with 5% S-BMO (wt/vol). (G) Weights (mean±SEM; normalized to weight at time of colonization) of gnotobiotic mice colonized at five-weeks-of-age with the entire 25-member Malawian
infant culture collection or with only *B. fragilis* and *E. coli* (Bf + Ec). Mice were fed the M8 diet with or without S-BMO (*n*=5 mice/group).

**Figure 5: S-BMO supplementation alters levels of serum and liver metabolites in gnotobiotic mice.** (A) Sum of the concentrations of medium- and long-chain serum acylcarnitines (chain length ≥10) in fasted or non-fasted gnotobiotic mice fed M8 or M8 supplemented with S-BMO or inulin. *p*<0.05, Student’s t-test. (B) Heatmap of liver metabolites whose concentrations are affected by S-BMO-treatment in mice fasted 8 hours prior to euthanasia. Rows represent replicate gnotobiotic mice, grouped by dietary treatment. Columns represent individual metabolites identified by targeted MS/MS. *p*<0.05 (two-tailed, unpaired Student’s t-test). All comparisons were made to controls. Red asterisks represent significant differences between the two groups of animals after Benjamini-Hochberg correction for multiple hypotheses (α=0.1). See Table S7A,B for concentrations of all measured metabolites. (C) O-PLS-DA score plot (inset) and O-PLS-DA S-plot of fasted serum metabolites in gnotobiotic mice fed M8 with or without S-BMO. Metabolites highlighted in purple represent branched-chain amino acid metabolites while those highlighted in green are ketone body metabolites. See also Figure S2 and Figure S3.

**Figure 6: S-BMO modulates growth and metabolism in gnotobiotic piglets.** (A) Experimental design. Full-term piglets were delivered into gnotobiotic isolators immediately after caesarian section and colonized with a 17-member consortium of the stunted Malawian infant’s culture collection at postnatal day 3. Piglets were weaned from a sterile sow’s milk replacement formula onto the M8 diet by day 11 and then split into two groups. For 6 days, starting at postnatal day 13, the treatment group received M8 supplemented with S-BMO while controls were fed the unsupplemented, isocaloric M8 diet. (B) Weight gain normalized to body weight at postnatal day 13 (mean±SEM). **p*<0.01, two-way, repeated measures ANOVA. (C,D) *N*-acetylneuraminic acid (NeuAc) concentrations in cecal contents and feces (panel C), as well as in proximal and distal small intestine (SI) and spiral colon mucosa (panel D), harvested from gnotobiotic piglets fed M8 ± S-BMO. All samples were obtained at the time of euthanasia. *p*<0.05, **p*<0.01 two-tailed, unpaired Student’s t-test. (E) Heatmap displaying acylcarnitine and fatty acyl CoA concentrations in the livers
of non-fasted gnotobiotic piglets, measured at the time of euthanasia on postnatal day 19. (F,G)
Amino acid concentrations in (F) serum and (G) skeletal muscle of non-fasted gnotobiotic piglets,
measured at the time of euthanasia. In panels E,F, columns represent individual metabolites, and
rows represent replicate gnotobiotic piglets, grouped by treatment. *p<0.05 (two-tailed, unpaired
Student’s t-test). All comparisons made to controls. Red asterisks denote statistical significance
after Benjamini-Hochberg correction for multiple hypotheses (α=0.1). See also Figure S1, Figure
S6, Table S7C and Table S8.
Figures

Figure 1.
Figure 2.
Figure 3.

A: Graph showing % initial weight over Days post colonization for M8, M8 + Inulin, and M8 + S-BMO conditions.

B: Bar graph showing lean mass gain [g] over Days on diet for M8, M8 + Inulin, and M8 + S-BMO conditions.

C: Graph showing % initial weight over Days on diet for M8 + S-BMO (germ-free) and M8 (germ-free) conditions.

D: Bar graph comparing BV/TV for M8 + S-BMO and M8 conditions, with separate bars for Cortical and Trabecular regions.

E: Images showing histological sections for M8 and M8 + S-BMO conditions, with inset images highlighting specific areas of interest.
Figure 4.

A. E. coli relative abundance [%] over days post colonization.

B. Expression of selected genes in different conditions.

C. KEGG pathway analysis showing enriched pathways.

D. B. Fragilis gene expression and annotation.

E. Spectral abundance of sialylactose and sialic acid.

F. E. coli OD600 over time with different conditions.

G. % initial weight over days post colonization.
Figure 5.
Figure 6.
Extended Experimental Procedures

Human studies

Human studies were conducted with approval of institutional review boards from the University of Malawi, Pirkanmaa Hospital (Finland) and Washington University School of Medicine in St. Louis. The mothers and infants included in this analysis were enrolled in two randomized, controlled, single-blind, parallel group clinical trials, LCNI-5 and iLiNS-DYAD-M [ClinicalTrials.gov identifiers: NCT00524446 and NCT01239693] of micronutrient fortified lipid-based nutrient supplements conducted in southern Malawi (Ashorn et al, 2015; Mangani et al, 2013). Anthropometric measurements of 6 month-old infants were performed in triplicate to determine height-for-age Z-scores (HAZ).

De-identified breast milk samples were collected from women enrolled in the two trials when their infants were six months old. The entire contents of one breast were manually expressed into a sterile container approximately 2 h after the last reported feeding on that breast. Aliquots were stored at -40°C to -80°C in Malawi and later shipped on dry ice to Washington University in St. Louis, where the samples were maintained at −80 °C until being shipped to UC Davis for HMO analysis. A fresh morning fecal sample was collected from the infants of these mothers on the day of their six-month clinic visit. The sample was divided into aliquots in a laboratory and stored at −20°C on the same day. Within three days, samples were transferred to -40°C or -80°C freezers in Malawi prior to shipment to Washington University.

Mass spectrometry of breast milk samples

Mass spectrometry analysis of breast milk HMO content was performed on all available samples from mothers of children enrolled in the LCNI-5 and iLiNS-DYAD-M study cohorts that met the anthropometric criteria stated in the main text (n=88 and 215 for LCNI-5 and iLiNS, respectively). Free HMOs from whole milk were extracted using previously reported methods (Ninonuevo et al, 2006; Wu et al, 2010; Wu et al, 2011; Totten et al. 2014). Briefly, 50 µL aliquots of breast milk
samples obtained from mothers in the LCNI-5 cohort and 25 µL aliquots of samples obtained from members of the iLiNS-DYAD-M cohort were distributed into 96-well plates, diluted and defatted by centrifugation. To precipitate proteins, two volumes of ethanol were added to the aqueous layer and the mixtures incubated for 1.5 hours at -80°C. After centrifugation, the supernatant was collected, and the resulting glycans were reduced with 1.0 M NaBH₄ for 1.5 hours at 65°C. Samples were then purified by solid phase extraction, loaded onto graphitized carbon cartridges, desalted with nanopure water and then eluted with 20% acetonitrile in water and 40% acetonitrile in 0.05% trifluoroacetic acid (v/v). The eluent fractions were combined and the solvent evaporated. After reconstitution, the samples were diluted to appropriate concentrations and subjected to nano-high performance liquid chromatography (HPLC)-chip/time-of-flight (TOF) mass spectrometry.

The Agilent 1200 series liquid chromatography unit utilizes a dual pump system with sample loading and analyte separation performed on a microfluidic chip. The chip has a 40 nL enrichment column and a 75 µL x 43 mm analytical column packed with porous graphitized carbon. This system is coupled to an Agilent 6220 series TOF mass spectrometer via chip-cube interface. The HPLC capillary pump loads the sample onto the chip’s enrichment column at a flow rate of 4.0 µL/min with a 1 µL injection volume, and the nano pump achieves separation with a binary gradient of aqueous solvent (3% acetonitrile/water (v/v) in 0.1% formic acid (FA) and organic solvent (90% acetonitrile/water (v/v) in 0.1% FA). The separation method was developed and optimized for HMO mixtures (Wu et al, 2010; Wu et al, 2011). To monitor instrument performance and minimize batch effects, an external HMO pool was run intermittently throughout sample analysis.

Data were collected using Agilent MassHunter Workstation Data Acquisition software, version B.02.01 and then analyzed using Agilent MassHunter Qualitative Analysis software version B.03.01. HMO compounds were identified using the ‘Find Compounds by Molecular Feature’ function and peak alignment was performed using custom scripts. Specific structures were assigned by matching exact mass (within 20 ppm mass error of theoretical values) and retention time to established HMO libraries (Wu et al, 2010; Wu et al, 2011). Absolute abundances in ion counts were normalized to total HMO abundance found in the external HMO pool run respective to that
sample’s batch. Glycan types were divided into four classes: fucosylated, sialylated, fucosylated and sialylated, and non-fucosylated neutrals. Relative class abundances were calculated by dividing absolute class abundance by each mother’s total HMO abundance. The same normalization and relative abundance calculations were performed for individual structures.

Phenotypic secretor status for all mothers was determined by analyzing abundances of individual HMO structures and examining their chromatographic profiles. LCNI-5 secretor status assignment was based on a previously developed method (Totten et al, 2012). Secretor status assignment for mothers in the iLiNS-DYAD-M cohort also used this method but with the following modifications: absolute abundances of compounds with known α1-2 fucose linkage [fucose(α1-2)galactose] were summed and normalized to each mother’s total oligosaccharide abundance in order to calculate relative α1-2 fucosylation for each mother. The structures used for this determination were 2’fucosyllactose (2’FL), lactodifucotetraose (LDFT), trifucosyllacto-N-hexaose (TFLNH), difucosyllacto-N-hexaose a (DFLNa) and fucosyl-iso-lacto-N-hexaose I (IFLNHI) (Newburg and Neubauer, 1995). Secretor status was confirmed by comparing the extracted ion chromatograms of those structures between assigned secretors and non-secretors.

Generating a clonally arrayed sequenced collection of bacterial strains from a 6-month-old severely stunted Malawian infant

A de-identified fecal sample, collected and immediately frozen in a dry cryogenic storage dewar (pre-charged with liquid nitrogen), from a 6-month-old infant that had been enrolled in a previously reported study of twins living in five rural southern Malawian villages (patient ID h264A in Yatsunenko et al, 2012), was used to produce a clonally arrayed collection of anaerobic bacterial strains with a protocol modified slightly from what we had described previously (Faith et al, 2013; Goodman et al, 2011; Ridaura et al, 2013). An aliquot of the frozen sample was pulverized in a Biosafety Class II hood with a ceramic mortar and pestle filled with liquid nitrogen. An aliquot (1g) of the pulverized material, sealed in a sterile screw-capped tube (Axygen SCT-200-C-S), was brought into an anaerobic Coy chamber (atmosphere: 20% CO₂, 5% H₂, and 75% N₂), immediately
suspended in 15 mL Gut Microbiota Medium (GMM; Goodman et al, 2011) and blended with a Waring MC1 blender set on maximum speed (four cycles of blending for 20 seconds followed by a 30 second pause). The sample was allowed to stand for 5 minutes so that particulate matter could settle by gravity; the resulting supernatant was passed through a 100 μm pore diameter filter (BD systems, Inc. Franklin Lakes, NJ) to remove remaining particulate material, mixed with an equal volume of pre-reduced GMM containing 30% glycerol (final concentration 15% glycerol) and placed in Wheaton crimp top tubes for storage at -80°C.

A vial of the -80°C anaerobic glycerol stock containing an aliquot of the clarified sample was diluted into pre-reduced GMM lacking resazurin in an anaerobic chamber and dispensed into a pre-reduced 384-well flat-bottomed polypropylene multi-well plate (170 μL per well; Costar #3964, Corning, NY). All liquid handling steps were performed using a set of custom interfaces for a Precision XS robot (BioTek) so that picking, arraying and archiving of the bacterial culture collection could be done with economy and speed under anaerobic conditions.

To determine the dilution at which approximately 30% of wells would receive one or more bacterial cells in the initial inoculation, two- and fourfold serial dilutions of the clarified fecal suspension were performed (from $10^{-6}$ to $10^{-10}$) in a trial inoculation (48 wells per dilution; 170 μL per well). Plates (Nunc 260251) were sealed with sterile foil lids (VWR 60941-076) and incubated in a Coy chamber for 3 days at 37°C under anaerobic conditions (atmosphere: 20% CO$_2$, 5% H$_2$, and 75% N$_2$). The dilution at which ~30% of wells were turbid ($OD_{630} > 0.1$) was chosen for a subsequent large-scale culturing effort. To do this, a second vial of the frozen clarified fecal sample was added to 500 mL of pre-reduced GMM at the calculated dilution and dispensed into 384-well plates as above (170 μL per well, 3,840 wells total) and incubated at 37°C under anaerobic conditions for 3 days. Frozen anaerobic glycerol stocks of each plate were prepared by resuspending cells (by pipetting) and transferring 25 μL aliquots to a pre-reduced 384-well plate containing an equal volume of pre-reduced GMM with 30% glycerol. Glycerol stock plates were frozen on dry ice inside the anaerobic chamber and stored in a -80°C freezer. The anaerobic status of frozen stocks was verified by including resazurin in the storage medium. An additional 60 μL
of the resuspended cell culture was taken from each well for measuring OD$_{630}$ and sequencing of bacterial genomic DNA.

Based on the results of sequencing amplicons generated from the V4 region of each isolate’s 16S rRNA gene, the library was subsequently condensed to include up to seven representatives of each unique V4 16S rRNA sequence. (Condensing at this step consisted of recovering the frozen 384-well stock plates in GMM, purifying each isolate on GMM agar plates and inoculating fresh 96-well plates with the isolates).

Genomic DNA from each isolate was subjected to shotgun sequencing in multiplex using a HiSeq2000 instrument [101 nt paired-end reads, 44.65±29.32 fold coverage (mean±SD), see Table S2A]. Genomes were assembled using Velvet and Velvet Optimizer, version 2.1.7. Genes encoding proteins, tRNAs and rRNAs were annotated with Glimmer3.0, tRNAscan 1.23 and RNAmmer 1.2, respectively.

Shared genome content at the nucleotide level was computed using Nucmer as described in an earlier report (Faith et al, 2013). Shared genome content was calculated between two genomes (A and B) by the following formula,

$$\text{Shared Genome Content} = \frac{X + Y}{\text{len}(A) + \text{len}(B)} \times 100\%$$

where X is the length of A that aligns to B, Y is the length of B that aligns to A, len(A) is the length of A, and len(B) is the length of B. Isolates were clustered at the level of >96% shared genome content using Nucmer (Delcher, 2002). A single representative of each cluster was selected based on draft genome sequence quality and used to generate the 25-member strain collection.

Each of the selected isolates listed in Table S2A was re-purified on LYHBHI agar plates (Sokol et al, 2008) and a full-length 16S rRNA amplicon was generated [50 μL PCR reactions contained 44 μL ReddyMix PCR master mix (Thermo Scientific#AB-0575/LD), 2 μL Forward primer (8F; 10 μM) and 2 μL Reverse primer (1392R; 10 μM); cycling conditions were 95°C for 10 minutes, followed by 35 cycles of 95°C for 45 seconds, 56°C for 45 seconds, and 72°C for 2 minutes,
followed by 72°C for 10 minutes]. Amplicons were sequenced using the dideoxy chain termination method (Retrogen, Inc; San Diego, CA). Our designation of a strain as *B. longum* subsp. *infantis* was based on (i) the presence of genes encoding urease subunits (LoCascio et al, 2010) and (ii) the presence of HMO utilization gene clusters containing all four glycoside hydrolases active on HMO linkages [β-galactosidase, sialidase, fucosidase and hexosaminidase] (Sela et al, 2008); see Table S2C.

The 25-strain library was stored in a “sparse” 96-well format where members were deliberately spaced across the plate in between un-inoculated wells employed as a ‘buffer’ to monitor potential contamination events. To prepare the strain collection for gavage into germ-free animals, a frozen archived ‘stock plate’ was thawed at room temperature inside the anaerobic chamber and a 30 µL aliquot of the contents of each well was transferred using a Precision XS robot (BioTek) into 600 µL of either fresh pre-reduced GMM or pre-reduced LYHBHI medium (Sokol et al, 2008) located in the wells of a new 96-well microtiter plate (Nunc, #260251). Following an overnight incubation at 37°C in an anaerobic chamber, cultures were independently assayed for growth (OD$_{630}$), an equal volume (200 µL) was taken from each well and the aliquots were pooled. An equal volume of pre-reduced PBS containing 30% glycerol was mixed with the pooled cultures, and the mixture was dispensed, in 1 mL aliquots, into Wheaton crimp cap tubes and stored at -80°C until use. Stocks of the pooled material were kept at -80°C and then thawed immediately prior to gavage of gnotobiotic animals.

**Purification and compositional analysis of S-BMO**

S-BMO was purified from commercial whey permeate after lactose was partially removed by concentration and crystallization. BMO in delactose permeate was first concentrated with a single stage ultrafiltration membranes (molecular mass cut off of 1000 Da; GE Waters, Inc.; Boulder, CO, USA) at 30°C. S-BMO was then extracted from this ultrafiltrate by anion-exchange chromatography on a column containing 100 L of dimethylamine functionalized chloromethylated copolymer of styrene and divinyl benzene (Bucher-Alimentech Ltd.; Auckland, NZ). Chromatography was
run at 150 liter/hour at 30°C. The S-BMO peak collected from anion-exchange chromatography was further concentrated by ultrafiltration at 30°C using the same type of membrane to remove minerals in the solution. The final concentrate was freeze-dried and stored in a vacuum desiccator at room temperature.

The total carbohydrate composition of S-BMO was determined on an Agilent 6520 accurate-mass Q-TOF LC/MS with a microfluidic nano-electrospray chip and high-performance anion-exchange chromatography with pulsed amperometric detection (Thermo Scientific HPAE-PAD ICS-5000, Sunnyvale, CA, USA).

**Design and preparation of the Malawi 8 (M8) diet**

This diet was formulated based on a dietary survey of the complementary feeding practices of 43 nine-month-old Malawian infants and children enrolled in the iLiNS-DOSE clinical study (#NCT00945698) that took place in the Mangochi district of Malawi. The United States Department of Agriculture Nutrient Database (http://ndb.nal.usda.gov/ndb/search/list) was used estimate the macro- and micronutrient content of components of their diets. We then used linear programming to calculate combinations of ingredients that resembled the mean energy and nutrient values for diets identified from the dietary survey. *Table S3B* lists the composition of food ingredients in M8.

The M8 diet was prepared for administration to gnotobiotic animals as follows: Meseca® corn flour was obtained from Restaurant Depot (College Point, NY). The remaining ingredients were purchased from Whole Foods Supermarkets. The diet was prepared in batches of 20 kg. A relish containing 2kg mustard greens, 1.5kg onions and 1.5kg tomatoes was pureed in a food processor (Robot Coupe Model R23, Jackson, MS) and the puree was cooked in 1L of water for 60 minutes on a Corning stirrer/hot plot (high setting, until browned). After cooking, the relish was combined with a pureed mixture of 1kg ground peanuts, 700g soaked red kidney beans, 1kg canned pumpkin and 2.5kg peeled bananas in an industrial mixer (Globe SP30P 30-quart pizza mixer; gear speed 1; Globe Food Equipment Company, Dayton, OH). Corn flour (5kg) and hot,
freshly autoclaved water (5L) were then added slowly and mixed using the industrial mixer for 5 min.

Purified powdered S-BMO or inulin (from dahlia tubers, Santa Cruz Biotechnology, Inc., Product Number 9005-80-5) were then added to the M8 diet at the concentrations noted in the main text. Batches of unsupplemented M8, S-BMO-supplemented M8 or inulin-supplemented M8 diets were allowed to cool in clean plastic containers in a 4°C cold room prior to vacuum packing in 500 g aliquots in FDA/USDA-compliant poly-nylon vacuum pouches (#S-7556; Uline, Pleasant Prairie, WI). Sealed diet aliquots were double-bagged and sterilized by irradiation (20-50 kGy) within 24 hours of production (Steris Co; Chicago, IL).

For experimental convenience and consistency, dry pelleted versions of M8 and S-BMO-supplemented M8 were also obtained from Dyets, Inc. (Bethlehem, PA). Pellets were extruded with a 1/2-inch diameter for gnotobiotic mouse studies and a 3/16-inch diameter for gnotobiotic piglet studies.

The nutritional content of all cooked and irradiated custom diets was defined by N.P. Analytical Laboratories [St Louis, MO (Table S3C)]. Measurement of chow consumption by gnotobiotic animals confirmed comparable caloric and nutritional consumption between paste and pelleted versions of the diets (data not shown). Irradiated food was stored at 4°C for up to six months. Sterility was determined by resuspending a small aliquot of each batch of food in pre-reduced GMM under anaerobic conditions and incubating the suspension for 3 days at 37°C. Sterility was further verified by subculture on pre-reduced anaerobic GMM agar plates.

Studies involving gnotobiotic mice

All gnotobiotic mouse experiments were performed using protocols approved by the Washington University Animal Studies Committee. Male germ-free C57Bl/6J mice were maintained in sterile, flexible, plastic gnotobiotic isolators (Class Biologically Clean Ltd., Madison, WI) under a strict 12-hour cycle (lights on at 0600, off at 1800 h). Mice were fed an autoclaved low-fat, polysaccha-
ride-rich chow (LF/HPP) diet (B&K University, East Yorkshire, U.K.; diet 7378000) from weaning until 3 days prior to the beginning of an experiment. At that time, 4.5-week old male animals were switched to the M8 diet or the M8 diet supplemented with S-BMO or inulin for the remainder of the experiment. Mice received a single oral gavage (200 µL) of the complete 25-member culture collection or subsets of the collection 3 days after switch to one of the M8-based diets.

**Quantitative magnetic resonance (qMR) analysis of body composition**

Body composition was defined using an EchoMRI-3in1 instrument (EchoMRI, Houston, TX). Each mouse was transported from its gnotobiotic isolator to the MR instrument in a HEPA filter-capped glass vessel.

**Sample collection**

Fecal samples were collected at defined times after gavage. At the time of euthanasia, blood, cecal contents, liver, brain and skeletal muscle (soleus or gastrocnemius) were collected and immediately frozen in liquid nitrogen. Soleus and gastrocnemius muscle samples were freeze-clamped after euthanasia *in situ* using pre-cooled forceps prior to harvest and storage.

**Micro-computed tomography**

Femurs were harvested from mice at time of euthanasia, cleaned of soft tissue, fixed for 24 hours in 10% formalin and stored at 4°C in 70% ethanol until scanning. Micro-computed tomography was performed using a µCT 40 desktop cone-beam instrument (ScanCO Medical, Brüttisellen, Switzerland). For cortical analyses, 200-300 slices were taken for each sample in the transverse plane with a 6 µm voxel size (high resolution). For all cortical scans, slices began at the midpoint of the femur and extended toward the distal femur. Boundaries of and thresholds for bone were drawn manually using µCT 40 software. Calculation of volumetric parameters (bone volume/tissue volume, bone mineral density and cortical thickness) was automated with custom scripts.
**Histological analysis of distal femur morphology**

Previously fixed femurs were decalcified by incubation at room temperature in 15 mL of 14% EDTA, pH 7.2 (#EDS; Sigma Aldrich) for 14 days with gentle agitation. The EDTA solution was replaced daily. Decalcified femurs were subsequently washed four times with deionized water, rehydrated with stepwise immersions in 30% ethanol, 50% ethanol and 70% ethanol (30 minutes each), embedded in paraffin, cut into 5 μm thick longitudinal sections and stained for tartrate-resistant acid phosphatase (TRAP) for osteoclasts (Oddie et al, 2000) and hematoxylin for nuclei. Sections were imaged at 100x magnification with an Olympus U-CMAD3 microscope. The resulting images were analyzed using BioQuant OSTEO 2012 software (BIOQUANT Image Analysis Co.; Nashville, TN). Personnel were blinded to sample identifications. The region of interest for analysis of trabecular bone was defined for each section as the 2-dimensional area beginning 750 μm proximal to the distal growth plate and extending to the very proximal edge of the growth plate and spanning the full width of the bone. Osteoclasts were identified as TRAP+ cells. Regions located adjacent to the growth plate where osteoclast density was too high to accurately count cells were excluded from the analysis.

**Multiplex sequencing of amplicons generated from bacterial 16S rRNA genes**

Genomic DNA was extracted by resuspending fecal pellets (~50 mg) or cecal contents (~50 mg) in a solution containing 500 μL of extraction buffer [200 mM Tris (pH 8.0), 200 mM NaCl, 20 mM EDTA], 210 μL of 20% SDS, 500 μL phenol:chloroform:isoamyl alcohol (pH 7.9, 25:24:1, Ambion) and 500 μL of 0.1-mm diameter zirconia/silica beads. Cells were mechanically disrupted using a bead beater (BioSpec Products, Bartlesville, OK; maximum setting for 4 min at room temperature), followed by extraction with phenol:chloroform:isoamyl alcohol and precipitation with isopropanol.

Amplicons (~365bp) spanning variable region 4 (V4) of the bacterial 16S rRNA gene were generated by polymerase chain reaction (PCR) using primers and cycling conditions described in an earlier report (Subramanian et al, 2014). PCR primers incorporated sample-specific barcodes.
allowing samples to be subjected to multiplex sequencing using an Illumina MiSeq instrument (paired-end 250 nt reads). Paired V4-16S rRNA sequences were trimmed to 200 bp and merged into a single sequence with Flash software (Magoc and Salzberg, 2011). Merged sequences were filtered for low quality reads and binned according to their sample-specific barcodes. Reads were clustered into 97%ID OTUs using UCLUST (Edgar, 2010) and the Greengenes references OTU database. Reads that failed to hit the reference dataset were clustered \textit{de novo} using UCLUST. A representative OTU set was created using the most abundant OTU from each bin. Reads were aligned using PyNAST (Caporaso et al, 2010). A custom dataset of manually curated NCBI bacterial taxonomy (Ridaura et al, 2013) was used to train the Ribosomal Database Project (RDP) version 2.4 classifier (Wang et al, 2007b) and to assign taxonomy to picked OTUs. Validation of this assignment strategy is described in a previous publication (Ridaura et al, 2013). For analysis of bacterial abundance and colonization efficiency, rare OTUs (<0.1% relative abundance on average across all samples collected from a given treatment group and timepoint) were removed.

**Microbial RNA-Seq**

Protocols for microbial RNA-Seq are described in our previous publications (Faith et al, 2011; McNulty et al, 2013; Turnbaugh et al, 2009). A slightly modified procedure was followed for this study. Cecal contents (100-300mg), collected 44 days after gavage, were suspended in 1 mL RNAprotect bacteria reagent (Qiagen), vortexed for 5 min at room temperature and centrifuged for 10 min at 5000 x g (4°C). After decanting the supernatant, pelleted cells were suspended in 500 μL extraction buffer (200 mM NaCL, 20 mM EDTA), 210 μL of 20% SDS, 500 μL of phenol:chloroform:isoamyl alcohol (pH7.9, 125:24:1, Ambion) and 250 μL of acid-washed glass beads (Sigma-Aldrich, 212-300 μm diameter). Microbial cells were lysed by mechanical disruption using a bead beater (Biospec; maximum setting; 5 min at room temperature), followed by phenol:chloroform:isoamyl alcohol extraction and precipitation with isopropanol. RNA was treated with RNase-free TURBO-DNAsel (Ambion) and 5S rRNA and tRNAs were removed (MEGA-Clear columns, Ambion). A second DNase treatment was performed (Baseline-ZERO DNase;
Epicenter). RNA integrity was checked by gel electrophoresis (100 ng, 1% agarose gel). 16S and 23S rRNA was depleted using the Ribo-zero meta-bacterial rRNA removal kit (Epicenter), followed by ethanol precipitation. cDNA was synthesized using SuperScript II (Invitrogen), followed by second strand synthesis with RNaseH, E. coli DNA polymerase (NEB) and E. coli DNA ligase (NEB). Samples were sheared using a BioRuptor XL sonicator (Diagenode); 200-300 bp fragments were gel selected and prepared for sequencing.

Multiplex microbial RNA-Seq for gnotobiotic mouse experiments was performed using an Illumina Hi-Seq2000 instrument to generate 10.78±1.45 million (mean±SD) unidirectional 101 nt reads per sample. Sequencing reads were split according to the 8-bp barcodes used to label each of the twelve samples pooled together per HiSeq lane (sequencing was performed over three lanes with reads pooled for analysis).

Reads were mapped to genes in a custom database of 25 unique draft genomes included in the defined Malawian infant bacterial community. If a read mapped to more than one location in a genome or to multiple genomes, the counts for each gene were added according to the gene’s fraction of unique-match counts. Read counts (rounded to the nearest integer) for each gene were split by genome and then normalized and analyzed with the R statistical package, DESeq (Anders and Huber, 2010).

**Community profiling by sequencing (COPRO-Seq)**

Protocols for determining bacterial community composition by COPRO-Seq are described in our previous publication (McNulty et al, 2011). Briefly, DNA samples were sheared using a BioRuptor Pico sonicator (Diagenode); 200-300 bp fragments were gel selected and prepared for sequencing. Samples were sequenced in multiplex using an Illumina NextSeq instrument to generate unidirectional 75 nt reads. Sequencing reads were split according to 8 bp barcodes assigned to each sample, processed and mapped to the draft genomes of members of the Malawian infant’s culture collection using COPRO-Seq software (https://github.com/nmcnulty/COPRO-Seq).
Mass spectroscopy of metabolites contained in biospecimens obtained from gnotobiotic animals

**Targeted ultra high performance liquid chromatography-mass spectrometry (UPLC-MS) of liver, muscle, brain, serum and cecal contents**

Liver, brain, muscle and cecal samples were homogenized in 50% aqueous acetonitrile containing 0.3% formic acid (50 mg wet weight tissue/mL solution) using a high-speed homogenizer (IKA #EW-04739-21) set at maximum speed for 30-45 seconds. Samples were maintained on ice or dry ice throughout this procedure.

Amino acids, acylcarnitines, organic acids, acyl CoAs and ceramides were analyzed using stable isotope dilution techniques. Amino acids and acylcarnitine measurements were made by flow injection tandem mass spectrometry using sample preparation methods described previously (An et al, 2004, Ferrara et al, 2008). Data were acquired using a Waters Acquity™ UPLC system equipped with a TQ (triple quadrupole) detector and a data system controlled by MassLynx 4.1 operating system (Waters, Milford, MA). Organic acids were quantified according to a previously published protocol (Jensen et al, 2006) using Trace Ultra GC coupled to ISQ MS operating under Xcalibur 2.2 (Thermo Fisher Scientific, Austin, TX). Acyl CoAs were extracted and purified as described (Deutsch et al, 1994; Magnes et al, 2005; Minkler et al, 2008), and analyzed by flow injection analysis using positive electrospray ionization on Xevo TQ-S, triple quadrupole mass spectrometer (Waters, Milford, MA). Heptadecanoyl CoA was employed as an internal standard. Ceramides were extracted using a previously reported procedure (Merrill et al, 2005) and analyzed by flow injection tandem mass spectrometry using a Xevo TQS spectrometer (Waters Milford, MA,) for precursors of m/z 264.

**Non-targeted gas chromatography/mass spectrometry (GC-MS) of serum**

Serum samples were spiked with perdeuterated myristic acid (D27-C14:0) as an internal standard for retention time locking (RTL IS). Following treatment with 7.5 volumes of methanol, the mix-
ture was centrifuged and the supernatant was decanted and dried. Derivatization of all dried supernatants for GC-MS followed a method adapted with modifications from Roessner et al. (2000). Reagents were from Sigma-Aldrich (St. Louis, MO), unless otherwise noted. Briefly, certain reactive carbonyls were first methoximated at 50°C with a saturated solution of methoxyamine hydrochloride in dry pyridine, followed by replacement of exchangeable protons with trimethylsilyl (TMS) groups using $N$-methyl-$N$-(trimethylsilyl) trifluoroacetamide with a 1% v/v catalytic admixture of trimethylchlorosilane (Thermo-Fisher Scientific), also at 50°C. GC-MS methods generally used a 6890N GC connected to a 5975 inert single-quadrupole MS (Agilent, Santa Clara, CA). A large-volume ProSep inlet (Apex Technologies, Inc.; Independence, KY) enabled programmed temperature vaporization and diversion of heavy contaminants away from the GC and MS. The two wall-coated, open-tubular GC columns connected in series were both from J&W/Agilent (part 122-5512, DB5-MS; 15 m in length, 0.25 mm in diameter, with a 0.25 µm luminal film). Prior to each run, initial inlet pressures were empirically adjusted such that the resulting retention time of the TMS-D27-C14:0 standard was set at ~16.727 min. Under these conditions, derivatized metabolites eluted from the column and reached the electron-ionization source in the MS at known times. A mid-column, microfluidic splitter (Agilent) provided a means for hot back-flushing of the upstream GC column at the end of each run while the oven was held at 325°C for a terminal “bake-out” [another antifouling and anti-carryover measure analogous to that described in (Chen et al, 2009)]. During this terminal “bake-out,” the inlet was also held at 325°C and purged with a large flow of the carrier gas, helium. Positive ions generated with conventional EI at 70 eV were scanned broadly from 600 to 50 m/z in the detector throughout the run.

Raw data from Agilent’s ChemStation software environment were imported into AMDIS (Automatic Mass Spectral Deconvolution and Identification Software; Stein, 1999). Deconvoluted spectra were identified, to the extent possible, using several commercial and public spectral libraries. Our primary source was the Fiehn GC-MS Metabolomics RTL Library (a gift from Agilent Technologies, Santa Clara, CA; part number G1676-90000). Additional spectra for comparison were gleaned from the Golm Metabolome Library (Kopka et al, 2005), the commercial NIST/
EPA/NIH Mass Spectral Library and our own purpose-built spectral library. Where indicated, peak alignment was performed with SpectConnect freeware (Styczynski et al, 2007). Chemometrics were performed with Mass Profiler Professional (Agilent), along with our own custom macros, written in Visual Basic for use in the Excel software environment.

Non-targeted analysis of brain metabolites using GC-MS

Left cerebral hemispheres of mouse brains were homogenized using the same procedures described above for UPLC-MS of liver, serum and cecal contents. Homogenates were centrifuged (20,800 x g for 10 min at 4°C). A 200µL aliquot of the resulting supernatant was transferred to a clean tube and combined with 400 µL ice-cold methanol. The mixture was subsequently vortexed and centrifuged, and a 500 µL aliquot of the resulting supernatant, together with 10 µL of lysine-$^{13}$C$_6$,$^{15}$N$_2$ (2 mM), was evaporated to dryness using a speed vacuum. To derivatize the sample, 80 µL of a solution of methoxylamine (15 mg/mL in pyridine) was added to methoximate reactive carbonyls (incubation for 16 h for 37°C), followed by replacement of exchangeable protons with trimethylsilyl groups using N-methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) with a 1% (vol/vol) catalytic admixture of trimethylchlorosilane (Thermo-Fisher Scientific) (incubation for 1h at 70°C). Heptane (160 µL) was added and a 1 µL aliquot of each derivatized sample was injected into an Agilent 7890A gas chromatography system, coupled with a 5975C mass spectrometer detector (Agilent, CA). The quantity of each metabolite was determined from spike-in internal standards.

Non-targeted UPLC-MS of brain metabolites

Frozen mouse left cerebral hemisphere samples were homogenized with 20 vol/wt of cold methanol and 1 vol/wt cysteine $^{13}$C$_6$,$^{15}$N$_2$ (4 mM). Samples were subsequently incubated at -20°C for 1h and centrifuged 10 min at 20,800 x g. The resulting supernatant (300µL) was collected and dried in a SpeedVac at room temperature. Dried samples were resuspended in 100 µL of 95:5 water:ethanol, clarified for 5 min by centrifugation at 20,800 x g for 10 min at 4°C and the super-
Experiments involving gnotobiotic piglets

All experiments involving pigs were performed under the close supervision of a veterinarian using protocols approved by the Washington University Animal Studies Committee.

Derivation of germ free piglets

Pregnant domestic sows (mixture of Landrace and Yorkshire genetic backgrounds, mated to Duroc breed domestic boars) were obtained from Oak Hill Genetics (Ewing, IL). Piglets were delivered one day prior to the anticipated date of farrow. To prepare for surgery, the sow was sedated with ketamine (20 mg/kg administered intramuscularly) and anesthetized with isofluorane (2-3%, delivered by mask or endotracheal tube). A local incisional block was achieved using 60-80 mL of 2% lidocaine injected subcutaneously. The sow’s paralumbar abdominal area was shaved, washed, and disinfected with povidone-iodine (1% titratable iodine). Piglets were delivered by making a paralumbar abdominal incision on the sow, exteriorizing the uterus and removing piglets individually into the sterile surgical field. Each piglet’s umbilical cord was tied off to prevent bleeding and then severed. Piglets were then immediately passed, prior to their first breath, through a sterile tank filled with 10% betadine into a sterile gnotobiotic ‘nursery’ isolator made of flexible plastic. Piglets were revived within the ‘nursery’ isolator and kept on a heated pad until the procedure was complete. At the completion of surgery, the sow was euthanized by pentobarbital overdose (>150 mg/kg intravenously).

Immediately after piglets were revived (breathing independently and moving), they were transported within the nursery isolators from the operating room to a gnotobiotic facility. Within 24 hours, all piglets were transferred from the initial nursery isolators to larger, steel gnotobiotic
isolator tubs (Class Biologically Clean; Madison, WI) where they were maintained for the duration of the experiment. Steel isolators were equipped with a flexible plastic canopy allowing trained personnel to examine and interact with the animals. Gnotobiotic piglets were group-housed (three piglets per isolator, complying with USDA animal housing regulations) and closely monitored throughout the experiment.

The room where gnotobiotic piglets were housed in their isolators was maintained at 85°C-90°C. Gnotobiotic isolators were maintained at 95°C-100°C using electric heaters positioned near their exterior (Cuori #HD904-A7Q 1500-watt electric oil-filled radiant portable heaters). Thermometer probes were positioned near inlets and outlets for airflow to monitor isolator temperatures. Piglet weights were measured using a hanging scale after carefully tying a rope around the piglet’s abdomen and forelimbs.

**Experimental design**

Piglets were initially fed a powdered sow’s milk replacement (Soweena Litter Life, Merrick #C30287N). The powdered sow’s milk replacement was prepared in 120g vacuum-sealed, gamma-irradiated packets for reconstitution in gnotobiotic isolators (120g/1L autoclaved water). Piglets were fed by bottle for the first 2-3 days of life as they were trained to drink independently from troughs. Bottle-feeding was withdrawn when all piglets were observed to be drinking independently. Piglets were slowly weaned from sow’s milk replacement formula to custom M8 pelleted chow starting at postnatal day 5 of life and completing on day 11. To do so, pelleted chow was introduced as creep feed in addition to formula, and the amount of formula provided to piglets was reduced stepwise until all piglets were observed eating pelleted chow. A small amount of sterile water was added to pelleted chow to encourage eating, and water was made available *ad libitum* to all animals in separate troughs. Behavior and the condition of all piglets were monitored multiple times each day to ensure their well-being. Piglets were fed at 3-hour intervals on days 1-2 after birth, at 4-hour intervals during days 3-4 and at 8-hour intervals from postnatal day 5 until the end of the experiment.
The germ-free status of piglets prior to colonization on postnatal day 3 was confirmed by aerobic and anaerobic culture of rectal swabs in LYHBHI medium (Sokol et al, 2008). Anaerobic cultures were taken by placing rectal swabs in anaerobic transport medium tubes (Anaerobe Systems ATM 6.0ML #AS-911) for transport to an anaerobic chamber (atmosphere: 20% CO₂, 5% H₂, and 75% N₂).

Gnotobiotic piglets were colonized at day 3 of life by oral gavage. Briefly, equal volumes of overnight pre-reduced LYHBHI cultures of the 17 strains from the Malawian infant’s bacterial culture collection that lacked more than three known virulence factor were mixed, diluted 1:10 in LYHBHI medium and mixed with an equal volume of pre-reduced PBS containing 30% glycerol [final glycerol concentration 15% (vol/vol)]. Aliquots of the bacterial consortium were stored at -80°C until use. For each piglet, 1 mL of this inoculum was resuspended in 10 mL of sow’s milk replacement formula and introduced by intragastric gavage using a Kendall Kangaroo™ 2.7 mm diameter feeding tube (Cat #8888260406; Covidien, Minneapolis, MN).

Gnotobiotic piglets were euthanized on day 19 of postnatal life. Piglets were removed from the gnotobiotic isolator, sedated with ketamine (20 mg/kg IM), anesthetized with isofluorane and euthanized by pentobarbital overdose (>150 mg/kg IV). Whole blood was collected under anesthesia in Vacutainer tubes (BD #366404), immediately prior to euthanasia, stored on ice for at least 30 minutes and then centrifuged at 4000 × g for 5 minutes. Serum was collected, separated into small aliquots, and stored at -80°C. Other tissues were collected immediately after euthanasia, frozen in liquid nitrogen and stored at -80°C. Muscle samples were collected from the biceps femoris immediately after euthanasia and frozen in liquid nitrogen within 3-4 minutes.

**Microbial RNA-Seq**

Fecal samples were collected at the time of euthanasia on day 19 of postnatal life, RNA was extracted and microbial RNA-Seq was performed using an Illumina NextSeq instrument to generate 42.5±15.2 million unidirectional 75 nt reads per sample. Reads were demultiplexed using the 8 nt sample-specific barcode adapters and processed/mapped to the draft genomes of strains in the
culture collection using a set of custom scripts. Read counts were normalized and analyzed using DESeq (Anders and Huber, 2010).

**Gut mucosa glycan analysis by triple quadrupole mass spectrometry**

Mucosal samples were collected from three segments of the gut: (i) proximal small intestine (defined as a 10 cm segment distal to the stomach); (ii) distal small intestine (a 10 cm segment proximal to the cecum), and (iii) the spiral colon (10 cm segment distal to the cecum). Each of these segments was opened with clean steel scissors and luminal contents were gently removed with a clean steel spatula. The mucosa was gently scraped with a separate clean steel spatula and scrapings from each segment immediately frozen in liquid nitrogen. In addition, cecal contents and feces were recovered.

Each sample was diluted in water to 0.1 mg/10 µL, vortexed and briefly centrifuged at 2,000 x g. 100 µL of the resulting supernatant was taken and mixed with 100 µL of 0.2 M trifluoroacetic acid (TFA), then heated at 80°C for 1 hour. The solvent was then dried using a centrifugal evaporator. Samples were subsequently reconstituted in 200 µL of water and centrifuged at 21,130 x g for 30 min. 10 µL of the resulting supernatant was then added to 40 µL water and centrifuged at 21,130 x g for 20 min.

Analysis was performed on an Agilent ultra high-pressure liquid chromatography-electrospray ionization (UPLC-ESI)-triple quadrupole (QqQ) mass spectrometer in dynamic multiple reaction monitoring (MRM) mode. The Agilent 1290 infinity UPLC system incorporates a binary solvent system of solvent A [3% acetonitrile/water (v/v) in 0.1% formic acid (FA)] and solvent B [95% acetonitrile/water (v/v)] with an injection volume of 1 µL and flow rate of 0.3 mL/min with a Waters ACQUITY UPLC BEH amide column. Samples were run under isocratic conditions of 85% solvent A and 15% solvent B with a total run time of 3 min. The UPLC system is coupled to an Agilent 6490 QqQ mass spectrometer that was run in the positive mode. The MRM transitions used for monitoring sialic acid were ([M + 1H]1+ 310.2 m/z 197.0 and m/z 121.1). MRM data were analyzed using Agilent MassHunter Quantitative Analysis B.05.02 software.
Measurement of serum leptin, insulin and triglycerides

Mouse serum triglycerides were measured using a DxC600 clinical analyzer (Beckman 445850; Brea, CA). Mouse insulin and leptin were measured using a duplex assay and an SI-2400 electro-chemiluminescent imager (Meso Scale Discovery K15124C-1; Rockville, MD).

Porcine insulin and leptin were measured using a cross-reactive duplex assay developed for human sera (Meso Scale Discovery K151BYC-1; Rockville, MD). Porcine leptin was also measured with a human leptin kit (R&D Systems DLP00; Minneapolis, MN) as well as a porcine-specific kit from Biomatik (EKU05595; Wilmington, DE) using an SpectraMax M2e plate reader (Molecular Devices; Sunnyvale, CA). Porcine leptin was below the limit of detection by both methods.

In vitro S-BMO consumption and growth assays

Frozen stock cultures were recovered by incubation overnight at 37°C in an anaerobic chamber in pre-reduced LYHBHI broth (Sokol et al, 2008). For incubation with S-BMO, strains were first grown overnight in 5 mL pre-reduced LYHBHI to an OD$_{600}$ of approximately 0.8, then pelleted by centrifugation at 2600 x g for 5 minutes and washed twice with pre-reduced sterile PBS. Cell pellets were resuspended in 5 mL pre-reduced PBS containing 5% (wt/vol) and incubated at 37°C for 24 hours under anaerobic conditions (atmosphere: 20% CO$_2$, 5% H$_2$, and 75% N$_2$). Cells were pelleted by centrifugation at 2600 x g for 5 minutes, and the resulting culture supernatants were collected for UPLC-MS or GC-MS analysis of metabolites.

For E. coli growth experiments, starter cultures of E. coli strain MC1 were grown overnight in pre-reduced LYHBHI to an OD$_{600}$ of approximately 0.8, pelleted, washed as above and resuspended in 5 mL pre-reduced sterile PBS. E. coli cultures were started using a 100-fold dilution in a 200 µL volume of pre-reduced Davis Minimal Broth without dextrose (Sigma-Aldrich #15758, Saint Louis, MO) supplemented with 0.5% glucose, 0.5% sialic acid (Santa Cruz Biotechnology, Santa Cruz, CA), 2% S-BMO, 2% B. fragilis conditioned S-BMO (prepared by a 24 hour incuba-
tion of *B. fragilis* with S-BMO, as described above) or PBS only. All culture assays (*n*=8 technical replicates) were performed with negative controls (no *E. coli* inoculum) in a 96-well plate format. The OD$_{600}$ of cultures was monitored over a 24-hour period using a Biotek Eon plate reader (Biotek Instruments, Winooski, VT) within the anaerobic Coy chamber.

**Statistical analysis**

Multivariate analyses (O-PLS-DA) of metabolomics data were performed using the R package, Muma (Gaude et al, 2013). Univariate analyses (t-tests and ANOVA) were performed using Microsoft Excel or Prism 6.0 (GraphPad Software, Inc.). Heat maps were generated using Gene-E software ([http://www.broadinstitute.org/cancer/software/GENE-E/index.html](http://www.broadinstitute.org/cancer/software/GENE-E/index.html)). Two-way ANOVA of HMO and secretor status data from the LCNI-5 and iLiNS studies was performed by treating both factors (mother’s secretor status and infant HAZ bin) as categorical variables. HMO abundance was treated as a continuous response variable. Spearman’s rank correlation analyses were performed in the R environment using the test.cor function.
Supplemental References


Supplemental Figure Legends

**Figure S1:** Colonization with the Malawian infant bacterial culture collection is reproducible across recipient mice and piglets (relates to Figure 2, Figure 3 and Figure 6). (A,B) 16S rRNA-based weighted UniFrac principal coordinates plots displaying the phylogenetic structure of the fecal microbiota in recipient gnotobiotic mice as a function of time after colonization with all 25-members of the culture collection and diet (unsupplemented M8 diet or the isocaloric M8 diet supplemented with inulin or S-BMO) \( n=4-5 \) mice/treatment group). Panel A displays the first principal component (mean±SEM; 33.27% variance explained). Panel B displays the second principal component (mean±SEM; 21.84% variance explained). (C,D) 16S rRNA weighted UniFrac-based principal coordinates plots showing the first principal component (mean±SEM; 70.04% variance explained) and the second principal component (mean±SEM; 7.65% variance explained) of the fecal microbiota from gnotobiotic piglets colonized with 17-members of the defined Malawian infant bacterial community and fed either the M8 diet or M8 supplemented with S-BMO \( n=3 \) piglets/treatment group). The x-axis displays days after introduction of the bacterial community. Vertical dashed lines indicate the start of the period of S-BMO treatment.

**Figure S2:** S-BMO supplementation modulates levels of liver and serum metabolites in gnotobiotic mice (relates to Figure 5). (A) Heatmap showing liver acylcarnitines and fatty acyl CoAs whose concentrations in non-fasted mice harboring the 25-member Malawian infant community are affected by S-BMO supplementation (measurements made 37 days post-colonization). Rows represent replicate gnotobiotic mice grouped by dietary treatment. Columns represent individual metabolites identified by targeted MS/MS. *\( p<0.05 \) (two-tailed, unpaired Student’s t-test). Red asterisks indicate that levels of the metabolite are significantly different between groups after Benjamini-Hochberg correction for multiple hypotheses \( \alpha=0.1 \). (B-D) Concentrations of serum insulin, leptin and triglycerides in gnotobiotic mice harboring the Malawian infant community, fed M8 with or without S-BMO supplementation; samples collected 37 days post colonization from non-fasted animals. **\( p<0.01 \) (two-tailed, unpaired Student’s t-test).
**Figure S3: S-BMO supplementation alters levels of brain metabolites** (relates to Figure 5). (A) O-PLS-DA score plot (inset) and O-PLS-DA S-plot of metabolites measured by untargeted UPLC-MS in the left cerebral hemispheres of gnotobiotic mice fed M8 or M8 supplemented with S-BMO. The mass of the most strongly S-BMO-associated metabolite is indicated (putative assignment, N-acetylneuraminic acid). (B) N-acetylneuraminic acid (NeuAc) and (C) inosine abundance in the left cerebral hemisphere of gnotobiotic mice colonized with the Malawian defined community and fed M8 with or without S-BMO supplementation (measured 37 days after colonization). **p<0.01 two-tailed, unpaired Student’s t-test. See Table S7E,F for abundances of all measured left cerebral hemisphere metabolites.

**Figure S4: S-BMO supplementation increases expression of B. fragilis genes involved in branched chain amino acid biosynthesis in gnotobiotic piglets colonized with a 17-member consortium of bacterial strains from a stunted Malawian infant** (relates to Figure 6). Representation of the KEGG pathway for ‘valine, leucine, and isoleucine biosynthesis’, with enzymes labeled by enzyme commission (EC) number. Genes encoding enzymes with EC numbers highlighted with a red box exhibit significantly increased expression in B. fragilis in the context of S-BMO supplementation (negative binomial test after correction using the Benjamini-Hochberg procedure).
Supplemental Figures

Figure S1.

A  Mouse microbiota

Days post colonization

B  Mouse microbiota

Days post colonization

C  Pig microbiota

Days post colonization

D  Pig microbiota

Days post colonization
Figure S2.

A

Liver acyl CoAs

Liver acylcarnitines

Z-score

-1
0
1

Non-fasted metabolite

B

C

D

Serum insulin [pg/mL]

Serum leptin [pg/mL]

Serum triglycerides [mg/dL]

M8

M8 + S-BMO

Non-fasted serum metabolites

**

**

**

*
Figure S3.

A

B

C

Orthogonal component
Component 1 correlation
Component 1 covariance

NeuAc

M8
M8 + S-BMO

M8
M8 + S-BMO

Inosine

Spectral abundance [au]
Spectral abundance [au]
Spectral abundance [au]

**

***

ND

346.0546 Da

M8
M8 + S-BMO
Figure S4.
Supplemental Tables

Table S1. Breast milk HMO content analysis from two cohorts of Malawian mothers.

Table S2. A clonally-arrayed, genome-sequenced bacterial culture collection generated from a stunted, 6-month-old Malawian infant.

Table S3. Components of the prototypic Malawian diet and S-BMO supplement.

Table S4. COPRO-Seq and 16S rRNA analysis of relative abundances of Malawian infant culture collection strains in the fecal microbiota of gnotobiotic mice and piglets.

Table S5. Femur morphological parameters in gnotobiotic mice fed M8 with or without S-BMO supplementation.

Table S6. Escherichia coli MC1 genes differentially expressed with S-BMO supplementation in gnotobiotic mice.

Table S7. Tissue metabolite concentrations in gnotobiotic mice and pigs colonized with the stunted Malawian donor’s culture collection and fed the M8 diet with or without S-BMO supplementation.

Table S8. Genes in 17-member bacterial community that are differentially expressed with S-BMO supplementation in gnotobiotic piglets.
Chapter 4

Future Directions
Chapter 4

Future Directions

The findings presented in this dissertation demonstrate that immature gut microbiota from Malawian children with undernutrition transmit growth and metabolic abnormalities to young, recently weaned gnotobiotic mice consuming a prototypic Malawian diet (Chapter 2). The results reported in Chapter 3 go on to show that dietary supplementation with a purified preparation of sialylated bovine milk oligosaccharides (S-BMO) promotes increased body weight and lean body mass gain in the context of gnotobiotic mice and gnotobiotic piglets harboring a defined consortium of bacteria isolated from a stunted Malawian infant’s gut microbiota and fed a Malawian diet. Critically, this effect was dependent on the presence of the gut microbiota. S-BMO supplementation was also associated with increased femoral bone volume, increased sialic acid concentrations in brain, and responses to the feeding/fasting transition that are consistent with enhanced metabolic flexibility. This future directions chapter focuses on experimental approaches for (i) obtaining a more detailed understanding of the effects of S-BMO on growth and metabolism, (ii) identifying individual milk oligosaccharide structures that recapitulate the effects of S-BMO, and (iii) characterizing the microbial determinants of S-BMO-associated growth phenotypes. In addition, this chapter discusses the potential for, and challenges facing, milk oligosaccharide supplementation as a treatment for childhood undernutrition.

Characterizing the effects of S-BMO on bone morphology and metabolic flexibility

S-BMO supplementation has various effects on growth and metabolism in the context of gnotobiotic mice and piglets. A more detailed understanding of these effects is required to appreciate their mechanistic underpinnings and therapeutic potential. S-BMO treatment was associated with increased trabecular and cortical bone volume (normalized to tissue volume) in mouse femurs (Chapter 3), but whether this represents an increase in bone formation or a decrease in bone loss in S-BMO supplemented animals was not determined. One could address this question by apply-
ing dynamic histomorphometry techniques (Slyfield et al, 2012). With this approach, gnotobiotic mice, harboring the 25-member Malawian infant defined community and fed a Malawian diet with or without added S-BMO, would be injected with a marker of bone formation, such as calcein, in two doses (one week prior and one day prior to euthanasia). The marker would be incorporated into mineralizing bone, creating ‘fronts’ that would be visible by fluorescence microscopy of sections prepared from femurs. If S-BMO supplemented mice exhibit an increase in bone formation, one would expect to observe a larger distance between fronts (bone formation rate) and a greater labeled bone surface area (mineralizing bone) than what is observed in unsupplemented animals. If the observed bone volume difference accompanying S-BMO supplementation were a reflection of reduced bone loss, however, one would expect to see no detectable difference between bone marker fronts in either group of animals (Kulak and Dempster, 2010).

Compared to gnotobiotic animals that were not fed an S-BMO supplemented diet, supplemented animals displayed enhanced metabolic flexibility [an increase in serum fatty acid metabolites (acylcarnitines and fatty acyl CoAs) upon fasting]. Moving forward, follow-on experiments in gnotobiotic mice and piglets, serially sampled during feeding-to-fasting transitions would allow us to determine the period of time required after initiation of S-BMO supplementation for enhancement of metabolic flexibility to occur. Additional follow-on experiments should include measurement of the respiratory quotient by indirect calorimetry (Even and Nadkarni, 2012) to determine which substrates (fats and/or carbohydrates) are used during feeding and fasting, with or without S-BMO supplementation. Metabolic inflexibility is a feature usually associated with overnutrition (metabolic syndrome) and is thought to be a result of high concentrations of both lipids and carbohydrates that lead to a regulatory ‘roadblock’ (Muoio 2014). The relationship, therefore, between metabolic inflexibility and undernutrition is unclear. One testable hypothesis is that the aberrant metabolic response of unsupplemented animals is the result of an inability to adequately store dietary nutrients as fats during feeding and subsequently no capacity to release stored fats upon fasting. A key regulator of lipid storage is malonyl-CoA, which serves as a precursor for de novo lipogenesis (McGarry 2002). Malonyl-CoA also acts as an allosteric inhibitor of
carnitine palmitoyltransferase-1 (CPT-1), a key enzyme for mitochondrial beta-oxidation of fatty acids (McGarry 2002). In the follow-on experiments described above, an observed increase in malonyl-CoA in liver upon feeding in S-BMO supplemented animals (metabolically flexible) but not in unsupplemented controls (metabolically inflexible), would support this hypothesis.

**Identifying milk oligosaccharide structures that promote growth in gnotobiotic animal models**

The effects of dietary supplementation with a mixture of sialylated milk oligosaccharides on growth and metabolism represent a promising avenue of research for the development of microbiota-directed therapeutics. However, the molecular mechanisms by which these milk-derived compounds act remain unclear. A critical path forward is to identify specific components of S-BMO that are necessary and sufficient for mediating effects on growth and metabolism in gnotobiotic mouse and piglet models. Identification of these bioactive structures would inform additional studies of the microbial and host factors that underlie oligosaccharide-mediated growth as well as the sourcing of oligosaccharides for therapeutic application.

The S-BMO preparation used for the studies reported in Chapter 3 is composed primarily of sialyllactose, which represents 88% of its oligosaccharide content. Sialyllactose, a triose consisting of glucose, galactose, and sialic acid monomers is present in two distinct configurations, defined by either an α-2,3 or an α-2,6 glycosidic bond between the sialic acid and galactose moieties [termed 3’SIL and 6’SIL, respectively (Bruggencate et al, 2014)]. The abundance of these structures in S-BMO, along with their commercial availability in kilogram quantities, makes them prime candidates for follow-on studies, conducted in gnotobiotic animal models, designed to identify bioactive milk oligosaccharide structures.

Testing the biological activity of 3’SIL and 6’SIL would follow an experimental design similar to the studies detailed in Chapter 3, wherein gnotobiotic mice are colonized with the 25-member defined consortium of bacteria isolated from a stunted Malawian infant and fed a representative Malawian diet with or without addition of 3’SIL and/or 6’SIL. One could then moni-
tor growth and metabolic phenotypes in recipient animals, including body weight gain, lean body mass, bone morphology, and metabolic flexibility, as a function of 3’SL or 6’SL supplementation. Furthermore, one could examine the effect of these oligosaccharides on gut microbiota structure (COPRO-Seq analysis of serially collected fecal samples to quantify the proportional representation of members of the defined community as a function of treatment), on bacterial gene expression (microbial RNA-Seq analysis of fecal samples as well as samples collected along the length of the gut at the time of sacrifice), and on levels of various metabolites, including sialic acid, in the gut, liver, brain, muscle, and serum (as defined by targeted mass spectrometry). Importantly, the proposed studies would reveal whether there was structural specificity for the effects of sialyllactose on growth and metabolic phenotypes. Lastly, oligosaccharide structures that effectively recapitulate S-BMO mediated growth in gnotobiotic mice could be evaluated in the context of gnotobiotic piglets, which bear greater anatomic and physiologic similarity to humans (Kararli, 1995; Miller and Ullrey, 1987).

The dose of oligosaccharides to be administered is an important consideration. It is logical to initially select a dose that recapitulates the concentration of sialyllactose present in S-BMO supplemented diets, but additional studies could incorporate 3’SL and/or 6’SL at higher levels to determine whether larger effect sizes can be achieved and whether such effects saturate. These studies are important components of a pipeline that tests milk oligosaccharide supplements as candidate therapeutic agents. Determining optimal dosing is critical to define not only efficacy and safety, but also for determining the economic feasibility of this approach.

**Dependence of milk oligosaccharide-mediated growth on members of the infant gut microbiota**

S-BMO promotion of growth did not occur in germ-free mice. In addition, S-BMO had no impact on growth in gnotobiotic mice harboring a two-member community composed of *Bacteroides fragilis* and *Escherichia coli*, though both of these organisms exhibited transcriptional responses to S-BMO supplementation. These observations indicate that S-BMO mediated growth is micro-
biota-dependent and not simply attributable to bacterial colonization. Gnotobiotic mice provide a highly manipulable system for determining which human gut bacterial species/strains are essential for mediating the effects of S-BMO on growth and metabolism. Once an intact culture collection from an undernourished donor is shown to be able to transmit growth faltering phenotypes to gnotobiotic mice and to support S-BMO amelioration of these phenotypes, follow-up experiments can be performed in which animals are colonized with different subsets of this collection (Faith et al, 2014). S-BMO associated growth was documented in the context of a 25-member Malawian infant bacterial community; thus one could interrogate the requirement of various members of this community by selective removal of species/strains. Infant-associated *Bifidobacteria* species, for example, are prodigious consumers of milk oligosaccharides, including sialyllactose, and constitute five strains present in the 25-member community (Pacheco et al, 2015). Members of this genus have also been documented to modify gene expression in colonic epithelial cells after growth on milk oligosaccharides (Chichlowski et al, 2012). The hypothesis that *Bifidobacteria* strains are important effectors of S-BMO mediated growth can be tested by colonizing young gnotobiotic mice with the defined community minus *Bifidobacteria* and documenting the effect of dietary S-BMO supplementation on growth and metabolic phenotypes. If no effect is observed in these animals, one could conclude (with appropriate positive controls) that these *Bifidobacteria* strains are required for S-BMO mediated growth. Furthermore, applying a similar strategy in additional follow-up experiments could identify which *Bifidobacteria* isolates are necessary to mediate S-BMO associated growth. This would be accomplished by their individual removal from the 25-member community. These studies could then be performed in gnotobiotic piglets to test the generalizability (and translatability) of the findings in mice.

This experimental paradigm could be extended to other members of the 25-member community, including potentially pathogenic strains (e.g., *Enterococcus faecalis*) whose representation and/or expressed functions could be enhanced or attenuated by sialylated oligosaccharides. Such experiments could be conducted in a targeted, hypothesis-driven manner as above or by ‘random’ subsetting of the bacterial community [since the combinatorial possibilities are great, the size of
the subset is an important parameter to consider, as is the need to have overlapping membership between the subsets to reduce the effort in identifying which microbes or groups of microbes are required to mediate S-BMO effects (Faith et al, 2014)].

Identification of the microbial determinants of S-BMO mediated growth would further enable the determination of required bacterial genes [e.g., using random mutagenesis techniques like INSeq (Goodman et al, 2009)] and give a greater mechanistic understanding of the role of milk oligosaccharides in promoting infant growth and nutrition.

The results of the studies detailed above are extraordinarily important with respect to the therapeutic potential of milk oligosaccharide supplementation. The structure of the gut microbiota varies substantially across geographically distinct populations and between individuals within a given population (Yatsunenko et al, 2012), and the gut microbiota structure of undernourished infants is distinct (more immature) from that of their healthy counterparts (Chapter 2). Thus, the representation of taxa that are critical for S-BMO mediated effects may vary between infants, and infants lacking those strains may be unresponsive (or more responsive) to treatment. The ability to screen children for the presence of required strains in their gut microbiota could guide recruitment for clinical studies and/or development of treatments that incorporate the administration of required bacterial species.

**Challenges facing milk oligosaccharides as novel therapeutics for the treatments of childhood undernutrition**

S-BMO and similar preparations of milk oligosaccharides are attractive candidates for next generation, microbiota-directed therapeutics for the prevention and/or treatment of childhood undernutrition. Key challenges must be addressed, however, to make milk oligosaccharide preparations a viable form of treatment.

First, scalable sources of milk oligosaccharides must be identified. Milk oligosaccharides occur naturally in bovine milk at low levels and can be purified from existing whey waste streams.
This approach may carry the benefit of the resulting products being viewed as foods, rather than as drugs, both by target populations and regulatory agencies. An important consideration, however, is that the sourcing and purification of milk oligosaccharide preparations must be economically feasible (profitable for dairy farmers). The economic viability of producing milk oligosaccharide preparations will likely depend on many factors, including the required purity of the mixtures produced. Enzymatic removal of lactose, for example, could be cost prohibitive, but inclusion of lactose in oligosaccharide mixtures may not preclude the biological activity of other milk oligosaccharides that are present. Alternatively, the identification of individual bioactive milk oligosaccharide structures, as described above, may allow a synthetic approach (by chemical engineering or through production in engineered microbes), wherein one or more structures can be synthesized at scale for the purpose of treating childhood undernutrition/stunting. A key advantage of the latter approach is that it would allow tight control of product quality and clear characterization of dosing and safety.

Secondly, the effectiveness and safety of oligosaccharide supplements for the treatment of childhood undernutrition/stunting must be rigorously established through randomized, controlled clinical trials. Preclinical studies, of the type described in this chapter, will help to elucidate the microbial dependencies of oligosaccharide-mediated growth, determine the generalizability of the effects (by testing S-BMO against microbiota from a number of children exhibiting varying degrees of linear growth defects from various regions of a given country or countries), and inform the design of clinical studies. For example, study cohorts could be designed to consider only those infants harboring species required for processing of milk oligosaccharides. Another approach might involve co-administration of required bacterial species with milk oligosaccharide supplements to encourage effectiveness of treatment.

An important role for preclinical studies in gnotobiotic animal models is to help define risk associated with various pathogenic members of the gut microbiota. Previous reports have documented expansion of enteric pathogens in mouse models accompanying the liberation of sialic acid from host glycans (Ng et al, 2013). This is consistent with the crossfeeding behavior observed
between *B. fragilis* and *E. coli*, reported in Chapter 3. Sialylated milk oligosaccharide supplementation may, therefore, carry substantial risk in the context of high enteropathogen burden. This information should guide de-risking strategies for the design of clinical trials using milk oligosaccharide supplements as well as avoiding treatment of infants with high enteropathogen burdens.

In conclusion, purified preparations of milk oligosaccharides are attractive candidates for next generation, microbiota-directed therapeutics for the treatment of childhood stunting/under-nutrition, and preclinical studies of the type described above will enable better understanding of their mechanisms of action and inform design of clinical trials to establish their safety and efficacy.
References


Appendices

Appendix A

The Long-Term Stability of the Human Gut Microbiota

Jeremiah J. Faith, Janaki L. Guruge, Mark Charbonneau, Sathish Subramanian, Henning Seedorf, Andrew L. Goodman, Jose C. Clemente, Rob Knight, Andrew C. Heath, Rudolph L. Leibel, Michael Rosenbaum, Jeffrey I. Gordon*

Introduction: Understanding the dynamics and stability of the human gut microbiota is important if its characterization is to play a role in the diagnosis, treatment, and prevention of disease. To characterize stability in related and unrelated individuals and its responsiveness to physiologic change (weight loss), we developed a method for bacterial 16S rRNA amplicon sequencing at high depth with high precision. We also sequenced the genomes of anaerobic bacteria represented in culture collections prepared from fecal samples collected from individuals over time.

Methods: Low-error amplicon sequencing (LEA-Seq) is a quantitative method based on redundant sequencing of bacterial 16S rRNA genes. A dilute, barcoded, oligonucleotide primer solution is used to create ~150,000 linear PCR extensions of the template DNA. The labeled, bottlenecked linear PCR pool is amplified with exponential PCR, using primers that specifically amplify only the linear PCR molecules. The exponential PCR pool is sequenced at sufficient depth to obtain ~20× coverage. Multiple reads enable the generation of an error-corrected consensus sequence for each barcoded template molecule. LEA-Seq can be used for a variety of other applications.

Results and Discussion: LEA-Seq of fecal samples from 37 healthy U.S. adults sampled 2 to 13 times up to 296 weeks apart revealed that they harbored 195 ± 48 bacterial strains, representing 101 ± 27 species. On average, their individual microbiota was remarkably stable, with 60% of strains remaining over the course of 5 years. Stability followed a power law, which, when extrapolated, suggests that most strains in an individual’s intestine are residents for decades (figure, panel A). Members of Bacteroidetes and Actinobacteria are significantly more stable components than the population average. LEA-Seq of four individuals sampled during an 8- to 32-week period during a calorie-restricted dietary study showed that weight stability is a significantly better predictor of microbiota stability than the time interval between samples (figure, panel B). After generating clonally arrayed collections of anaerobic bacteria from frozen fecal samples collected from six weight-stable individuals sampled 7 to 69 weeks apart, we produced draft genome sequences for 534 isolates representing 188 strains and 75 species. A targeted approach focused on Methanobrevibacter smithii isolates from two sets of twin pairs and their mothers and Bacteroides thetaiotaomicron strains from nine donors including sister-sister and mother-daughter pairs. Strains, defined as isolates sharing >96% of their genome content, were maintained over time within an individual and between family members but not between unrelated individuals. Thus, early gut colonizers, such as those acquired from our parents and siblings, have the potential to exert their physiologic, metabolic, and immunologic effects for most, and perhaps all, of our lives.

A 0.4 0.5 0.6 0.7 0.8 0.9 1.0

Time between samples (years)

Jaccard index

B

0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00

∆lnBMI between samples

∆time between samples (weeks)

The list of author affiliations is available in the full article online.

*Corresponding author. E-mail: jgordon@wustl.edu

FIGURES AND TABLE IN THE FULL ARTICLE

Fig. 1. Multiplex bacterial 16S rRNA gene sequencing using LEA-Seq; comparison with previous methods using mock communities composed of sequenced gut bacterial species.

Fig. 2. Measuring the stability of an individual’s fecal microbiota over time with LEA-Seq.

Fig. 3. Relationship among weight stability, time, and fecal microbiota stability.

Fig. 4. Comparison of genome stability in fecal bacterial isolates recovered from individuals over time.

Table 1. Species composition of the sequenced arrayed culture collections from six donors.

SUPPLEMENTARY MATERIALS

Materials and Methods

Fig. S1 to S8
Table S1 to S15
References

Relationship among time, physiology, and microbiota stability. (A) Stability of fecal microbiota follows a power-law function (n = 37 females sampled over time; <1 week to >5 years). Dashed lines show 95% confidence bounds over 10- and 50-year extrapolations (inset). (B) Microbiota stability is inversely related to the stability of each individual’s body mass index.
The Long-Term Stability of the Human Gut Microbiota

Jeremiah J. Faith,† Janaki L. Guruge,§ Mark Charbonneau,¶ Sathish Subramanian,¶ Henning Seedorf,¶ Andrew L. Goodman,¶ Jose C. Clemente,§ Rob Knight,§*,‡ Andrew C. Heath,§ Rudolph L. Leibel,§ Michael Rosenbaum,‡ Jeffrey I. Gordon‡†

A low-error 16S ribosomal RNA amplicon sequencing method, in combination with whole-genome sequencing of >500 cultured isolates, was used to characterize bacterial strain composition in the fecal microbiota of 37 U.S. adults sampled for up to 5 years. Microbiota stability followed a power-law function, which when extrapolated suggests that most strains in an individual are residents for decades. Shared strains were recovered from family members but not from unrelated individuals. Sampling of individuals who consumed a monotonous liquid diet for up to 32 weeks indicated that changes in strain composition were better predicted by changes in weight than by differences in sampling interval. This combination of stability and responsiveness to physiologic change confirms the potential of the gut microbiota as a diagnostic tool and therapeutic target.

Our growing understanding of the human gut microbiota as an indicator of and contributor to human health suggests that it will play important roles in the diagnosis, treatment, and ultimately prevention of human disease. These applications require an understanding of the dynamics and stability of the microbiota over the life span of an individual. Amplicon sequencing of the bacterial 16S rRNA gene from fecal microbial communities (microbiota) has revealed that each individual harbors a unique collection of species (1–3). Estimates of the number of species present in an individual’s microbiota have varied greatly. Culture-based techniques (4) indicate ~100 such species, whereas culture-independent deep shotgun sequencing of fecal community DNA (5) indicates ~160 such species. Several times these numbers of species are suggested by the results of 16S rRNA amplicon sequencing, even after in silico attempts to remove chimeric molecules formed in the course of a polymerase chain reaction (PCR) and errors introduced during sequencing (2). These artifacts complicate tracking of individual bacterial taxa across time by inflating the set of strains in each sample with false positives. Shotgun sequencing of the community’s microbiome is another approach for defining diversity (6), but it is difficult to associate gene sequences with their genome of origin.

With these limitations in mind, we have developed a method for amplicon sequencing to assay the bacterial composition of the gut microbiota of individuals at high depth with high precision over time. When combined with high-throughput methods for culturing and sequencing the genomes of anaerobic bacteria, these results reveal that the majority of the bacterial strains in an individual’s microbiota persist for years, and suggest that our gut colonizers have the potential to shape many aspects of our biological features for most and in some cases all of our lives.

A Method for Low-Error Amplicon Sequencing (LEA-Seq) of Bacterial 16S rRNA genes

A 16S rRNA sequencing method for assaying the stability of an individual’s microbiota over time would ideally retain high precision at high sequencing depth [precision = (true positives)/(true positives + false positives)]. Low-precision data complicate comparison of sequences between samples, as it becomes difficult to differentiate species (typically defined as isolates that share ≥97% sequence identity in their 16S rRNA genes) and strains (isolates of a given species with more minor variations in their 16S rRNA gene sequences) from sequencing errors. Standard amplicon sequencing is limited in its precision by the overall error rate of the sequencing method. If sequencing depth is low, it becomes impossible to determine whether a strain has dropped out of a given individual’s microbiota or has fallen below the limits of detection at the sampling depth used.

In many applications it would be advantageous to exchange sequencing depth for improved sequence quality. Despite several optimizations we developed to increase the precision of standard amplicon sequencing at shallow depths, we found that sequencing a sample beyond 10,000 reads did not substantially increase the lower detection limit possible at high precision (7). Exchanging sequence quantity for sequence quality is inherent in shotgun genome sequencing, where redundant sequencing of genomes at 10× to 50× coverage enables a far lower error rate than is attainable from single reads alone. In general, to redundantly sequence DNA fragments, it is necessary to create a finite DNA pool that is smaller than the amount of sequencing available (i.e., create a bottleneck) and to have a method of labeling the molecules in the pool (8–10). To adapt these techniques to redundantly sequence PCR amplicons, the initial template DNA could be diluted to create a bottleneck. However, this dilution would likely need to be empirically determined for every input sample (e.g., using quantitative PCR), and one would still need to label each template molecule. As an alternative, we developed a method called low-error amplicon sequencing (LEA-Seq).

As outlined in Fig. 1A, LEA-Seq is based on redundant sequencing of a set of linear PCR template extensions of 16S rRNA genes to trade sequence quantity for quality. In this method, we create the bottleneck with a linear PCR extension of the template DNA with a dilute, barcoded, oligonucleotide primer solution. Each oligonucleotide is labeled with a random barcode positioned 5′ to the universal 16S rRNA primer sequence (Fig. 1A and fig. S1). We then amplify the labeled, bottlenecked linear PCR pool with exponential PCR, using primers that specifically amplify only the linear PCR molecules. During the exponential PCR, an index primer is added to the amplicons with a third primer to allow pooling of multiple samples in the same sequencing run (fig. S1). This exponential PCR pool is then sequenced at sufficient depth to redundantly sequence (~20× coverage) the bottlenecked linear amplicons. The resulting sequences are separated by sample, using the index sequence, and the amplicon sequences within each sample are separated by the unique barcode; the multiple reads for each barcode allow the generation of an error-corrected consensus sequence for the initial template molecule. In LEA-Seq, the linear PCR primers are diluted to a concentration that generates ~150,000 amplicon reads at 20× coverage per amplicon on an Illumina HiSeq DNA sequencer.

To empirically test LEA-Seq against existing 16S rRNA amplicon sequencing methods, we first generated nine in vitro “mock” communities composed of different proportions of strains from a 48-member collection of phylogenetically diverse, cultured human gut bacteria whose genomes had been characterized (7) (table S1). To
calculate precision, we compared amplicons generated using two sequencing platforms (Illumina MiSeq and 454 FLX instruments), targeting different variable (V) regions of the 16S rRNA gene with different PCR primers. We defined a true positive sequence as 100% identical across 100% of its length to the 16S rRNA gene sequence(s) in the reference genome. We calculated precision at different abundance thresholds by including only those sequences representing at least a minimal portion of the total sequencing reads (0.5%, 0.1%, 0.05%, 0.01%, or 0.005%).

Relative to the existing standard approaches, LEA-Seq produced amplicon sequences with higher precision at lower abundance thresholds (Fig. 1B). For 16S rRNA sequences representing ≥ 0.01% of the reads, LEA-Seq enabled a precision of 0.83 ± 0.02 (V4) and 0.63 ± 0.03 (V1V2) versus 0.08 ± 0.064 and 0.09 ± 0.005, respectively, for the same regions with standard amplicon sequencing (table S2). These performance improvements are dependent on generating the consensus sequence from the redundant amplicon reads (table S2, method “LEA-Seq without consensus”). LEA-Seq also produced slower saturation in performance (precision of ≥ 0.7 for reads representing 0.001% of the total; fig. S2 and table S2). Similar results were obtained using the several different mock communities [see (7) for additional details of the analysis, including V1V2 versus V4 comparisons]. On the basis of this assessment of its attributes, we used LEA-Seq to quantify the stability of the gut microbiota within individuals as a function of time and change in body mass index (BMI) while they consumed controlled monotonous or free diets.

### Applying LEA-Seq to Define the Stability of the Fecal Microbiota of 37 Healthy Adults

**Stability of a Microbiota Best Fits a Power-Law Function**

We used LEA-Seq to characterize the microbiota in 175 fecal samples obtained from 37 healthy adults residing throughout the United States; 33 of these donors were sampled 2 to 13 times up to 296 weeks apart (1, 11) (table S3). The remaining four individuals were sampled on average every 16 days for up to 32 weeks while consuming a monotonous liquid diet as part of a controlled in-patient weight loss study (see methods) (12–14). None of the individuals took antibiotics for at least 2 months before sampling. All fecal samples were cooled to −20°C immediately after they were produced and then to −80°C within 24 hours. DNA was isolated from all samples by bead beating in phenol and chloroform.

Using an Illumina HiSeq 2000 instrument to sequence amplicons from the V1V2 region of bacterial 16S rRNA genes, we generated 108,677 (mean ± SD) LEA-Seq reads per fecal DNA sample. Reads were then filtered using a minimum sequence abundance threshold cutoff of eight reads (i.e., to detect strains present in the fecal microbiota at an average relative abundance of 0.007%). On the basis of our mock community data, the precision at this threshold for the V1V2 region is 0.63. We defined the number of strains in a sample as the number of unique amplicon sequences, and the number of species-level operational taxonomic units (OTUs) in the sample as the number of clusters with 97% shared sequence identity. To correct for false positives, we multiplied the number of strains by the precision (i.e., if we detect 100 unique sequences, we expect 63 of them to be true). For individuals sampled over multiple time points, we calculated the number of species and strains for each sample individually and averaged them. The results indicated that individuals in this cohort harbored 195 ± 48 bacterial strains in their fecal gut microbiota, representing 101 ± 27 species.

To study each individual’s microbiota over time, we took all possible pairs of samples from the time series of each individual (table S3) and calculated the time in weeks between the sample dates as well as the fraction of shared strains between them, as measured by the binary Jaccard index (an unweighted metric of community overlap):

\[
\text{Jaccard index(sample A, sample B)} = \frac{\text{sample A} \cap \text{sample B}}{\text{sample A} \cup \text{sample B}}
\]

Control experiments using mock communities (table S1) established that LEA-Seq of V1V2 16S rRNA amplicons produced highly accurate estimates of the Jaccard index [correlation between known and measured Jaccard index = 0.996; see (7)]. To characterize the stability of an individual’s microbiota, we binned fecal samples into intervals (<3 weeks, 3 to 6 weeks, 6 to

---

**Fig. 1.** Multiplex bacterial 16S rRNA gene sequencing using LEA-Seq; comparison with previous methods using mock communities composed of sequenced gut bacterial species. (A) Schematic of how the LEA-Seq method is used to redundantly sequence PCR amplicons from a set of linear PCR template extensions of bacterial 16S rDNA. This approach results in amplicon sequences with a higher precision than standard amplicon sequencing at lower abundance thresholds. (B) Performance of 16S RNA amplicon sequencing methods assayed as the precision obtained for different sequence abundance thresholds. Standard methods for amplicon sequencing using the 454 pyrosequencer and the Illumina MiSeq instrument exhibit increased precision as less abundant reads are filtered out. By redundantly sequencing each amplicon with LEA-Seq, the precision of amplicon sequencing is increased at lower abundance thresholds for both the V1V2 region of the bacterial 16S rRNA gene (compare red and green lines) and the V4 region (compare magenta and blue lines), thereby enabling detection of lower-abundance bacterial taxa at higher precision.

---

1237439-2

5 JULY 2013 VOL 341 SCIENCE www.sciencemag.org
9 weeks, 9 to 12 weeks, 12 to 32 weeks, 32 to 52 weeks, 52 to 104 weeks, 104 to 156 weeks, 156 to 208 weeks, 208 to 260 weeks, and >260 weeks apart (mean ± SE for each bin is shown). The decay in the Jaccard index as a function of time between two samples best fits a power law (blue line). (B) Four individuals losing 10% of their body weight in the study involving consumption of a monotonous low-calorie liquid diet (magenta) had significantly less stable microbiota than the mean of the 33 remaining individuals (blue). Means ± SE for the Jaccard index are plotted. (C) At the phylum level, Bacteroidetes (blue) and Actinobacteria (red) were more stable components of the microbiota than Proteobacteria and Firmicutes (hypergeometric distribution).

To define the stability of a given strain as a function of its relative abundance in the microbiota, we used all pairwise combinations of fecal samples obtained from each individual to calculate (i) the mean abundance of the strains shared by two or more samples, and (ii) the mean abundance of strains that were not shared between any two samples. Strains that were shared across two time points were roughly 3 times as abundant as those that were not shared [0.030 ± 0.013 fraction of the community versus 0.011 ± 0.004 (mean ± SD); \( P = 2.2 \times 10^{-9} \) (t test); fig. S3A]. We also binned the strain abundances for each donor using five fractional abundance thresholds of 0.1, 0.01, 0.001, 0.0001, and <0.0001 (e.g., bin 0.01 contains all strains ≤0.1 and >0.01) and calculated the probability that strains in a given bin were shared between samples. We found that the higher the fractional abundance of a strain, the more likely the strain was shared between samples (\( r = 0.96, P < 0.0087; \) fig. S3B). Together, these results suggest that the more stable components of the microbiota are also the most abundant members.

**Effects of a Monotonous Low-Calorie Diet and Associated Weight Loss on Diversity**

To explore the role of weight loss on the microbiota, we applied LEA-Seq to the fecal microbiota of four individuals sampled over the course of an 8- to 32-week period in a three-phase study that used different caloric intakes of a defined monotonous liquid diet to first stabilize initial weight, then decrease weight by 10%, and finally maintain weight at the 10% reduced level (Fig. 2B and table S3). Daily caloric intake was 2988 ± 290 kcal, 800 kcal, and 2313 ± 333 kcal for the three phases of the study, respectively (14, 15). While on this diet, these four individuals experienced significantly reduced stability of their microbiota, as measured by the Jaccard index (Fig. 2B). For each individual, we found no significant correlation between time and diversity/richness (i.e., number of strains in a sample; minimum \( P = 0.17 \)). Additionally, we found no significant correlation between the change in composition of the microbiota (Jaccard index between two samples) and the change in diversity/richness (absolute difference in the number of species or strains between two samples) \( (P = 0.09 \) and 0.44 for strains and species, respectively). Considering family-level taxonomic bins, there were several groups whose abundance was strongly positively or negatively correlated with time during the weight loss period, including Clostridiaceae \( (\) average correlation \( r = 0.60, P < 0.0001 \) across donors during weight loss = 0.60), Coriobacteriaceae \( (r = 0.53), Bifidobacteriaceae \( (r = 0.55), Enterobacteriaceae \( (r = 0.58), Lachnospiraceae \( (r = -0.65), Oscillospiraceae \( (r = -0.53), \) and Oxalobacteraceae \( (r = -0.74).\)

**Modeling the Relationship Among Time, Body Composition, and Microbiota Stability**

Given the correlation between weight loss and changes in the microbiota of individuals consuming a monotonous 800 kcal/day diet, we took a broader view across all 37 individuals in our study to determine whether this correlation was due to the monotonous diet that the four individuals had consumed, or if there is a generalizable and quantifiable relationship between weight stability and microbiota stability. To explore this question, we not only calculated the time \( \Delta \text{time} \) and Jaccard index between all pairs of fecal samples collected from an individual (Fig. 2), but also the absolute value of the change in natural logarithm of the BMI value (abbreviated \( \Delta \text{BMI} \)) between all pairs. We found a significant negative correlation between \( \Delta \text{BMI} \) and Jaccard index \( (\text{Fig. 3A}; r = -0.68, P = 2.98 \times 10^{-17}) \) that was even greater than that between \( \Delta \text{time} \) and Jaccard index \( (\text{Fig. 3B}; r = -0.42, P = 1.45 \times 10^{-12}).\) These relationships held when we removed the data generated from the four individuals on the monotonous diet \( (\Delta \text{BMI}; r = -0.69, P = 3.27 \times 10^{-34}; \Delta \text{time}; r = -0.65, P = 9.05 \times 10^{-86})).\)

To quantify the relationship among \( \Delta \text{time}, \Delta \text{BMI}, \) and the Jaccard index between pairs of samples (Fig. 3C), we fit the following model:

\[
\text{Microbiota stability} = \beta_0 + \beta_{\Delta \text{BMI}} \times \Delta \text{BMI} + \beta_{\Delta \text{time}} \times \Delta \text{time}
\]
where microbiota stability is the Jaccard index between samples, $\Delta \ln \text{BMI}$ is the change in $\ln \text{BMI}$ between any two samples collected from the individual (time, $\beta_\text{lnBMI}$ and $\beta_\text{time}$ are the linear regression estimated parameters for $\Delta \ln \text{BMI}$ and $\Delta \text{time}$, respectively. Remarkably, this model explained 46% of the variance in the stability of the microbiota ($R^2 = 0.46, P = 2.82 \times 10^{-7}$). Together, these relationships among time, BMI, and the stability of an individual’s microbiota highlight the role that longitudinal surveys of a microbiota could play in health diagnostics.

Sequenced Collections of Fecal Bacteria Obtained from Individuals over Time

As in previous studies (1, 16–19), we found that each individual’s microbiota at a given time point was most similar to their own at other time points (Jaccard index, 0.82 $\pm$ 0.02), followed by their family members (Jaccard index, 0.38 $\pm$ 0.02), and then unrelated individuals (Jaccard index, 0.30 $\pm$ 0.005). The Jaccard index estimates with LEA-Seq suggest that on average any two unrelated individuals share ~30% of the strains in their microbiota. However, it is possible that unrelated individuals on average share no strains in their microbiota, and that this 30% represents the lower resolving limit of 16S rRNA amplicon sequencing of the targeted variable region (V1V2) and currently available maximum read lengths on the Illumina HiSeq 2000 instrument (paired-end, 101 base pairs).

Whole-genome alignments between bacteria isolated and sequenced from different samples provide many orders of magnitude of additional resolving power to determine which strains (now defined at the level of whole-genome sequence identity rather than 16S rRNA identity) remain in an individual’s microbiota over time, or reside in two unrelated individuals. Isolation and sequencing of extensive collections of organisms from the human gut microbiota (20) can provide a practical method to look at the plasticity and evolution of the gene content of microbial
strains harbored in individuals’ intestines over time. Therefore, adapting a high-throughput method we had developed for generating clonally arrayed collections of anaerobic bacteria in multwell format from frozen fecal samples (20), we produced draft genome sequences for 444 bacterial isolates recovered from the frozen fecal microbiota of five donors who had been sampled across periods from 7 to 69 weeks apart (n = 1 to 4 time points per donor; 11 total samples; mean coverage per microbial genome = 118×; see tables S5 and S6) (7). These genomes span a broad phylogenetic range within the four dominant bacterial phyla that constitute the human gut microbiota (Bacteroidetes, Firmicutes, Proteobacteria, and Actinobacteria; table S6).

To look for changes in bacterial genome content across time in each individual, we performed whole-genome alignment with nucmer (21) and calculated the fraction of DNA sequence aligned between each pair of genomes [alignment score = (X_{ab} + Y_{ab})(X + Y), where X and Y are the lengths of genome X and Y, respectively, and X_{ab} and Y_{ab} are the number of aligned bases of genomes X and Y, respectively] (7, 22). We found that the shared genome content between isolates from unrelated individuals was broadly distributed for taxa from the same genus (coverage score = 0.30 ± 0.20) or species (0.77 ± 0.12), with a maximum of 0.956 (Fig. 4A, blue; fig. S4). We then compared the shared genome content between isolates within each fecal sample (i.e., self-versus-self at a single time point) and found isolates that shared a very high proportion of their content (0.965 to 0.999) (Fig. 4A, red). Remarkably, we found the same high proportion of shared genome content between isolates from a given donor between different time points (i.e., self-versus-self over time; Fig. 4A, green), which suggests that the same strains of bacteria persisted in these individuals over the course of the sampling period.

Defining replicative bacterial strains as those with a coverage score of >0.96 and species as those with a coverage score of >0.5 (fig. S4), we subsequently clustered the genome isolates by sample and by individual (table S5); this effort yielded a total of 165 strains and 69 species across the five donors (Table 1). Across the four donors with multiple time points, on average 36% of an individual’s bacterial strains were isolated from multiple time points. This fraction of shared bacterial strains across time at the level of the genome is lower than that measured by LEA-Seq; however, this likely reflects the increased sampling depth and culture independence of LEA-Seq [detecting isolates at depths of 1:10,000 to 1:100,000 (0.01 to 0.001%) compared with 0.14 to 0.06% for high-throughput culturing]. For the most deeply sampled individual (F3T1 in table S3), where isolates were sequenced from four samples taken over the course of ~16 months, more than 60% of the strains were isolated from multiple samples.

Stability Viewed from the Perspective of Phylum-Level Membership

When we assigned phylum-level taxonomy to all LEA-Seq 16S rRNA amplicons from each of the 37 individuals in our study (23), we found that members of the Bacteroidetes and Actinobacteria were significantly more stable components of the microbiota relative to the population average (hypergeometric distribution comparing the total number of shared/not shared strains within a given phylum for all samples versus the total number of shared/not shared strains across all phyla, except the phylum of interest; P = 7.54 × 10^{-28} and 0.0068, respectively), whereas the Firmicutes and Proteobacteria were significantly less stable (Fig. 2C; P = 1.83 × 10^{-11} and 0.0015, respectively). The cultured bacterial strains manifested similar trends for the Bacteroidetes and Firmicutes, where 52% and 21%, respectively, of the strains were isolated and sequenced across multiple time points (table S7), thus demonstrating at a whole-genome level the strain...
stability initially identified when only the 16S rRNA gene was targeted for analysis.

Strains Shared Between Members of Human Families

The power-law response of the Jaccard index as a function of the sample collection interval makes it possible to extrapolate beyond the sampling time frame of the current study and suggests that the majority of strains in the microbiota represent a stable core that persists in an individual’s intestine for his or her entire adult life, and could represent strains acquired during childhood from parents or siblings (fig. S5). Therefore, we used LEA-Seq to measure the fraction of shared strains between family members (sister-sister or mother-daughter). As in previous studies (1), we found the microbiota of related individuals was more similar than unrelated ones, with a significantly larger proportion of shared V1V2 16S rRNA sequences [Jaccard index = 0.38 ± 0.020 (related), 0.30 ± 0.005 (unrelated); P = 0.00053].

To determine whether this increased similarity between family members manifested itself at the level of their gut microbial genome sequences, we used a targeted approach to look at genome content differences in (i) two families using previously sequenced Methanobrevibacter smithii isolates (24) from two sets of twin pairs and their mothers (six total donors, 19 genomes; table S3), and (ii) five families where 26 Bacteroides thetaiotaomicron strains were isolated with a species-specific monoclonal antibody (7, 25) from nine donors including sister-sister and mother-daughter pairs (all isolates were from a single sample from each donor; table S3). M. smithii, a methanogen, is the dominant archaeon in the human gut microbiota and facilitates fermentation of polysaccharides by saccharolytic bacteria such as B. thetaiotaomicron by virtue of its ability to remove hydrogen (24). As with our untargeted large-scale genome sequencing of personal bacterial culture collections described above, we found that unrelated individuals had no pair of isolates of either species that shared >96% of their genome content. However, within an individual we once again found replicate isolates of the same strain (Fig. 4, B and C, blue and red). Strikingly, we also found replicate strains of M. smithii or B. thetaiotaomicron shared across family members (Fig. 4, B and C, brown; table S3).

In contrast with the results obtained using this taxon-targeted whole-genome sequencing approach, our untargeted sequencing of the clonally arrayed personal bacterial culture collections had only involved two related individuals (female dizygotic co-twins 1 and 2 from family 60; F60T1 and F60T2, table S3) and had revealed no strains with >96% of their genomes aligned. Therefore, we isolated and sequenced an additional 89 genomes from two time points of the dizygotic twin sister (F61T1) of subject F61T1 (yielding a total of 188 strains and 75 species across the six donors). As with the previous do-

Table 1. Species composition of the sequenced arrayed culture collections from six donors. Alternative names for species are in parentheses.

<table>
<thead>
<tr>
<th>Species ID</th>
<th>Species</th>
<th>F3T1</th>
<th>F58T1</th>
<th>F60T1</th>
<th>F60T2</th>
<th>F61T1</th>
<th>F61T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alistipes indistinctus</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Anaerococcus vaginalis (A. hydrogenalis)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Anaerostatistis stercorihammonis</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Anaerostatistis stercorihammonis</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Bacteroides</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Bacteroides caccae</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Bacteroides finegoldii</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Bacteroides fragilis</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Bacteroides intestinalis (B. cellulosilyticus)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Bacteroides massiliensis</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Bacteroides ovatus</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Bacteroides ovatus</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Bacteroides thetaiotaomicron (B. faecis)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Bacteroides uniformis (B. acidifaciens)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Bacteroides vulgatus (B. dorei)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Barnesiella intestinihominis</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Bifidobacterium adolescentis</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Bifidobacterium bifiduni</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Bifidobacterium longum</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Bifidobacterium pseudocatenulatum</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Blautia</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Blautia schinkii</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Butyricimonas virosa</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Clostridiales</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Clostridiales</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Clostridiales</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Clostridiales</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Clostridiales</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Clostridiales</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Clostridiales</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Clostridium</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Clostridium</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Clostridium</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Clostridium bolteae</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Clostridium hylemonae</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>Clostridium leptum</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>Clostridium scindens</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>Clostridium scindens</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>Collinsella</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>Collinsella aerofaciens</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>Coprococcus comes</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>Dorea formicigeners</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>Dorea longicatena</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>Dorea longicatena</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>Eggerthella lenta (Subdoligranulum variabile)</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>Escherichia coli</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>Eubacterium biforme</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>Eubacterium callanderi</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>Eubacterium contortum</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>Eubacterium eligens</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>Finegoldia magna (Dialister invisus)</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>Lactobacillus</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>Lactobacillus casei</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>Megasphaera elsdonii</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>Odoribacter splanchnicus</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>Parabacteroides distasonis</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>Parabacteroides goldsteinii</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>Parabacteroides merdae</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>Peptostreptococcus harei</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>Roseburia intestinalis</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
nors, we were able to isolate numerous strains shared across the two time points (8 of 25 = 32%). In addition, we were able to isolate two strains (B. thetaiotaomicron and Escherichia coli) in both of the sisters, showing that even nontargeted genome isolation and sequencing can retrieve the same strain across family members. We did not explicitly sample members of our cohort of females during significant physiological transitions such as menarche and menopause. However, the presence of the same bacterial strain in mothers and their adult daughters who had progressed through one or both of these life-cycle milestones suggests that components of the microbiota are retained during these events.

Prospectus

The objects we touch and consume during the course of our lives are covered with diverse microbial life. Despite this, LEA-Seq revealed that on average 60% of the approximately 200 microbial strains harbored in each adult’s intestine were retained in their host over the course of a 5-year sampling period. Our results are supported by a microarray-based profiling of fecal microbiota collected from three males and two females over ~8 years but differ from those of a similar analysis using standard 16S rRNA amplicon sequencing, which found high variability in microbiota composition in two individuals sampled for up to 15 months (26). This difference likely reflects the fact that the sequencing depth and precision limitations of standard 16S rRNA amplicon sequencing are overcome to some extent with microarrays where amplicons are mapped and hybridized to a finite pool of target sequences (i.e., sacrificing resolution for precision). The differences could also be due to true differences in the stabilities in microbiota of the individuals, as both studies surveyed only a small number of individuals. Our findings are also supported by a recent report that mapped deep shotgun sequencing data sets of the fecal microbiome to a set of reference bacterial genomes (6) and found that the gut communities of these individuals were more similar to each other at the microbiome level than to unrelated individuals (average maximum time between samples = 32 weeks with two individuals sampled over a period of >1 year). Applying LEA-Seq to longitudinal surveys of the fecal microbiota of 37 twins sampled for up to 5 years allowed us to show that the stability of an individual’s microbiota follows a power-law function. Using this function, we could extrapolate the stability of the microbiota over decades. The resolution and accuracy of these predictions should improve as advances in sequencing chemistry enable longer reads of 16S rRNA genes to be characterized. LEA-Seq itself can be generalized to any application that requires deep amplicon sequencing with high precision (e.g., the VDJ regions of immunoglobulin and T cell receptor genes, or targeted searches for variants in candidate or known disease-producing genes).

Our study also illustrates how a highly personalized analysis of the gut community at strain-level microbial genome resolution can be conducted using collections of cultured bacteria (or archaea) generated from frozen fecal samples collected over time from a given subject. This strain-level analysis can be part of a broad phylogenetic survey, or it can target a particular species.

The stability of the microbiota that we document in healthy individuals has important implications for future use of the microbiota (and microbiome) as a diagnostic tool as well as a therapeutic target for individuals of various ages. Our findings suggest that obtaining a routine fecal sample as part of a yearly physical examination designed to promote disease prevention would be sufficient to monitor changes in the composition and stability of an individual’s fecal microbiota. For example, in the case of inflammatory bowel diseases, the concordance for Crohn’s disease and ulcerative colitis among monozygotic twin pairs is only 38% and 15%, respectively (27). Our results suggest that these twins likely share identifiable unique subsets of their microbiota that represent long-term environmental exposures for their immune systems that should be considered when trying to predict disease risk, or infer which species or strains may have a causal role in disease initiation, progression, relapse, and treatment responses. Moreover, the effects of travel, changes in diet, weight gain and loss, diarrhea disease, antibiotics, immunosuppressive therapy, or clinical trials designed to deliberately manipulate the microbiota (e.g., through administration of existing or new prebiotics, probiotics, symbiotics, antibiotics, or transplantation of microbiota from healthy individuals to those with various diseases attributed to a dysfunctional microbiota) can be more accurately quantified by applying the methods we describe. Finally, the stability we document highlights the impact of early colonization events on our microbiota in later life; earlier colonizers, such as those acquired from our parents and siblings, have the potential to provide their metabolic products and exert their immunologic effects for our entire lives.

Methods

Diet Studies

Four obese (BMI > 30 kg/m²) female subjects with a mean (± SD) age of 26 ± 3 years were admitted to the General Clinical Research Center at Columbia University Medical Center and remained as in-patients throughout the study. The protocol for recruitment and for the weight loss study was approved by the Institutional Review Board of New York Presbyterian Medical Center and is consistent with guiding principles for research involving humans. Written informed consent was obtained from all subjects. The diet protocol has been described in detail (12, 13). Briefly, subjects were fed a liquid-formula diet with 40% of energy as fat (corn oil), 45% as carbohydrate (glucose polymer), and 15% as protein (casein hydrolysate). Diet composition but not quantity was constant throughout the study. The diet had a caloric density of 1.25 digestible kcal of energy per gram and was supplemented with vitamins and minerals in quantities sufficient to maintain a stable weight, defined as an average daily weight variation of <10 g/day for ≥2 weeks. This weight plateau is designated as W_initial. The four individuals in this study consumed 2600 to 3300 kcal/day of the diet to maintain W_initial. After a brief period at W_initial, subjects were provided 800 kcal/day of the same liquid-formula diet until they had lost ~10% of W_initial. The duration of the weight loss phase ranged from 36 to 62 days (table S3). Once 10% weight loss had been achieved, intake was adjusted upward until subjects were again weight-stable. Weight maintenance calories were disproportionately reduced (−22%) below those required to
maintain initial weight and ranged from 2050 to 2800 kcal/day for the four individuals. Subject F72 also received triiodothyronine (25 μg/day) during this second weight-stable period (table S3). Fecal samples were obtained throughout the study (table S3) and frozen at –80°C until processed for DNA extraction (I).

**Twin Participants**

Twins were selected from a general population cohort of female like-sex twin pairs, born in Missouri to Missouri-resident parents between 1 July 1975 and 30 June 1985, and first assessed at median age 15 with multiple waves of follow-up (28, 29). Selected twins were drawn from (i) a study, which included biological mothers where available, contrastingly concordant lean twin pairs (both twins had BMIs in the range 18.5 to 24.9 by self-report at all completed assessments) and concordant obese twin pairs (both twins had BMIs ≥ 30, but with pairs prioritized where at least one twin had BMI > 35, to maximize separation from the concordant lean pairs) (I); (ii) a small-scale study of concordant lean monozygotic twin pairs contrasting free diet with free diet supplemented by twice-daily consumption of a fermented milk product (II); and (iii) an ongoing study of twin pairs selected for BMI discordance (either discordant lean/obese or quantitatively discordant).

**Other Protocols**

See (7) for procedures for (i) creating mock bacterial communities to benchmark standard methods for 16S rRNA sequencing and LEA-Seq, (ii) generating robotically arranged personal bacterial culture collections from human fecal samples, (iii) isolating *B. thetaiotaomicron* strains from fecal samples collected over time from individuals and family members, and (iv) sequencing microbial genomes.

**References and Notes**