Gut microbiome westernization in Hmong and Karen refugees and immigrants in the United States

A Dissertation
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

PAJAU VANGAY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dan Knights

August 2018

Acknowledgements

First and foremost, I would like to thank the participants of our research study for their generous time, support, and trust. This research would not have been possible without Shannon Pergament, who was so patient in helping me navigate community-based research, and Bwei Paw, Mary Xiong, and Pimpanitta Saenyakul, who were absolutely central to our success. In addition, a warm-hearted thank you to our community advisory boards for your commitment throughout this entire study: Bu Bu, Jamiey Cha, Yoha Christiansen, Pa Chua Vang, Duachi Her, Ku Ku Paw Lynn, Mayly Lochungvu, Mudah Takoni, Aye Mi San, Yeng Moua, Ko Nay Oo, Donna Vue Lee, Houa Vue-Her, Pakou Xiong, and Shoua Yang. My work in Thailand would not have been possible without Chaisiri Angkurawaranon and Rose McGready, who truly enabled our research to take place internationally. A big thank you to our Thai-based community researchers as well: Ntxawm Lis, Yi Lis, Blooming Zion, Htoo Lay Paw, Moo Kho Paw, See Thoj, and Wirachon Yangyuenkun. I am also indebted to Grant Kim, Austin Kim, Nurul Quratulaini Abd Salim Nast, Dominique Sabas, and Max Abramson, who volunteered so much of their precious time on this research. In addition, I would especially like to thank Jennifer Moss for working so hard to turn our research results into an animated video, as well as Nancy Yang, who was exceptional in leading community discussions and managing dissemination events.

Thank you to our collaborators who spent countless hours with our experiments: Sarah Lucas, Ryan Hunter, Lisa Till, Purna Kashyap, Lalit Beura, Emily Thompson, Dave Masopust. In addition, my Ph.D. committee who not only enthusiastically served, but were also so generous with their time: Mike Sadowsky, Chery Smith, Ran Blekhman, John Carlis, and Claudia Neuhauser.

A huge thank you to my amazing labmates in the Knights Lab: Tonya Ward, Robin Shields-Cutler, Abby Johnson, Gabe Al-Ghalith, Ben Hillmann, Jonathan Clayton, and Jesse Kennedy—my days never felt like work with such smart, fun, and truly good-hearted individuals.

I am incredibly privileged to have been mentored by two brilliant individuals: Kathie Culhane-Pera and Dan Knights. Kathie was instrumental in connecting me to my own Hmong community in Minnesota and our collaborators in Thailand, and introduced me to the world of community-based research. Dan, my extraordinary Ph.D. advisor, who instilled a sense of adventure in all of us, encouraged every idea I came to him with, no matter how ambitious. Thank you for your commitment to science, your generosity and genuine care for others, and most of all, for lifting all who surround you.

And finally, I am especially proud to share the completion of my Ph.D. with my family. I am eternally indebted to my life partner, Eric, who willingly trudges through my hurdles with me; to my siblings, who unconditionally support me in every way imaginable; to my nieces and nephews, who are constant reminders to have fun; and to my parents, who have sacrificed the world for me to pursue my dreams.

Dedication

I dedicate my dissertation to the generations of Hmong who came before and will come after me. Our culture is so precious, as are our microbes.

Table of Contents

Table of Contents iv
List of Tables viii
List of Figures ix
Chapter 1: General Introduction 1
Introduction 2
Background and Literature Review 3
The U.S. is home to a growing immigrant population whose health is declining over time 3
Central role of the human microbiome in human health 4
Western diet induces pronounced deleterious effects on the gut microbiome 4
Loss of co-evolved microbes contributes to rise in chronic diseases 5
Other western exposures contribute to disruption of gut homeostasis 5
Gut microbiome mediates obesity 6
Effects of fiber on gut health and obesity 6
Study objectives and overall approach 7
Specific Aim 1: Determine changes in the gut microbiome attributed to the length of residence in the U.S. and obesity risk. 7
Specific Aim 2: Characterize gut microbiome recovery immediately after immigration. 8
Specific Aim 3: Identify dietary components with utility of preserving the native gut microbiome. 8
Community-based research 8
Dissertation Organization 9
Chapter 2: Antibiotics, pediatric dysbiosis, and disease 11
Introduction 12
Overprescription trends 12
Additional harmful effects of antibiotic exposure 13
Normal host-microbiome development 14

Gastrointestinal development 14
Microbiome development 15
Important host-microbiome interactions 16
Other major influences on microbiome development 17
Approaches for studying pediatric dysbiosis and related disease 19
A Dysbiosis-centric view 19
A Disease-centric view 20
An Age-centric view 21
A Response-centric view 21
A Recovery-centric view 22
Current evidence for disease mechanisms 23
Obesity 23
Allergy and Atopic Disorders 24
Autoimmune diseases 26
Infectious diseases 27
Future directions 28
Chapter 3: A machine learning repository for microbiome datasets 30
Introduction 31
Workflow 32
Methods 33
Pre-processing of sequencing reads 33
Selection of classification tasks 33
Website generation 34
Case Study Benchmarking 34
Publicly available web-based interface 35
Benefits of curated microbiome-based prediction tasks 35
Comparison to similar databases 36
Case studies 36
Future Work 37
Conclusions 38

Chapter 4: U.S. immigration westernizes the human gut microbiome 39
Introduction 40
Methods 42
Study setting, population, and recruitment. 42
Application of Community-based Participatory Action Research methods 43
Cross-sectional specimen and data collection, U.S. 44
Longitudinal specimen and data collection, U.S. 44
Specimen and data collection, Thailand 45
Stool sample collection 45
Dietary data processing workflow 45
16S sample processing and sequencing 46
Deep shotgun metagenomic sample processing and sequencing 47
Food-Microbiome Procrustes distance associations 47
Predicted biosynthetic gene clusters 48
Mouse experiment specimen and data collection 48
Histology 49
Cell isolation, staining and flow cytometry 49
Other limitations 50
Results 50
Assembly of a multi-generational Asian American immigrant cohort 50
U.S. immigration is associated with loss of gut microbes 52
Bacteroides strains displace Prevotella strains across generations in the U.S. 53
U.S. immigrants lose enzymes associated with plant fiber degradation 54
Dietary acculturation partly explains microbiome acculturation 55
Gut biodiversity decreases according to duration of residence in the U.S. 57
Prevotella displacement continues for more than one decade 57
Microbiome Westernization begins within 9 months after immigration 58
Western diet and Western microbes lead to adverse health outcomes 59
Discussion 59
Chapter 5: Concluding remarks and future work 63

Broader Impacts 65
Comprehensive assessment of microbiome structure in two minority ethnic groups 65
International project spanning migratory paths of the Hmong and Karen 66
Cutting-edge tools enable novel insights into microbiome analyses 66
Potential for translatable results with impact to public health in Minnesota 67
Bibliography 103
Appendix 128
Appendix A. Survey form. 128
Appendix B. Fecal sampling instructions 132

List of Tables

Chapter 2 Tables 68
Table 2.1. References synthesizing mechanistic and epidemiological evidence linking antibiotics, changes in the gut microbiome, and disease. 68
Chapter 3 Tables 69
Table 3.1. Dataset descriptions 69
Chapter 4 Tables 70
Table 4.1. Sample group recruitment stratified by BMI threshold of 25. 70
Table 4.2. Sample Group Characteristics. 71
Table 4.3. Foods that were individually researched, then entered as custom foods into SuperTracker and assigned a unique identifier for the food tree. 72
Table 4.4. OTU prevalences in HmongThai and Hmong1st. Related to Figure 2. 76
Table 4.5. NCBI Genome IDs of Bacteroides and Prevotella strains. Related to Figure 3. 77
Table 4.6. Foods (summarized at level 3) that are significantly correlated with PC1 of the diet-based unweighted Unifrac PCOA ($q < 0.01$). Related to Figure 4. 78

List of Figures

Chapter 2 Figures 79
Figure 2.1. Disease model of host-microbiome development. 79
Figure 2.2. Trajectories for infant recovery after antibiotic exposure. 80
Figure 2.3. Percent decrease in gut microbiome biodiversity across studies with different antibiotic exposures. 81
Figure 2.4. Predicted Microbiome Maturity Index (MMI). 82
Chapter 3 Figures 83
Figure 3.1. Workflow of data and website generation. 83
Figure 3.2. Website screenshots of MLRepo homepage, task, and dataset pages. 84
Figure 3.3. ROCs comparing random forest and SVM with different kernels. 85
Figure 3.4. ROCs comparing NCBI RefSeq and Greengenes 97 databases. 86
Figure 3.5. Summary statistics of framework and database comparisons. 87
Chapter 4 Figures 88
Figure 4.1. Assembly of a multi-generational Asian American cohort, while accounting for BMI and diet 88
Figure 4.2. Loss of diversity and native bacterial taxa with time in the U.S. 89
Figure 4.3. Bacteroides and Prevotella strain diversity and abundances 90
Figure 4.4. Dietary acculturation is detectable using novel food tree and partially explains microbiome variation 91
Figure 4.5. Gut biodiversity decreases with time spent in the U.S. 92
Figure 4.6. Prevotella displacement is observable within one decade in the U.S. 93
Figure 4.7. Longitudinal changes immediately pre- and post-arrival to the U.S. 94
Chapter 4 Supplemental Figures 96
Figure 4.S1. Geographical locations of recruitment sites in Thailand. Related to Figure 4.1. 96
Figure 4.S2. Alpha diversity boxplots of obese and lean individuals, separated by ethnicity. Related to Figure 4.4. 97
Figure 4.S3. Functional annotations. Related to Figure 4.4 98

Figure 4.S4. Macronutrient pairwise comparisons. Related to Figure 4.4. 99
Figure 4.S5. Bipartite network of participant dietary records and food items. Related to Figure 4.4. 100
Figure 4.S6. Procrustes of diet and microbiome distances. Related to Figure 4.4. 101
Figure 4.S7. PCoA of unweighted Unifrac distances of longitudinal samples. Related to Figure 4.7. 102

Chapter 1: General Introduction

Introduction

Immigrants to the U.S. arrive relatively free of endemic chronic diseases, but their health eventually deteriorates over time to match that of their American counterparts. This "healthy immigrant effect" (Kennedy et al., 2006) unfolds at different rates and with different severities, as some groups develop disproportionately higher risks of disease than others (Antecol and Bedard, 2006). Given that immigrants are the fastest growing group in the U.S. and are projected to make up 20% of the population by 2050 (Passel and Cohn, 2008), prevention of obesity and related diseases in this population will have large implications on public health and the U.S. economic health burden. Behavioral and social risk factors (Bates et al., 2008; Cairney and Ostbye, 1999; Goel et al., 2004; Kaplan et al., 2004; Lauderdale and Rathouz, 2000; McTigue et al., 2002; Walker et al., 2008), such as sedentary lifestyles, western diet, and low socioeconomic status, have been well-described and are important considerations for preventing and managing obesity. Unfortunately, the etiology of obesity is further complicated by host-specific factors such as genetics and more recently, the gut microbiome (Ley et al., 2005, 2006; Schwiertz et al., 2010; Turnbaugh and Gordon, 2009; Turnbaugh et al., 2006). The gut microbiome plays a critical role in host metabolism and because it is largely shaped by the external environment, varies based on an individual's geographical origin (Yatsunenko et al., 2012). New and severe environmental exposures can lead to disruptions in gut homoestasis (David et al., 2014; Dethlefsen and Relman, 2011; Turnbaugh et al., 2009a), which have been associated with a variety of diseases, including obesity. Thus, there is a critical need to determine how the gut microbiome adapts to the permanent and often severe environmental changes characteristic of immigration between under-developed and western countries. In the absence of such knowledge, the strategies used to combat obesity and related diseases among U.S. immigrant and minority populations will only be partially effective.

Background and Literature Review

The U.S. is home to a growing immigrant population whose health is declining over time

An estimated 40 million immigrants (foreign-born individuals) reside in the U.S. (13% of the population as of 2010), representing the fastest growing segment of the population (Goel et al., 2004). Immigrants are projected to make up 20% of the population by 2050 (Passel and Cohn, 2008). Minnesota is home to the highest number of refugees per capita in the U.S., and is currently expecting the largest wave of refugees in the last decade (Koumpilova, 2015). The Hmong, a minority ethnic group from China and Southeast Asia, make up the largest refugee group in Minnesota (22,033 total refugees as of 2014) (Minnesota Department of Health), and also form the largest centralized Hmong community in the U.S. (70,000 total individuals) (Pfeifer and Thao, 2013). Karen refugees from Burma (Myanmar) have been arriving in large numbers in recent years (Minnesota Department of Health), also from a similar region in Asia. Past work reveals that length of residence in the U.S. increases the risk of obesity, with some groups experiencing up to a four-fold increase in obesity risk after 15 years (Bates et al., 2008; Cairney and Ostbye, 1999; Goel et al., 2004; Kaplan et al., 2004; Lauderdale and Rathouz, 2000; Walker et al., 2008). This "healthy immigrant effect" has been well-documented in western countries (Antecol and Bedard, 2006), yet occurs at varying degrees among different groups (e.g. Mexican-born females are at highest risk of developing obesity (Barcenas et al., 2007)). In Minnesota, increasing levels of dietary acculturation has been correlated with increasing BMI among Hmong (Mulasi-Pokhriyal et al., 2012; Smith and Franzen-Castle, 2012), and the prevalence of overweight and obesity is highest among Hmong compared to other Asian ethnic groups (Arcan et al., 2014; Franzen and Smith, 2009; Himes et al., 1992; Mulasi-Pokhriyal et al., 2012). Limited data exists for the more recently arrived Karen, yet these trends suggest that this group will soon be at risk. With the recent crises in the middle east, the U.S. is expected to increase the number of accepted refugees by 25% over the next two years (Morello, 2016). There is a need to gain a basic understanding of how these migrations impact

human health in order to ensure that these populations transition to life in the U.S. as seamlessly as possible.

Central role of the human microbiome in human health

The human microbiome, the trillions of microorganisms that live inside and on our bodies, is an important contributor to human health. These microbial communities aid in immune system development, protection against pathogenic infections, and host metabolism. Furthermore, distinct gut microbiome compositions have been associated with various disease states, such as Crohn's disease (Gevers et al., 2014), cancers (Wang et al., 2012), diabetes (Qin et al., 2012), allergy and asthma (Abrahamsson et al., 2012; Atarashi et al., 2013; Bisgaard et al., 2011), and obesity (Ley et al., 2006; Turnbaugh and Gordon, 2009; Turnbaugh et al., 2006). There is enormous potential for the gut microbiome to serve as a biomarker for the diagnostics and prevention of diseases, and to also serve as a target for prebiotic and probiotic therapeutics in combating disease.

Western diet induces pronounced deleterious effects on the gut microbiome

With a critical role in host metabolism, the gut microbiome composition is heavily influenced by an individual's long-term diet (Hildebrandt et al., 2009; Wu et al., 2011), yet can also quickly respond to drastic dietary changes (David et al., 2014; Turnbaugh et al., 2009a). African Americans and rural South Africans who exchanged diets for two weeks also exchanged their respective cancer risks, as indicated by mucosal biomarkers and microbiome shifts (O'Keefe et al., 2015). Aside from being animal-based, high-fat, and low-fiber, the western diet harbors other food components that have recently been under scrutiny, such as artificial sweeteners inducing glucose intolerance (Suez et al., 2014) and dietary emulsifiers inducing metabolic syndrome (Chassaing et al., 2015). Although diet has been shown to be a major determinant of gut microbiome composition, it also exhibits resilience in some states, and therefore may show limited response to diet alone (Smith et al., 2013). Perhaps more importantly for immigrants and their future generations, diet modifications are unable to recover microbes that have been lost over several generations (Sonnenburg et al., 2016).

Loss of co-evolved microbes contributes to rise in chronic diseases

Although our study captures diet acculturation as a western exposure driving risk of obesity, we also explore evidence of how westernization contributes to the "disappearing microbiota" hypothesis (Blaser and Falkow, 2009), which suggests that the loss of indigenous organisms that have coevolved with humans may be contributing to the rise in chronic diseases. This hypothesis is supported by several studies, which found that greater diversity and novel taxa are present in non-western versus western human gut microbiomes (Clemente et al., 2015; Obregon-Tito et al., 2015; Yatsunenko et al., 2012). Furthermore, there may be a strong health-promoting relationship between a specific set of microbes and the respective population that harbors it. Kodaman et al. found that two clusters of *Helicobacter pylori* strains were found among African and Amerindian populations, and were benign when strain ancestry and human ancestry matched, but deleterious for gastric cancer risk when an individual with African ancestry harbored an Amerindian *H. pylori* strain and vice-versa (Kodaman et al., 2014).

Other western exposures contribute to disruption of gut homeostasis

Exposure to broad-spectrum antibiotics, commonly used in western medicine, can lead to imbalances in the gut microbiome. Evidence for antibiotics-induced obesity is primarily characterized by shifts in functional capability, or more specifically, long-lasting metabolic shifts that result from incomplete recovery back to the normal trajectory. Recent work found that mice given sub-therapeutic levels of antibiotics after weaning exhibited increased adiposity, large taxonomic changes in their gut microbiomes, and increased levels of short-chain fatty acids (SFCAs) as well as counts of bacterial genes involved in SFCA metabolism (Cho et al., 2012a). These mice also had lower caloric output in their faecal pellets despite dietary intake similar to controls, suggesting their gut microbiota developed the ability to extract increased energy from indigestible components (Cho et al., 2012a). Furthermore, use of multiple courses of broad-spectrum antibiotics can perturb the gut microbiome such that complete recovery is unattainable (Dethlefsen and Relman, 2011).

Gut microbiome mediates obesity

Obesity risk can be attributed partially to host genetics, but recently has been found to be considerably associated with differences in the gut microbiome (Ley et al., 2005, 2006; Schwiertz et al., 2010; Turnbaugh and Gordon, 2009; Turnbaugh et al., 2006). The gut microbiome is much more dynamic than the human genome; it exhibits substantial inter-individual variability (unrelated individuals share only 30% of their gut microbiomes as opposed to 99.9% of their genomes), and intra-individual variability (an individual's gut microbiome changes over days, weeks, and months (Bäckhed et al., 2005; Faith et al., 2013)). Gut microbiomes transplanted from lean or obese individuals into mice directly induces weight loss or gain, respectively, implicating the gut microbiome as a causal factor in obesity (Ridaura et al., 2013; Smith et al., 2013; Turnbaugh et al., 2006). Additionally, weight loss is dependent on the initial gut microbiome composition, and stability of an individual's weight is directly associated with the stability of his or her gut microbiome (Faith et al., 2013; Santacruz et al., 2009).

Effects of fiber on gut health and obesity

Although the gut microbiome may serve as an important biomarker for obesity, it also holds enormous potential for modifying host metabolism. Consumption of dietary fiber is important for promoting gut microbial diversity; it is fermented into beneficial short-chain fatty acids, and as a result is important for combating various gut conditions associated with obesity, such as increased gut permeability and low-grade inflammation (Cani et al., 2009a; Maachi et al., 2004). Studies show that gut microbiota fermentation of dietary fiber promotes satiety (Cani et al., 2009b; Parnell and Reimer, 2012) and is involved in a variety of signaling pathways that maintain glucose and energy homeostasis (De Vadder et al., 2014). It is well established among epidemiological studies that consumption of dietary fiber is important for weight loss (Howarth et al., 2001; Liu et al., 2003; Slavin, 2005), and that non-Western populations with fiber-rich diets have lower incidences of obesity (Cerqueira et al., 1979; Lindeberg and Lundh, 1993). Characterizing the effect of dietary fiber consumption in preserving or restoring native gut microbiomes among immigrant populations has significant implications for both the

prevention and treatment of obesity, and subsequent development of obesity-related diseases, in these at-risk populations.

Study objectives and overall approach

The long-term goal of the research contained in this thesis is to identify key factors that support a healthy transition to life in the U.S., so that improved strategies can prevent the development of endemic chronic diseases among new Americans. The main objective is to understand the role of the gut microbiome in mediating increased risk of obesity within Hmong and Karen refugees and immigrants as they adapt to western diet and environment. We hypothesize that immigration and subsequent exposure to westernization induces dramatic and permanent shifts in the gut microbiome directly associated with increased risk of obesity. The rationale for the main human study described in this thesis is that its successful completion would provide a strong conceptual framework for the implementation of a comprehensive dietary or probiotic intervention targeted at multiple groups.

Specific Aim 1: Determine changes in the gut microbiome attributed to the length of residence in the U.S. and obesity risk.

Hypothesis: Obesity risk is characterized by low diversity, overgrowth of pathogens, or increased energy extraction in the gut microbiome, and becomes more pronounced with U.S. residence.

- Analyze gut microbiomes from a cross-section of lean and overweight/obese Hmong and Karen women prior to immigration, newly arrived in the U.S., resided longer-term in the U.S., and who were born in the U.S.
- Use taxonomic marker gene amplicon sequencing and deep shotgun metagenomics to measure shifts in functional and taxonomic compositions, and taxonomic biodiversity
- Correlate microbiome features with anthropometric measurements used to assess obesity risk and length of residence in the U.S., while statistically controlling for intake of a subset of dietary nutrients, migration history, and antibiotic exposures

Specific Aim 2: Characterize gut microbiome recovery immediately after immigration.
Hypothesis: Immigration induces compositional and functional shifts observable within the first 6 months in the U.S.

- Characterize gut microbiome adaptation to residency in the U.S. at 1-month intervals within the same individual, and statistically controlling for intake of a subset of dietary nutrients, migration history, antibiotic exposures, and anthropometric measurements
- Use taxonomic marker gene amplicon sequencing and deep shotgun metagenomics to measure shifts in functional, taxonomic, taxonomic biodiversity, and rate of change in the gut microbiome

Specific Aim 3: Identify dietary components with utility of preserving the native gut microbiome.
Hypothesis: Dietary fiber promotes maintenance of the native gut microbiome and protects against obesity risk.

- Transplant gut microbiomes from pre-immigration and second-generation immigrant individuals into gnotobiotic mice, and apply a series of dietary interventions including varying diversity and concentration of fibers to determine their ability to preserve or deplete the native gut microbiome.
- Dietary intervention groups will be compared to assess differences in body composition, metabolic markers, and metagenomic compositions to infer mechanistic connections between the microbiome, diet, and metabolism.

Community-based research
Community involvement was a critical aspect of the human study described in Chapter 4 as the research team is a composed of community and academic members. In addition to the co-Principal Investigators and co-investigators listed above, our U.S. research team is composed of Hmong and Karen community researchers, all of whom have been trained on CBPAR methods and are current leaders in their communities. Their main roles have

been to consult with the development of recruitment materials, translate recruitment materials, contribute to the study design, promote and recruit for the study, prepare gut health educational materials, and deliver the dissemination event. To enhance community input, we have formed a community advisory board (CAB) with community leaders, community health professionals, and at-large community members from both Hmong and Karen communities. Early in the project, the community-academic research team held discussions with the CAB to discuss knowledge and awareness of the project topic, identify optimal recruitment strategies, review the study design and recruitment materials, and pilot the consent forms and surveys. The information from these discussions directly informed our study design, methodology, and dissemination strategies. Two additional advisory board meetings were held throughout the study period: (1) to check the progress of recruitment, troubleshoot, and improve the process, and (2) to prepare for the dissemination of results at the end of the study.

Dissertation Organization

This dissertation is organized as three manuscripts which have either been published or are being prepared for submission. It contains a general introduction, a description of the overall study objective and approach, one perspective paper, one technical paper, and one research paper, and an overall conclusions chapter. Note that these distinct article formats differ in organization; the technical paper does not contain a discussion section, and the research paper, which is formatted for submission to a specific journal, has a results section that embodies limited methods, results, and some discussion. Figures and references for all chapters can be found at the end of this paper.

Chapter 2 is a perspective paper that combines both an in-depth literature review with a proposed model of studying antibiotic-induced dysbiosis in the infant gut microbiome. This chapter lays the groundwork for how critical the gut microbiome is in immune system development and for maintaining long-term health, and discusses an important Western-associated factor, antibiotics, in the context of gut health. Chapter 3 is a technical note describing a machine learning repository of curated microbiome datasets

for the computational community. This chapter is the result of important validation work done with published datasets, which was necessary for estimating sample sizes, selecting computational and statistical methods, and validating bioinformatics tools used for the human study found in Chapter 4. Chapter 4 forms the focal point of this dissertation, and describes human gut microbiome changes in immigrants and refugees in the United States (U.S.). Chapter 5 provides a conclusion section that discusses the findings, broader impact, and future work resulting from this thesis.

Chapter 2: Antibiotics, pediatric dysbiosis, and disease

Pajau Vangay, Tonya Ward, Jeffrey Gerber, and Dan Knights

*A version of this chapter has been published in:
Cell Host Microbe. 2015 May 13;17(5):553-64.

This chapter discusses an exposure commonly associated with developed countries: antibiotics. Here, we describe the important relationship between the gut microbiome and the immune system, and discuss the ramifications of overuse of antibiotics and its deleterious effects on gut microbiome development.

Introduction

Epidemiological studies have identified links between antibiotic usage in early infancy and diseases such as obesity, diabetes, and asthma. Longitudinal studies of antibiotic usage have demonstrated profound short- and long-term effects of antibiotics on the diversity and composition of the gut microbiota. While a large and growing number of studies implicate dysbiosis in numerous diseases (Biedermann and Rogler, 2015), there are currently few studies directly linking antibiotics, pediatric dysbiosis, and the later development of disease. The vast majority of antibiotic use occurs in the outpatient setting, where up to a third of prescriptions are unnecessary. Furthermore, even when antibiotics are indicated, the use of broad-spectrum antibiotics has increased dramatically over the past two decades, which could influence the structure and function of the developing microbiome more dramatically than targeted antibiotics. Understanding the short- and long-term effects of early life antibiotic use on the diversity and composition of the gut microbiota is critical in identifying the risks associated with these emerging prescription trends.

Overprescription trends

Antibiotics are by far the most common prescription drugs given to children (Chai et al., 2012). In 2010, children received 74.5 million outpatient antibiotic prescriptions—one for every child in the US—accounting for one fourth of all medications for children (Hicks et al., 2013). Numerous studies have demonstrated that antibiotics are often prescribed unnecessarily (Gonzales et al., 2001; McCaig et al., 2003; Nash et al., 2002), with some estimates as high as 50% (Kronman et al., 2014). Nearly 30% of children receive an antibiotic prescription during an outpatient primary care visit (McCaig et al.,

2003), most often inappropriately, for viral upper respiratory tract infections (Gonzales et al., 2001; Nash et al., 2002; Nyquist et al., 1998). Overuse of broad-spectrum antibiotics for conditions responsive to narrow-spectrum agents has been dramatically increasing (Hersh et al., 2013). Even after adjusting for differences in patient age, comorbidities, and sociodemographic factors, children with the same infections can receive vastly different rates of antibiotic prescriptions depending upon the practice or clinician visited (Fierro et al., 2014; Gerber et al., 2014). This phenomenon also seems to be universal: per capita antibiotic prescribing rates vary widely across US states (Hicks et al., 2013) and European countries (Goossens et al., 2005) , without reasonable cause for geographic differences in bacterial infection rates. These prescribing patterns suggest that (1) antibiotics are often overprescribed and (2) benchmarking data available in the form of clinical practices and geographical regions can be used to prescribe lower rates of antibiotics to help guide more judicious prescribing elsewhere.

Additional harmful effects of antibiotic exposure

In addition to the gut microbiome-mediated effects discussed below, inappropriate prescribing of antibiotics can lead to both drug-related adverse effects and the promotion of antibiotic resistance. More than 140,000 emergency department (ED) visits occur annually in the US for antimicrobial-related adverse effects, comprising almost 20% of all ED visits for drug-related adverse effects (Shehab et al., 2008). In addition to this direct patient harm, antibiotic use has been associated with the emergence of antimicrobial resistance, identified by the World Health Organization (WHO) as “one of the three greatest threats to human health.” Infections with resistant bacteria increase morbidity and mortality, and greatly increase the cost of medical care; the Institute of Medicine estimated that, in 2010, roughly $20 billion was spent on the treatment of antibiotic-resistant infections. Knowledge of these facts, however, has done little to curb antimicrobial use. Improving our awareness of the long-term implications of both necessary and unnecessary antibiotic exposure is important to better inform the risk/benefit ratio for antibiotic prescribing and to improve child health.

Normal host-microbiome development

Gastrointestinal development

Gastrointestinal (GI) development occurs throughout embryonic life, and its basic structure is first formed by the end of the first gestational trimester (Montgomery et al., 1999). Tight junctions are present by 10 weeks of gestation and intestinal villi are formed by weeks 12-19 (Maheshwari and Zemlin, 2009; Montgomery et al., 1999; Polak-Charcon et al., 1980). Postnatally, an abrupt shift in exposure from amniotic fluid to first foods occurs in the GI tract. This induces many changes along the GI tract, including a change in pH of the stomach. For example, some reports state the pH of the stomach is initially in the range of 6 – 8 (Avery et al., 1966), likely due to buffering by the amniotic fluid, which decreases to that of an adult (pH 1.5 - 2.5) within the first hours following birth (Kelly et al., 1993; Lebenthal and Lebenthal, 1999; Ménard, 2004). However, due to the consumption of milk, and its buffering capabilities, the pH of the infant stomach often returns to a circum neutral level of 7-7.6 (Hibberd et al., 1982). The higher pH of the stomach early in life has a meaningful impact, including a higher absorption rate of nutrients and a diminished digestive capacity compared to later in life, which may support transit of ingested bacteria to colonize the lower GI tract. Throughout postnatal development, the infant GI tract also increases in size in both length and in diameter, and loses most of its early-stage porosity within days post birth due to milk-borne growth factors and hormones that stimulate growth and development (Cummins and Thompson, 2002).

Development of the GI-associated lymphoid tissue (GALT), including mesenteric lymph nodes, Peyer's patches, and lymphocytes in the lamina propria, is complete in full-term infants at birth (Forchielli and Walker, 2005). For example, goblet cells, responsible for mucin production, are functional by 12 weeks of gestation (Montgomery et al., 1999), as are Paneth cells, which can secrete defensins and lysozymes by gestational weeks 13 and 20, respectively (Louis and Lin, 2009; Maheshwari and Zemlin, 2009; Rumbo and Schiffrin, 2005). Although full-term infants are born with fully developed digestive

tracts, exogenous stimulation through exposure to dietary antigens, hormones, growth factors, and bacteria is required to elicit proper function throughout life (Forchielli and Walker, 2005).

Microbiome development

Although the gastrointestinal tract of a healthy infant is generally considered to be sterile before birth, recent work suggests that initial colonization may take place in-utero (Aagaard et al., 2014; Funkhouser and Bordenstein, 2013; Matamoros et al., 2013). Hours after birth, microorganisms from the mother's vaginal, fecal, and/or skin microbiome and the environment are important colonizers of the infant gut (Penders et al., 2006), with actual contributions depending on mode of delivery. Several other factors including prematurity, infant diet (breast milk or formula), hygiene, and use of antibiotics will ultimately impact the composition of the infant gut microbiome. Despite a seemingly chaotic colonization, with large swings in composition over time, gut microbiome development is governed by Darwinian dynamics: microbes best adapted for the changing conditions of the gut will be most likely to survive. We can see this clearly throughout the first few weeks of life, as the colonization of facultative aerobes reduces the availability of oxygen, which then permit the growth of strict anaerobes (Bezirtzoglou, 1997). As illustrated in Figure 2.1, compositional changes in response to diet and host development ocurr throughout the first year of life. In the United States, the infant gut is initially colonized with Proteobacteria and Firmicutes, followed by a gradual increase in Actinobacteria (potentially due to the introduction of breastmilk (Favier et al., 2003; Sela et al., 2008; Yoshioka et al., 1983)). By six months of age, Bacteroidetes dominate while Proteobacteria and Actinobacteria gradually decline, which may be attributed to the abundance of carbohydrates in solid foods that coincides with weaning (Koenig et al., 2011; Vaishampayan et al., 2010). By the end of the first year of life, the infant gut is dominated by bacterial phyla Bacteroides and Firmicutes (Figure 2.1). The healthy infant gut continues with dramatic compositional changes throughout the first two years of life before becoming indistinguishable from an adult gut microbiome at age three (Yatsunenko et al., 2012).

Important host-microbiome interactions

Maturation of the intestinal immune system is contingent on parallel development of the gut microbiome (Figure 2.1). Germ-free animals have been found with significant immunological defects in the gut-associated lymphoid tissues (Falk et al., 1998; Macpherson and Harris, 2004) as well as improper development of Peyer's patches and mesenteric lymph nodes (Round and Mazmanian, 2009). Peyer's patches and the mesenteric lymph nodes develop prenatally and isolated lymphoid follicles develop postnatally, but all of these tissues require interaction with key members of the gut microbiome in order to ensure proper differentiation and specification, and complete development of adaptive immunity (Cherrier and Eberl, 2012; Maynard et al., 2012). The immune system must maintain an anti-inflammatory state (Tsuji and Kosaka, 2008) in the gut, especially during exposure to the considerable number of innocuous antigens from commensals, hormones, and food. The interactions of diverse cell types are necessary to carry out the complex functions of the immune system (Adkins et al., 2004); we highlight several immune cell types with important dependencies on the gut microbiome. Dendritic cells (DCs), one of the most important types of antigen-presenting cells, sample the lumen and are responsible for orchestrating inflammatory or tolerogenic responses. To help the immune system carry out appropriate responses, DCs can suppress or induce the activation of antigen-specific T cells, and have the unique ability to differentiate naive T cells into effector or regulatory T cells to target specific antigens (Lanzavecchia and Sallusto, 2001; Macatonia et al., 1995). T helper cells are critical in processing presented antigens into specific cytokines that provide direction for other immune cells and to eventually generate an immunological response. Members of the gut microbiome have been found to differentiate The Th17 class of T helper cellssecrete IL-17 to produce defensins (Kao et al., 2004) and recruit neutrophils (Aujla et al., 2007) to fight infections at mucosal surfaces (Atarashi et al., 2008; Ivanov et al., 2009). Pro-inflammatory Th17 cells must maintain balance with anti-inflammatory regulatory T cells, particularly for the prevention of autoimmune disorders. Certain Clostridia strains have been found to help with expansion and differentiation of regulatory T cells (Atarashi et al., 2013), and have a

direct role in reducing intestinal epithelial permeability by stimulating innate lymphoid cell and T cell production of cytokine IL-22 (Stefka et al., 2014). Innate lymphoid cells help induce pro-inflammatory responses and serve as the main source of IL-22 (Sawa et al., 2010); this cytokine is important for inducing mucus production from goblet cells, stimulating the production of antibacterial proteins, protecting cells from damage, and regulating cell differentiation (Sabat et al., 2014). A number of studies found that microbial signals modulate the amount of IL-22 produced by innate lymphoid cells (Sanos et al., 2009; Satoh-Takayama et al., 2008; Sawa et al., 2010; Sonnenberg et al., 2012; Stefka et al., 2014; Vonarbourg et al., 2010), suggesting the importance of the gut microbiome in host defense mechanisms against infectious and inflammatory diseases (Rutz et al., 2013). Furthermore, *Bifidobacterium longum* has been found to assist in the maturation of DCs in Peyer's Patches and the development of T cells in the thymus (Dong et al., 2010). Specific microbial signals have been deemed necessary for proper education of regulatory T cells and invariant natural killer T (iNKT) cells (Hansen et al., 2012; Olszak et al., 2012), which are a subset of T cells capable of quickly inducing an abundance of cytokines that can stimulate or suppress a variety of immune responses. Additional important microbe-host interactions and mechanisms will be presented later in the context of our proposed model. Considering how critical the various immune cells and their intricate signaling networks are for supporting immune health, disruptions hindering their development may have lasting deleterious effects.

Other major influences on microbiome development

Diet plays a large role in the colonization of the modern infant GI tract due to the vast compositional differences between human milk and infant formula. The most notable difference in the microbiome of breastfed versus formula fed infants is the predominance of *Bifidobacteria* and *Lactobacilli* in breastfed infants, while formula-fed infants harbor more *Enterococci* and Enterobacteria (Palmer et al., 2007). There are also easily detected differences in total community membership between breastfed and formula-fed infants when looking at twin cohorts (Yatsunenko et al., 2012). Human milk is able to modulate bacterial colonization in the infant gut with distinct components not found in formulas:

the human milk microbiome, factors that stimulate bacterial growth (prebiotics), and factors that prevent bacterial growth (antimicrobials). The human milk microbiome consists primarily of Proteobacteria and Firmicutes (Nichols et al., 1974), and has of a core group of taxa found in most human milk samples that include *Staphylococcus*, *Streptococcus*, *Serratia*, *Pseudomonas*, *Corynebacterium*, *Ralstonia*, *Propionibacterium*, *Sphingomonas*, and Bradyrhizobiaceae (Hunt et al., 2011). The human milk microbiome also changes over time, and is dependent on the mother's weight (Cabrera-Rubio et al., 2012). For example, *Weissella*, *Leuconostoc*, *Staphylococcus*, *Streptococcus*, and *Lactococcus* are predominant in milk immediately after giving birth, and milk from obese mothers is less diverse than that of non-obese mothers (Cabrera-Rubio et al., 2012). These ingested bacteria provide a constant source of community members to help colonize the GI tract. Milk-borne prebiotics that modulate the bacteria present in the GI tract include human milk oligosaccharides (HMOs), which are sugars produced solely for consumption by microbes. These include the "original" HMO, bifidus factor (Nichols et al., 1974), that stimulates *Bifidobacterium bifidum* and hundreds of other sugars (all within a family of unconjugated glycans containing lactose at the reducing end) which primarily promote the growth of *Bifidobacterium longum* subsp. *infantis (Bode, 2012)*. Antimicrobials in human milk that also influence the microbes within the GI tract include secretory immunoglobulin A (SIgA), which provides antigen-specific protection against microbes that the mother has already encountered (Rogier et al., 2014), and innate immune proteins, such as lactoferrin and lysozyme, that harbor bactericidal activity (Arnold et al., 1980). Milk obtained from mothers of preterm infants had highest concentrations of cytokines and immunoglobulins immediately after giving birth, further supporting the importance of breast milk consumption in early life (Moles et al., 2015).

Mode of birth delivery has an impact on the microbiome of infants, as the total microbiome (skin, oral mucosa, and nasopharyngeal aspirate, and meconium) of vaginally delivered infants resembles the maternal vaginal and intestinal microbiome, while infants delivered by cesarean section have total microbiomes resembling the maternal skin microbiome (Dominguez-Bello et al., 2010). Specifically, the microbiomes

of vaginally delivered infants consist mostly of *Lactobacillus*, *Prevotella*, *Atopobium*, or *Sneathia* spp, whereas the microbiome of cesarean section delivered infants contain *Staphylococcus* spp (Dominguez-Bello et al., 2010) and less *Bifidobacterium (Biasucci et al., 2010)*.

Approaches for studying pediatric dysbiosis and related disease

The mechanisms and health consequences of pediatric dysbiosis are complex and multifactorial, and are further complicated when also considering infant development (gut microbiome, immune system, and their interactions). Using a systems approach, we consider five interdependent models for understanding dysbiosis that focus on different aspects of disease mechanisms. We discuss these conceptual models in terms of their relative merits for clarity, potential for organization, and ability to express multi-factorial disease pathways. We restrict our hypotheses to those that assume long-term health effects of one or more short discrete courses of antibiotics, since that is by far the most common type of antibiotic exposure in human children (Gevers et al., 2014). Each perspective has strengths in its ability to generalize certain aspects of pediatric dysbiosis. In general, we find a combination of the dysbiosis-centric and disease-centric perspectives to be the most useful for discussing disease mechanisms.

A Dysbiosis-centric view

The gut microbiome is in constant flux; the community composition continuously adapts to environmental exposures and host developmental changes (Caporaso et al., 2011; Human Microbiome Project Consortium, 2012a). This adaptability is essential for maintaining gut homeostasis, but drastic changes, such as those induced by antibiotics, can potentially lead to negative health consequences. Pediatric dysbiosis can be characterized by these drastic changes in the microbial community, represented here as four distinct types. Since broad-spectrum antibiotics are designed to eradicate multiple bacterial taxa, the gut microbiome may be impacted by: (1) an unintended loss of keystone taxa that are critical for maintaining homeostasis or proper host development (e.g., immune system), or (2) an overall loss of biodiversity, which can have inherent

health risks on its own (e.g., the hygiene hypothesis) and also lead to other dysbiosis types (Figure 2.1). Taxa that have been eradicated from their niches leave vacancies to be filled by (3) blooms of pathogens and pathobionts. Even if the infant gut microbiome can recover from these dysbiotic states to arrive at some form of homeostasis, improper or partial recovery can result in a (4) shift in functional capability; for example, becoming more efficient at extracting energy (Figure 2.1). These dysbiosis types sometimes overlap, further adding to the complexity of the system and the challenge of building a unified conceptual model for pediatric dysbiosis research. Viewing pediatric dysbiosis from the perspective of different dysbiosis types is particularly important for understanding how small changes to the relatively simple infant gut can manifest as larger repercussions during adulthood. Such a dysbiosis-type model is crucial for understanding the community dynamics within the gut microbiome, but is limited in its ability to easily address several factors such as the age of the infant, the overlap and transition between dysbiosis types, the many-to-many relationship between dysbiosis types and disease phenotypes, and the parallel development of the immune system.

A Disease-centric view

In the context of different aspects of host development and specific taxa affected, the previously described pediatric dysbiosis types can give rise to a variety of health consequences. Deconstruction of the health outcome with a top-down approach is another model for understanding dysbiosis. In this disease-centered model, health outcomes are generalized by disease class, and then further characterized by specific mechanisms and interactions with subsystems of the model (host immune system, gut microbiome, host development, etc.) (Figure 2.1). For example, obesity-related pediatric dysbiosis in the context of this model begins with antibiotic treatment at any time point during the first two years of life. Biodiversity is depleted during treatment but rebounds after treatment ends, inducing large changes in taxonomic composition. In the case of obesity, these compositional changes also result in functional changes affecting metabolism; the microbiome becomes more efficient at extracting energy from multiple sources, and hence predisposes the host to obesity (Turnbaugh et al., 2006). Antibiotic exposure at a

younger age exacerbates predisposition to disease (Cox et al., 2014) and compounded disturbances may lead to unanticipated consequences (Paine et al., 1998). Other disease classes may include allergies and atopic diseases, autoimmune disorders, diabetes, and infectious disease. This model encapsulates major interdependencies within each disease class while accounting for temporal factors. The main shortcoming of the disease-centric view of pediatric dysbiosis is that it does not easily allow synthesis of common mechanisms across diseases.

An Age-centric view

Dysbiosis types can result in complete recovery with minimal impact to host health, or can have drastic unintended consequences depending on the stage of host development. Development of the microbiome and the host immune system can be categorized conveniently, although approximately, into four general stages: (1) 0 to 6 months, (2) 6 to 12 months, (3) 12 to 24 months, and (4) 24 months and older. The infant is most vulnerable to developing immunological defects during Stage 1, when adaptive immunity interaction with keystone taxa is most critical (Prescott et al., 1999; Rautava et al., 2004; Van Der Velden et al., 2001). By Stage 4, the gut microbiome establishes a new-formed stasis as it reaches maturity, carrying forth any existing functional shifts that could predispose the host to future diseases. Although the vulnerabilities of each stage of development are important considerations for understanding dysbiosis, considering segregated stages hinders characterization of mechanisms that span multiple stages.

A Response-centric view

The gut microbiome transitions through several stages in response to a course of antibiotics: pre-treatment, during treatment, recovery, and long-term stasis (Figure 2.2a). Dysbiosis types that emerge during treatment include loss of keystone taxa and short-term metabolic shifts, both of which would be compounded with multiple courses of antibiotics. Immediately after the antibiotic course, the gut microbiome begins to recover, but not without several potential complications. The loss of diversity imposed by

antibiotics may allow for blooms of pathogens and pathobionts; the adaptive immune system may be underdeveloped and keystone taxa may still not have recovered (further delaying immune development); and metabolic shifts may begin to take place. Eventually the gut microbiome reaches a form of stasis, which may be different from its pre-treatment stage (Figure 2.2). At this stage, permanent metabolic shifts may have been established, a loss of biodiversity accompanied by a bloom of pathobionts may persist, and the host may be predisposed to an increased risk of infectious disease. Although the dynamics of the community structure in response to antibiotics are useful for identifying short-term vulnerabilities, mechanisms of dysbiosis typically start during one stage (e.g., treatment stage) and end in another (e.g., recovery stage), making this model difficult and confusing to work with.

A Recovery-centric view

Although adult gut microbiomes experience day-to-day changes, they are relatively stable when compared to infant gut microbiomes, which are characterized by large swings in taxonomic composition, especially throughout the first year of life. Regardless of the seemingly random shifts, there exists a clear trajectory of healthy development in the infant gut microbiome when assessing biodiversity and relative abundances of specific taxa (Figure 2.2). This model defines dysbiosis in terms of how the microbiome recovers back to this trajectory: fast recovery, slow recovery, or incomplete recovery (Figure 2.2a). During fast recovery, there may be a short-term loss of diversity but keystone taxa are preserved and the gut microbiome quickly rebounds back to normal with little impact to the host. With a slow recovery, there may be loss of keystone taxa during a critical time for interaction with the immune system, therefore causing a delay in immune development. Biodiversity may be low and it may take some time before keystone taxa can reestablish and interact with the immune system before getting back on the normal trajectory. The host is most vulnerable to infectious disease during this prolonged state of recovery, with both an immature immune system and a low-diversity microbiome. Despite eventually recovering and reestablishing a healthy gut microbiome, the adaptive

immune system may have developed antibodies against commensals during the long recovery period, predisposing the host to autoimmune diseases. During an incomplete recovery, the compositional changes are so drastic that the gut microbiome reaches a completely new form of stasis, placing it on a trajectory completely different than expected (Figure 2.2a). These changes are accompanied by functional and metabolic shifts in the gut microbiome and come with disease risks of their own. The types of recovery in this model are not mutually exclusive since it is possible that either a fast or slow recovery rate may lead to an incomplete recovery. This model also does not address how a recovery type may be dependent on a specific development stage, as considered by the age-centric view.

Current evidence for disease mechanisms

In considering several alternative lenses through which to discuss and organize pediatric dysbiosis, we have decided to use a combination of the dysbiosis-centric and disease-centric perspectives for summarizing and synthesizing existing knowledge about potential disease mechanisms (Table 2.1). This combined model allows us to map multiple causes to the same disease, while keeping track of different developmental and treatment stages that underlie the various known or proposed mechanisms. Although the causal pathway between dysbiosis and disease can take many forms, we present four important disease classes in major contributing dysbiosis types.

Obesity

Evidence for antibiotics-induced obesity is primarily characterized by shifts in functional capability, or more specifically, long-lasting metabolic shifts that result from incomplete recovery back to the normal trajectory. Recent work found that mice given sub-therapeutic levels of antibiotics after weaning exhibited increased adiposity, large taxonomic changes in their gut microbiomes, and increased levels of short-chain fatty acids (SFCAs) as well as counts of bacterial genes involved in SFCA metabolism (Cho et al., 2012b). These mice also had lower caloric output in their faecal pellets despite dietary intake similar to controls, suggesting their gut microbiota developed the ability to extract

increased energy from indigestible components (Cho et al., 2012b). Furthermore, low-dose antibiotics started even earlier in life (prior to weaning) resulted in mice showing a more pronounced increase of adiposity, and induced adipogenesis synergistically with a high fat diet; fecal transplantation into germ-free mice lead to increased fat mass relative to transplantation from mice without antibiotics, implicating the gut microbiome in the causal pathway of obesity (Cox et al., 2014). Some epidemiological studies further substantiate the long-lasting effects of early exposures, finding that antibiotic exposures among infants younger than six months are significantly associated with increased BMI later on in life, although in general these findings are somewhat mixed and warrant follow-up in a prospective study (Ajslev et al., 2011; Bailey et al., 2014; Trasande et al., 2013). The 0-6 month window is a time of rapid host and microbiome development, yet also represents a period when the microbiome may be most susceptible to adopting long-term changes. Additional studies, especially with human subjects, are necessary to understand how exposures during various developmental windows can alter the gut microbiome and host metabolism.

Allergy and Atopic Disorders

A considerable number of epidemiological studies link early antibiotic exposures, especially multiple courses, to atopic diseases later in life (Droste et al., 2000; Farooqi and Hopkin, 1998; Johnson et al., 2005; McKeever et al., 2002; Ong et al., 2014; Wickens et al., 1999). As mentioned previously, normal development of the immune system is dependent on key members of the gut microbiome for the development of regulatory components of the immune system as well as maintaining homeostasis at the gut epithelium. Allergic and atopic disorders are primarily caused by impaired components of the adaptive immune system that rely largely on the gut microbiome, for example B cell maturity (Lundell et al., 2014) and regulatory T cell differentiation and expansion (Atarashi et al., 2013). Distinct compositions of infant gut microbiomes have been associated with the development of atopic diseases later in life (Abrahamsson et al., 2012; Atarashi et al., 2013; Bisgaard et al., 2011; Björkstén et al., 2001a; Kalliomäki et al., 2001); and therefore it is conceivable that early exposure to antibiotics, especially

broad-spectrum antibiotics, could be responsible for shaping the gut microbiome towards predisposition to allergy and atopic diseases. We hypothesize that two dysbiosis types may be responsible for allergy and atopic diseases: loss of keystone taxa and blooms of pathogens and pathobionts. Evidence for loss of keystone taxa has been shown in mouse studies, where antibiotic exposure led to changes in the gut microbiome, which eventually impacted the immune system. Reductions in regulatory T cell counts (Russell et al., 2012), and increases in serum IgE concentrations and basophil-associated TH2 cell responses (Hill et al., 2012) were observed with onset of the allergic disease phenotype. These observations agree with previous studies that found that an overabundance of IgE and the cytokine IL-4, produced by TH2 cells, are associated with allergies (Haas et al., 1999; Jujo et al., 1992). Another study found that antibiotics given to neonatal mice reduced the abundance of Clostridia and as a result induced food allergies; clostridia colonization is important for stimulating IL-22 production to prevent food antigens from crossing the gut epithelium (Stefka et al., 2014). Microbial taxa considered important for immune development may differ from one developmental stage to the next, therefore warranting further investigation into the importance of timing of antibiotic exposure in atopic disease. Although some antibiotic exposures may only create short-term dysbiosis and eventually allow the microbiome to recover, if the period of dysbiosis coincides with critical developmental time points, there is potential for long-term impact on immune health. Several studies have indicated the first six months of life as the most critical for immune development (Prescott et al., 1999; Rautava et al., 2004; Van Der Velden et al., 2001), suggesting the importance of host-microbiome interactions during this time. Germ-free mice have been shown to develop immune defenses against allergic asthma if colonized as neonates, but not if colonized in adulthood (Olszak et al., 2012). Similarly, *Helicobacter pylori* colonization in neonatal mice stomachs provided increased protection against asthma, compared to adult colonization (Arnold et al., 2011). Furthermore, Russell et al. induced asthma in mice with early antibiotic exposure but failed to reproduce the same phenotype with antibiotic exposure in adult mice (Russell et al., 2012). These studies suggest that antibiotic exposure during this critical window of development may have the most pronounced and long-lasting consequences.

In addition to the loss of keystone taxa, antibiotic exposure commonly results in an immediate reduction of biodiversity which may allow for unusual blooms of rare members of the gut microbiome. Blooms of certain strains of Clostridia, despite the importance of this taxa of bacteria in immune development, may actually contribute to atopic disease (Penders et al., 2013). Additionally, severe dysbiosis in a developing neonatal gut may allow for bacterial translocation of commensals and hence the development of systemic antibodies against these otherwise innocuous microbes. As seen in Crohn's disease (Adams et al., 2008), it is highly plausible that inappropriate immune responses against commensals could also lead to hypersensitivity to common antigens, eventually leading to allergy and atopic diseases.

Autoimmune diseases

Although autoimmune diseases such as type 1 diabetes, rheumatoid arthritis, and multiple sclerosis have a large genetic component, the gut microbiome has recently been found to be a potential major mediator of these diseases (Brown et al., 2011; Cani et al., 2008; Giongo et al., 2011; Lee and Mazmanian, 2010; Sellitto et al., 2012; Vaarala et al., 2008; Valladares et al., 2010; Wen et al., 2008). Autoimmune diseases result from an improperly developed immune system, which is, in part, mediated by the gut microbiome; there is evidence that germ-free mice are incapable of developing rheumatoid arthritis and multiple sclerosis (Lee and Mazmanian, 2010; Wu et al., 2010). In support of the hygiene hypothesis in autoimmune disease, one study found that the incidence of diabetes in non-obese diabetic mice raised in conventional breeding environments doubled when compared to pathogen-free breeding environments (Bach, 2002), which suggests that antibiotics could exacerbate the onset of diabetes. Recent work found that the number of courses of antibiotics during childhood is associated with risk of juvenile rheumatoid arthritis (Horton et al., 2014), and the risk of inflammatory bowel disease (Hviid et al., 2011). There is also evidence that antibiotics are associated with celiac disease (Mårild et al., 2013). Studies examining the effects of antibiotic exposure on type 1 diabetes have yielded inconsistent results: one study found that

antibiotics given to non-obese diabetic mice during pregnancy modulated type 1 diabetes development in offspring (Tormo-Badia et al., 2014), but other studies found antibiotics to be protective (Brugman et al., 2006; Cani et al., 2008). There is currently limited evidence linking antibiotic exposures to autoimmune disorders, but we hypothesize that the underlying mechanisms are driven by loss of keystone taxa and blooms of pathogens and pathobionts, similar to those of allergy and atopic disorders due to the critical role of the immune system in these diseases.

Infectious diseases

Antibiotics are used to eradicate one or more bacterial taxa, therefore a temporary reduction in biodiversity is expected. Current studies report a large range of percent losses of biodiversity after antibiotic exposure (Figure 2.3), suggesting that some subjects may take longer to recover to baseline than others (Figure 2.2). The recovery period represents a vulnerable time for the host since not all members of the microbial community are present to suppress, potentially, blooms of pathogens and pathobionts, and hence prevent infection. A number of studies support this theory, showing an increased susceptibility of infection after antibiotic exposure (Croswell et al., 2009; Deshmukh et al., 2014; Lawley et al., 2008; Sekirov et al., 2008), with a number of studies highlighting the proliferation of antibiotic-resistant strains (Ayres et al., 2012; Brandl et al., 2008; Buffie et al., 2012; Donskey et al., 2000; Ubeda et al., 2010). *Clostridium difficile* infection in adults is an appropriate example of how loss of biodiversity enables blooms of pathogens in the gut. Necrotizing enterocolitis in pre-term infants has also been linked to antibiotic use prior to onset of disease (Alexander et al., 2011; Cotten et al., 2009) and the gut microbiomes of children about to succumb to necrotizing enterocolitis exhibit decreased biodiversity and blooms of Gammaproteobacteria (Mai et al., 2011a; Wang et al., 2009). Although pre-term infants have a distinct set of health risks, this mode of infection can be extended to other disease agents in full-term infants as well (Figure 2.2b). This need for ecological checks and balances in the gut microbial community extends beyond its bacterial members; antibiotic-induced dysbiosis has been shown to impair innate antiviral immunity against

the influenza virus (Abt et al., 2012) as well as to enable blooms of opportunistic fungi, such as *Candida albicans (Noverr et al., 2004; Seelig, 1966; Sharp, 1954; Torack, 1957)*. Longer-duration antibiotic therapy appears to be correlated with length of recovery period (Fouhy et al., 2012), which also increases the risk of infection (Alexander et al., 2011). Identifying when a microbiome is fully recovered will be challenging given the inter-individual deviations of the adult gut microbiome, and will be even more difficult with the highly variable, developing infant gut microbiome. Lawley et al. found that mice exposed to antibiotics still exhibited increased colonization of *Salmonella* Typhimurium despite recovery of bacterial counts (Lawley et al., 2008), suggesting that not only is microbiome recovery challenging to define, but that current methods for measuring biodiversity may be insufficient for assessing infection risk. Although previous studies have focused on short-term risks for infection, it also is plausible that antibiotic exposure could lead to an incomplete, yet stable and permanent, recovery of the microbiome (Dethlefsen et al., 2008), potentially predisposing the recipient to infectious disease later in life.

Future directions

The model presented here links together the existing epidemiological and mechanistic studies on antibiotics and various gut-mediated disease outcomes. Large, integrated studies designed to focus on short- and long-term impact of antibiotics, both in terms of microbiome composition and in terms of disease risk, with careful consideration of the factors presented here, will be critical as we move toward an increased understanding of related disease etiologies. Such studies will enable important applications, such as the development of diagnostic tools to discover complex microbial biomarkers for dysbiosis risk. To demonstrate the potential importance, using a machine learning model trained on existing data (Knights et al., 2011a; Yatsunenko et al., 2012), we developed a Microbiome Maturity Index capable of accurately predicting the age of healthy infant gut microbiomes within 1.3 months (standard error) (Figure 2.4). Similar models have already been shown to be successful in identifying dysbiosis; Subramanian et al. used similar methods and found that children with severe acute malnutrition had gut

microbiomes that were significantly immature compared to healthy children (Subramanian et al., 2014). There is enormous potential for the microbiome field to revolutionize diagnostics and therapeutics, yet published human infant studies have not been designed to infer causality. Establishment of a large and diverse baseline cohort to define healthy development of the infant microbiome in presence and absence of perturbation by caesarian delivery, breast-feeding alternatives, and antibiotic usage is essential to refine our understanding of "normal development" so that pediatric dysbiosis can be identified robustly. Additionally, longitudinal and cross-sectional studies assessing the short-term, mechanistic, and longer-term health impact of antibiotics will be necessary to advance the diagnosis, interpretation, and treatment of pediatric dysbiosis, and to provide evidence-based recommendations regarding safe practices for antibiotic usage in infants.

Chapter 3: A machine learning repository for microbiome datasets

Pajau Vangay, Ben Hillmann, and Dan Knights

The microbiome research field consists of a wide variety of complex datasets; as a result, there does not yet exist a single gold standard bioinformatics pipeline for analyzing and interpreting data generated from these studies. In preparation for our human study in Chapter 5, we validated a series of statistical models and computational tools against published datasets to measure effect sizes and to select methods that were most robust. As a result of these efforts, we generated a large set of curated microbiome datasets, which we have made publicly available for other computational scientists.

Introduction

Machine learning is widely used as a method for classification and prediction, with a growing number of applications in human health (Jordan and Mitchell, 2015). The use of machine learning in biological fields (Furey et al., 2000; Shipp et al., 2002), and more specifically the microbiome research field (Aagaard et al., 2012; Knights et al., 2011b; Smith et al., 2013; Yatsunenko et al., 2012), has grown exponentially due to the robustness of these algorithms to high dimensional data. In addition, large-scale meta-analyses often requires manual curation of metadata and standardized processing of raw sequence data, resulting in variation in chosen datasets across studies (Pasolli et al., 2016; Sze and Schloss, 2016). Unfortunately, the resulting processed data is often not shared due to the use of previously published data inherent to meta-analyses. This results in inefficiencies in the manual steps of metadata curation. In addition, microbiome research data can be challenging to access for expert machine learning algorithm developers, who often do not have the domain expertise required to parse the metadata in complex studies. The University of California Irvine (UCI) Machine Learning Repository (Asuncion and Newman, 2007) revolutionized machine learning methods development by giving developers access to many curated datasets; its widespread usage and impact can be seen from its thousands of resulting citations. Currently, we are unaware of any machine learning repository that offers access to microbiome datasets. We constructed a complementary database to address this deficiency, and hope that it will promote the development of and usage of improved machine learning methods for the microbiome community.

Workflow

We present the Microbiome Learning Repo (ML Repo), a repository of 33 curated classification and regression tasks using human microbiome data. Our 33 tasks are curated from 15 publicly available human microbiome datasets, which include 12 amplicon-based and 3 shotgun sequencing datasets (Table 3.1). These datasets vary across sequencing technology platforms, 16s hypervariable regions, and study design, in order to help developer ensure robustness of algorithms across data types. We streamlined the microbiome data using a single post-processing workflow (Figure 3.1A). We downloaded trimmed and quality filtered sequencing reads for n=8 datasets from QIITA (Qiita Development Team), and raw sequences for n=7 datasets from public repositories. We preprocessed raw sequences using SHI7 (Al-Ghalith et al., 2018) or QIIME (Caporaso et al., 2010) according to individual technologies and characteristics of each study. Full details regarding the data preprocessing are provided for each data set in the repo. We picked Operational Taxonomic Units (OTUs) from all quality filtered sequences using a closed-reference method with the BURST (Al-Ghalith and Knights, 2017) aligner against both the NCBI RefSeq 16S ribosomal RNA project (O'Leary et al., 2016) and the Greengenes 97 database (McDonald et al., 2012). Samples with depths lower than 1000 sequences per sample were dropped for n=10 datasets, while we applied a lower threshold of 100 sequences per sample for n=5 datasets which had lower expected bacterial load. As a result, for each dataset we generated RefSeq-based OTU and taxa abundance counts, and Greengenes-based OTU and taxa abundance counts. We excluded additional post-processing filtering and normalization steps so that these parameters can be included in future benchmarking use cases as needed. We also limit our data to OTU and taxa tables as other metrics such as alpha and beta diversity can be subsequently generated as needed.

Sample metadata from individual studies were manually curated to generate viable prediction tasks. When available, published study exclusion criteria was applied accordingly and confounders were removed by dropping samples or stratification. Studies that were cross-sectional by design but contained several samples per subject were

filtered to contain one sample per subject. Well-known confounders, such as geography, were accounted for when constructing prediction tasks for other human-associated conditions. Longitudinal studies were reduced to single time points of interest to minimize the effect of high intra-individual similarities. Hence, each prediction task is made available as an individual, compartmentalized metadata file that contains sample identifiers, responses to predict, and optionally, confounder variables to control for. As a result, we generated 33 distinct tasks for predicting human-associated responses.

Methods

Pre-processing of sequencing reads

When available, preprocessed FASTA files were downloaded from QIITA (or previously, the QIIME database). For all other datasets, raw FASTQ files were downloaded from sources listed in Table 3.1. Sequences were trimmed and quality filtered using SHI7 (Al-Ghalith et al., 2018) or QIIME (Caporaso et al., 2010). OTUs were picked from processed FASTA files using BURST (Al-Ghalith and Knights, 2017) with Greengenes (McDonald et al., 2012) 97 or the NCBI RefSeq Targeted Loci Project 16s project (O'Leary et al., 2016) (accessed on 17-07-04). Samples with sequencing depth lower than 1000 sequences per sample were dropped for all studies, except for five datasets (David et al., 2014; Gevers et al., 2014; Human Microbiome Project Consortium, 2012b; Kostic et al., 2012; Turnbaugh et al., 2009b), where the minimum threshold was 100 sequences per sample.

Selection of classification tasks

Classification tasks were selected based on reported study results, biologically relevant high-level phenotypes, and sufficient sample sizes. Original metadata files and research methods were rigorously and manually curated in order to subset samples with minimal confounders. For confounders that were inherent to the study, we include an additional variable to control for in the task metadata files. Presence of control variables can be found by examining "control_vars" in the Tasks table.

Website generation

Website templating was developed using Jinja2 (Ronacher, 2017) and custom Python scripts. Individual webpages were generated by iterating through items in the Tasks and Datasets tables, and dynamically populating templates in order to generate individual Markdown (Gruber et al., 2004) pages. The resulting Markdown pages are hosted as GitHub Pages.

Case Study Benchmarking

Case study results were generated with custom R (Team and Others, 2013) scripts, which can be found in the */example* folder in the MLRepo Github repository. To compare machine learning models, we iterated through tasks with binary responses. OTU counts were converted to relative abundances, filtered at a minimum of 10% prevalence across samples, and collapsed at a complete-linkage correlation of 95%. We then constructed a 5-fold cross-validation for tasks containing more than 100 samples, or a leave-one-out cross-validation for tasks with smaller sample sizes. For n-fold cross validation, samples were assigned to folds such that classes were equally balanced within each fold (e.g. if our task contained 40% healthy and 60% diseased samples, our folds would also be selected to represent this distribution). For tasks that contained control variables, we selected folds such that samples with the same control variable value were contained within the same fold. For example, for a task dataset containing matching stool and oral samples from subjects, the Subject Identifier would be listed as the control variable and we should assign samples to folds such that all samples from a specific subject were contained within a fold. This step is crucial to avoid biasing or overfitting the training model; test folds should contain not only new samples, but also samples that are independent from those in the training set. Models were constructed using the 'caret' package (Kuhn and Others, 2008). This process was bootstrapped 100 times, and the mean class probabilities were used to calculate the resulting AUCs and ROCs. To compare classification accuracies using different reference databases, we used a similar procedure but held the model constant and predicted using different base OTU tables. This model enables comparison of a myriad of machine learning models available in the

'caret' package, and can be easily expanded to compare different OTU-picking algorithms, or normalization and filtering techniques.

Publicly available web-based interface

We expect two types of users: (1) machine-learning algorithm developers with limited knowledge of microbiome study designs and (2) microbiome researchers interested in obtaining additional datasets for meta-analysis. Generally, we expect that methods developers will be most interested in sweeping through the full set of prediction tasks for benchmarking, and hence would prefer to download a single compressed file containing all tasks and data. On the other hand, we expect that microbiome researchers will be more selective in downloading specific datasets and tasks depending on their research domain. Hence, researchers may prefer to browse specific details about tasks and datasets prior to downloading.

Based on these expected use cases, we created a publicly available web-interface for MLRepo hosted by GitHub Pages and available at: https://knights-lab.github.io/MLRepo. Tasks are organized by relevant response categories (Figure 3.3A). Task pages contain descriptive details such as Sample Size and Response Type that are specific to the selected prediction task, as well as links for downloading OTU tables, taxa tables, and sample metadata (Figure 3.3B). Dataset pages contain important details about the entire dataset, including links to the original research study, as well as original metadata files and quality filtered sequences (Figure 3.3C). We also provide a single compressed file containing the entire set of available tasks (OTU tables, taxa tables, and relevant metadata) for download from the main home page.

Benefits of curated microbiome-based prediction tasks

We expect MLRepo to be beneficial for both the machine-learning community as well as the microbiome research community. MLRepo will be a powerful complement to UCI's machine learning repository, as it will allow for benchmarking curated classification tasks with high-dimensional data, and hence enable the subsequent development of novel

algorithms for these complex datasets. Our streamlined approach in generating OTU and taxa tables offers a rich set of 15 datasets that microbiome researchers can use directly for further comparison with their own studies, for teaching and learning purposes, or for large meta-analyses. We expect that our provided OTU and taxa tables will also be beneficial for researchers with limited access to high-performance computing resources or bioinformatics skills necessary for processing raw sequencing data. In addition, we expect microbiome-specific methods development will also benefit from our repository prediction tasks. The subsetted samples found in each prediction task metadata file replaces the work of rigorously deciphering metadata and nuances from individual research studies. Hence, new methods, such as OTU-picking algorithms, can be evaluated not only on metrics such as speed and accuracy, but also based on overall impact to study findings.

Comparison to similar databases

Although a number of microbiome repositories exist, many are intended as data archival repositories (Hunter et al., 2014; Leinonen et al., 2011) or function as resources for aggregating across studies (Forster et al., 2016). Resources such as QIITA (Qiita Development Team) offer an extensive collection of datasets, and mock-community-based Mockrobiota (Bokulich et al., 2016a) is well-suited for benchmarking upstream methods, but neither offer support for the metadata interpretation necessary for predicting high-level phenotypes. MLRepo differs from all of these resources in that we provide well-defined tasks for predicting responses from manually curated metadata and standardized data from published microbiome research studies.

Case studies

We compare the performance of three machine learning models: random forest (Breiman, 2001), and support vector machine (Cortes and Vapnik, 1995) (SVM) with either a radial or linear kernel. Sweeping through available tasks with binary responses, we compare our models by examining receiver operating curves (ROCs) and areas under the curve (AUC) (Figure 3.4). Through direct comparison of ROCs, we can see that random forest

outperforms or ties the other two models in 21 out of the 28 tasks. The choice of kernels appears to have limited impact on overall mean accuracy, yet a linear kernel can perfectly classify penicillin-treated and vancomycin-treated mouse cecal contents when the other models could not; further examination of the microbial features in these samples may be warranted to further understand the strengths of this kernel. We also directly compared AUC and accuracy of the models across all tasks and, although not statistically significant (P=0.065 and P=0.15, respectively), found that random forest in general does better than the other two models (Figure 3.5A). Our results support the broad usage (Aagaard et al., 2012; Karlsson et al., 2013; Pasolli et al., 2016; Yatsunenko et al., 2012) and acceptance of random forest as a robust classifier (Knights et al., 2011b) with high-dimensional microbiome data.

We also used the classification tasks to assess the impact of reference database choice on classification accuracies by comparing random forest using OTUs picked with the Greengenes 97 database or the NCBI RefSeq Targeted Loci Project 16s project. We find that there is limited impact of database choice to overall classification accuracies (Figure 3.5B). This may be due to (1) large effect sizes that are driven mainly by several well-characterized bacterial taxa present in both databases (e.g. stool versus tongue samples), or (2) small effect sizes such that classification is difficult regardless of the database (e.g. male versus female stool).

Future Work

In the future, we expect and hope that the broader microbiome research community will add new datasets and prediction tasks to MLRepo. We provided instructions on our GitHub repository to guide users to create a fork from our repository, add the appropriate data and files, and update the master task and dataset lists. Users can then submit a pull request for our review, and if properly formatted, will be accepted and merged into the repository. We expect that data submissions will come from either the original researchers or those well-acquainted with the datasets, and hence will expect that sample selection and subsetting will have undergone rigorous review for prediction tasks.

Conclusions

We developed MLRepo, a repository of curated microbiome datasets made available for the computational community, and presented several case studies for how it can be a valuable resource. We hope that this repository will promote the development of and usage of improved machine learning methods for the microbiome community.

Chapter 4: U.S. immigration westernizes the human gut microbiome

Pajau Vangay, Abigail J. Johnson, Tonya Ward, Gabe Al-Ghalith, Robin Shields-Cutler, Ben Hillmann, Sarah Lucas, Lalit Beura, Emily A. Thompson, Lisa Till, Rodolfo Batres, Bwei Paw, Shannon Pergament, Pimpanitta Saenyakul, Mary Xiong, Austin Kim, Grant Kim, David Masopust, Eric Martens, Chaisiri Angkurawaranon, Rose McGready, Purna Kashyap, Kathleen A. Culhane-Pera, and Dan Knights

Although broad-spectrum antibiotics, commonly used in western medicine, can lead to imbalances in the gut microbiome, this chapter builds upon the examination of other western exposures that may be detrimental to the gut microbiome.

Introduction

Previous work has established that diet and geographical environment are two principal determinants of microbiome structure and function (De Filippo et al., 2010; Febinia, 2017; Gomez et al., 2016; Kwok et al., 2014; Obregon-Tito et al., 2015; Rothschild et al., 2018; Schnorr et al., 2014; Yatsunenko et al., 2012). Rural indigenous populations have been found to harbor substantial biodiversity in their gut microbiomes, including novel microbial taxa not found in industrialized populations (Clemente et al., 2015; Gomez et al., 2016; Obregon-Tito et al., 2015; Schnorr et al., 2014; Smits et al., 2017; Yatsunenko et al., 2012). This loss of indigenous microbes or "disappearing microbiota" (Blaser and Falkow, 2009) may be critical in explaining the rise of chronic diseases in the modern world. Despite the frequent migration of people across national borders in an increasingly interconnected world, little is known about how human migration may affect intricate human-microbe relationships.

The United States (U.S.) hosts the largest number of immigrants in the world (49.8 million or 19% of the world's total immigrants and approximately 21% of the U.S. population) (Department of Economic and Social Affairs, Population Division, 2017). Epidemiological evidence has shown that residency in the U.S. increases the risk of obesity and other chronic diseases among immigrants, with some groups experiencing up to a four-fold increase in obesity after 15 years (Bates et al., 2008; Cairney and Ostbye, 1999; Goel et al., 2004; Kaplan et al., 2004; Lauderdale and Rathouz, 2000; Walker et al., 2008). This "healthy immigrant effect" has been well-documented in Western countries (Antecol and Bedard, 2006), and is attributed to many complex, interacting factors, the effects of which vary depending on the immigrant subpopulation (Barcenas et al., 2007). Refugees, in particular, appear to be more vulnerable to rapid weight gain (Heney et al., 2014; Hervey et al., 2009), with Southeast Asian refugees exhibiting the highest average

increases in body mass index (BMI) (Careyva et al., 2015) after relocation to the U.S. Minnesota is home to the highest number of refugees per capita in the U.S., and has experienced the largest wave of refugees during the last decade (Koumpilova, 2015). The Hmong, a minority ethnic group from China who also reside in Southeast Asia, make up the largest refugee group in Minnesota (22,033 total refugees as of 2014) (Minnesota Department of Health), and also form the largest centralized Hmong community in the U.S. (70,000 total individuals) (Pfeifer and Thao, 2013). The Karen, an ethnic minority from Burma, have been arriving in large numbers in more recent years (Minnesota Department of Health). Although the Hmong and Karen originate from different countries, have distinct backgrounds, and arrived in the U.S. at different times, many in these groups share a common path through refugee camps in Thailand; they may also share similar disease disks in the U.S. Refugee children from Burma exhibited the steepest BMI increase after relocation, compared with other refugee and non-refugee children (Dawson-Hahn et al., 2016); to our knowledge, disaggregated data on long-term health changes in ethnic Karen from Burma do not yet exist. Overweight status and obesity rates are highest among Hmong compared to other Asian ethnic groups in Minnesota (Arcan et al., 2014; Franzen and Smith, 2009; Himes et al., 1992; Mulasi-Pokhriyal et al., 2012), and Western diet acculturation, previous exposure to food insecurity, and physical inactivity have been identified as contributing factors (Franzen and Smith, 2009; Mulasi-Pokhriyal et al., 2012; Smith and Franzen-Castle, 2012).

The gut microbiome plays a critical role in host metabolism and is heavily influenced by an individual's long-term diet (Hildebrandt et al., 2009; Wu et al., 2011), yet can also quickly respond to dramatic dietary changes (David et al., 2014; Turnbaugh et al., 2009a). Hence, the gut microbiome serves as an important window into the consequences of diet and lifestyle changes associated with migration. To study the short- and long-term impact of migration on the microbiome, we measured gut microbiomes and dietary intake from Hmong and Karen immigrants and refugees (henceforth referred to as immigrants) across cross-sectional and longitudinal cohorts undergoing relocation to the U.S. We characterized gut microbiome species, strains, and functional profiles among Hmong and

Karen individuals still living in Thailand and after U.S. immigration. The cohort was stratified by BMI to include cross-sectional samples from individuals with high (≥25) and low (<25) BMI in both pre- and post-immigration groups. The first-generation immigrant (foreign-born U.S. residents) group included individuals with duration of U.S. residence ranging from a few days to more than 40 years. This range allowed us to test for changes in the gut microbiome associated with long-term residence and duration of residence. We then studied second-generation (born in the U.S. to first-generation immigrants) Hmong immigrants to determine whether the effects of U.S. immigration are compounded across generations by birth in the U.S. Finally, we followed a longitudinal cohort of 19 Karen refugees for 6 months beginning immediately before or after arrival in the U.S to measure the immediate short-term effects of U.S. immigration.

Methods

Study setting, population, and recruitment.

Our inclusion criteria included individuals who were Hmong or Karen, female, at least 18 years old, and either were born and are currently living in Thailand, were born in Southeast Asia and moved to the U.S., or were born in the U.S. but whose parents were born in Southeast Asia. Our inclusion criteria for controls included Caucasian females at least 18 years of age who were born in the U.S. and whose parents and grandparents were also born in the U.S. Our exclusion criteria consisted of use of any antibiotics in the previous 6 months, current use of probiotic supplements, known presence of gastrointestinal, cancer, immunodeficiency or autoimmune disorders, adults lacking capacity to consent, or pregnancy. Additionally, control subjects could not have traveled outside of the U.S. within the last 12 months. We recruited using multiple methods which included flyers, emails, social media, oral presentations, tabling, letters followed by phone calls to West Side Community Health Services (West Side) patients who met criteria, and by word of mouth. We recruited throughout the Minneapolis-St. Paul metro area at local community centers, faith-based organizations, adult education centers, health care centers, and health fairs. We recruited in Thailand at Khun Chang Khian (KCK), a

rural Hmong village located one hour from Chiang Mai city, as well as from Mae La (ML) Camp, a Burmese refugee camp in Tak province located on the Myanmar-Thailand border (Figure 4.S1). Interested subjects were then screened and interviewed privately or as a group, as preferred by the participants. Interviews and body measurements were conducted by trained Hmong and Karen community researchers and a graduate student researcher. This study was approved for human subject research by the University of Minnesota Institutional Review Board (1510S79446), and the Thailand-based portion of the study was additionally approved for human subject research by the Chiang Mai University Institutional Review Board (475/2015) and the Chiang Mai Public Health Office (0032.002/9930).

Application of Community-based Participatory Action Research methods

This project used a community-based participatory action research (CBPAR) approach, with a multidisciplinary team composed of academic researchers, Hmong and Karen community researchers, and staff from the Somali, Latino and Hmong Partnership for Health and Wellness (SoLaHmo). SoLaHmo is a multi-ethnic, community-driven CBPAR program of West Side Community Health Services, Inc, whose mission is to build upon the unique cultural strengths of ethnic communities to promote health and wellness through research, education and policy. All SoLaHmo members are trained in qualitative research processes using a previously developed training curriculum (Allen et al., 2011). In addition, all phases of our project were further guided by community advisory boards (CABs) composed of Hmong and Karen health professionals and community experts. The study design, recruitment methods and strategies, and dissemination of results were developed in partnership with both academic and community researchers, and through multiple discussions with the CABs. Based on insight from the Hmong CAB and research team members that substantially more Hmong women than men were relocating to U.S. in recent years, we limited our study to women. In Thailand, we used a modified CPBAR approach in that Thai community researchers were members of the communities that we worked with, and were trained with qualitative research methods, recruitment, and sample and data collection, but were not directly

involved with study design. We note that Hmong refugee camps have long been closed (Bureau of Population, Refugees and Migration, 2004), hence Hmong in Khun Chang Khian are not refugees but serve as acceptable pre-immigration representatives available for US-based Hmong.

Cross-sectional specimen and data collection, U.S.

Research team members obtained informed consent and conducted interviews in the participants' preferred languages (English, Hmong, or Karen), and recorded participants' responses onto an English paper survey (Appendix A). Weights were measured using standard electronic scales, heights were measured against a wall using a pre-positioned measuring tape, and waist circumferences were measured with a tape measure at the uppermost lateral border of the iliac crest (Center For Disease Control, 2014). 24-hour dietary recalls were conducted using a multiple pass system (Tippett et al., 1999) with food models and measuring cups and spoons for portion size estimations. Participants were provided with a stool collection kit and instructions describing how to collect a stool sample. Stool samples were collected into preservative (see below) and were either returned to the research staff by mail or were stored at room temperature for up to 5 days before they were collected by the research team.

Longitudinal specimen and data collection, U.S.

Procedures for consent, interviews, anthropometrics, and stool sampling were as described above for the cross-sectional specimen and data collection. Once per month over six months, 24-hour dietary recalls were conducted as described previously. Month 1 and 6 samples were stored in a home freezer and picked up within 24 hours of stool collection. These samples were transported with an ice pack and immediately placed in a -80C freezer. Month 2-5 samples were stored in preservative (see below), mailed to the research team in prepaid mailers at room temperature, and placed in a -80C freezer upon receipt.

Specimen and data collection, Thailand

Procedures for consent, interviews, anthropometrics, and stool sampling were as described above for the cross-sectional specimen and data collection. 24-hour dietary recalls and sample collections were conducted as described previously. Stool samples from KCK were transported on dry ice then placed in a -20C freezer for 2 days then transferred to a -80C freezer. Stool samples from ML were placed in a -20C freezer for up to 8 hours then transferred to a -80C freezer. All samples collected in Thailand were shipped overnight on dry ice from Thailand to the U.S., and stored in a -80C freezer in the U.S.

Stool sample collection

Research team members instructed participants in stool collection, using an instructional video, written visual instructions, and verbal reinforcement. Participants placed their stool sample onto a FecesCatcher (Tag Hemi VOF) and 1 gram was collected using a sterile swab into a 1.5 ml cryogenic tube pre-filled with 900 ul of RNALater™ and mixed thoroughly, (Appendix B). Larger samples (longitudinal first and last month samples) were collected using a Sarstedt Inc 80.9924.014/CS500 tube and scoop without mixing or RNALater. Large samples collected in the U.S. were aliquoted into 1.5 ml tubes with and without 50% glycerol upon arrival, and stored at -80C. Large samples collected in Thailand were stored at -80C until arrival to the U.S., at which point they were thawed over ice, aliquoted, and stored in the same manner.

Dietary data processing workflow

De-identified survey data was entered into an electronic spreadsheet. Foods and portions from 24-hour dietary recalls were entered into the USDA SuperTracker system (Britten, 2013). Foods that were not found in the USDA database were researched individually (Speek et al., 1991) for macronutrient content and entered in as custom foods. SuperTracker macronutrient and food grouping summaries, as well as foods and their respective portions were downloaded directly from the SuperTracker website, or using custom Python (van Rossum and Drake, 2011) scripts. Foods and portions were mapped

to the SuperTracker and USDA databases to obtain respective food and portion identification numbers; food and portion identification numbers were used in tree-based food analysis. Custom foods not in the USDA database were manually assigned appropriate existing or new food identification numbers by group consensus. Micronutrients were excluded from dietary analyses due to the high number of custom foods with limited information on micronutrients. Food tree visualizations were generated with Graphlan (Asnicar et al., 2015). Dietary record and food item associations were generated using custom scripts, then visualized in Cytoscape (Shannon et al., 2003).

16S sample processing and sequencing

All fecal samples were submitted to the UMN Genomics Center for DNA extraction, amplification, and sequencing. 16S ribosomal rRNA gene sequences were extracted and amplified following the UMGC-developed protocol (Gohl et al., 2016). We trimmed and processed all marker-gene sequencing data for quality using SHI7 (Al-Ghalith et al., 2018) and picked *de novo* operational-taxonomic units (OTUs) as follows. We first filtered for reads with at least 100 exact duplicates as representative sequences, and assigned taxonomy by alignment at 0% to the NCBI RefSeq 16s reference database (O'Leary et al., 2016) using the BURST (Al-Ghalith and Knights, 2017) OTU-picking algorithm in CAPITALIST mode, which ensures optimal alignment of sequences and minimizes the set of aligned reference genomes. All original sequences were then re-aligned with BURST (Al-Ghalith and Knights, 2017) in CAPITALIST mode at 98% identity against this representative set, resulting in 93.54% of all available sequences aligned. Singleton OTUs and samples with depth less than 2,143 were removed using the Quantitative Insights Into Microbial Ecology (QIIME) software package (Caporaso et al., 2010). Using QIIME, we measured within-sample biodiversity (alpha diversity) with rarefied OTU tables (at 2,143 sequences/sample) using whole-tree phylogenetic diversity (Faith, 1992) and a custom generated phylogeny constructed with the representative sequences using aKronyMer (Al-Ghalith and Knights, 2018). To quantify differences in composition between subjects, we calculated the phylogeny-based UniFrac distance (Lozupone et al., 2011) between all pairs of samples. To visualize between-subject

differences (beta diversity) and to obtain principal components for subsequent statistical testing, we performed dimensionality reduction using principal coordinates analysis (Caporaso et al., 2010). Aitchison's distances were calculated by first imputing zeros from an abundance OTU table, then applying a centered log ratio transform using the *robCompositions* R package (Pawlowsky-Glahn and Buccianti, 2011). To enable tests for shifts in the relative abundances of *Bacteroides* and *Prevotella*, we collapsed the reference-based OTUs according to taxonomy at the genus level.

Deep shotgun metagenomic sample processing and sequencing

Shotgun DNA sequencing was performed on the Illumina HiSeq platform. All fecal samples were submitted to the UMN Genomics Center for DNA extraction, amplification, and sequencing. Amplification, quantification, and normalization of extracted DNA was performed using the Illumina NeoPrep Library System. A HiSeq 2x125 cycle v4 kit was used to sequence samples. Sequences were identified at the species level via genomic alignment against a custom database created from aligning human samples from various public datasets against the comprehensive NCBI RefSeq database (Tatusova et al., 2013) release 87, and all matched bacterial species, as well as all species in matched representative genera, were included from NCBI RefSeq database (Tatusova et al., 2013) release 87. Genome coverage estimates were calculated using the *bcov* utility from BURST (Al-Ghalith and Knights, 2017). Functional annotations were obtained using the HUMAnN2 (Abubucker et al., 2012) pipeline with UniRef50 (Suzek et al., 2015). Resulting functional pathways were mapped to and colored by the top level categories of the MetaCyc (Caspi et al., 2008) ontology. CAzyme annotations were obtained using metaSPAdes (Nurk et al., 2017), filtered for scaffolds with minimum 1000 bp, then further processed with Prokka (Seemann, 2014), dbCAN (Yin et al., 2012) with E-value < 1e−5, and the CAZy database (Lombard et al., 2014).

Food-Microbiome Procrustes distance associations

Procrustes: P-values are from the `vegan` implementation in function `protest ()` with 999 permutations (performed for each of the permuted data structures). Distances plotted

are the Euclidean distances between food and diet samples after rotation of distance matrices with Procrustes. The representative Procrustes plot with permuted labels was chosen based on median overall Procrustes distance (M12 = square-root of 1 minus the sum of squares) out of 10 permuted Procrustes rotations.

Predicted biosynthetic gene clusters

Raw shotgun reads were quality controlled using SHI7 (Al-Ghalith et al., 2018) and aligned using BURST (Al-Ghalith and Knights, 2017) at 95% identity against a reference database of 21,186 putative biosynthetic gene clusters (BGC) predicted by antiSMASH or deposited in the MIBiG database (Blin et al., 2017; Weber et al., 2015). The per-sample metagenomic coverage of each BGC was calculated using in-house Python and R code and filtered to pathways with a ratio of actual coverage to expected coverage (expected coverage probability is defined as 1-exp ((N * L_read) ⁄ L_BGC), where N = number of reads, L_read = median read length, and L_BGC = BGC sequence length) of at least 0.75. Differentiating BGCs were identified by comparing BGC presence/absence frequency between the treatment groups using Fisher's exact test with FDR correction at $q < 0.15$. To collapse homologous BGCs we used custom Python and C code to hierarchically cluster the pathways based on amino acid identity and open reading frame composition (Rashidi et al., 2018; Shields-Cutler et al., 2018; Zhang et al., 2018). Cluster annotations and taxonomic assignments were derived from their antiSMASH references.

Mouse experiment specimen and data collection

We thawed 2 mL of previously collected frozen human stool over ice, then added it to a 15ml conical tube containing 3 ml of pre-reduced PBS inside of a COY anaerobic chamber and vortexed for 1 minute. 4-8 week-old germ-free C57BL/6 female mice were fasted overnight then removed from germ-free isolators and gavaged with 300 uL of prepared donor material. Fasting blood glucose measurements were taken with a cheek bleed using a StatStrip Xpress glucometer, and mice were ear punched for identification. Mice were subsequently placed in cages with Sani-chips™and Crink-l'Nest™ bedding, which have been autoclaved at 128C for 30 minutes with a 15 minute dry cycle, and with

nanopure drinking water which had been autoclaved in 1L pyrex bottles at 121º for 2 hours. Prior to experiment start, mice were fed a standard autoclaved chow (LabDiet 5K67), and immediately after humanization, mice were placed on irradiated LabDiet 5061 or Harlan TD.86489. Cages were sealed and place into an Arrowmight Maxi Seal IVC System, housed within a germ-free facility. Cages were changed, mice and chow were weighed, and pellets were collected every two weeks using sterile handling methods throughout the duration of the study. At study endpoints, mice body composition analysis was performed as previously described (Schafer et al., 2016), mice were fasted overnight, then euthanized using CO_2 asphyxiation. Heart blood was collected immediately post-euthanasia for fasting blood glucose tests.

Histology

The duodenum was collected into 10 mL 10% formalin and sent to the UMN Comparative Pathology Shared Resource, where after 24 hours, the tissue was transferred to 70% ethanol. 4 µm formalin-fixed, paraffin-embedded sections of tissue were deparaffinized and rehydrated, followed by Hematoxylin and Eosin staining. Images were taken using a NIKON Eclipse E 800M microscope with 10x objective, and measurements were made using the NIS Basic Program.

Cell isolation, staining and flow cytometry

A two-inch section of the jejunum was collected in 3% PFA, and the remainder of the small intestine was emptied of its contents and stored in 35 mL of CMF. Tubes were stored in ice during transport. Leucocyte isolation from small intestine was performed as previously described (Thompson et al., 2016). Briefly, for isolation of intra-epithelial lymphocytes (IELs) from small intestine, fecal contents were removed, and Peyer's patches were excised and the gut was cut longitudinally and then into 1 cm pieces. Intestine pieces were incubated in 10% 1X HBSS/HEPES bicarbonate containing 15.4 mg/100 ml dithioerythritol (30 min at 37°C, 450 rpm) to extract IEL. After separating IELs, gut pieces were treated further with 100 U/ml type I collagenase (Worthington Biochemical, Lakewood, NJ, USA) for lamina propria lymphocyte (LPL) isolation.

Single cell suspensions were surface-stained with antibodies to detect various leukocyte populations in gut. The stained samples were acquired using LSR Fortessa flow cytometer (BD) and analyzed with FlowJo software (Treestar).

Other limitations

Although this study will provide a cross-sectional perspective of the change in gut microbiomes before, immediately after, and long after migration, it is important to note that not all subjects recruited will have followed the same migration path. For example, the large wave of Hmong who arrived to the U.S. approximately 10-15 years ago are primarily from Wat Tham Krabok, a temple that became home to over 30,000 Hmong refugees after the last official Hmong refugee camp was closed in Thailand in 1992. This large refuge has since been closed and evacuated, therefore making it infeasible for us to collect representative pre-immigration samples for this group. We mitigate this by sampling from a rural Hmong village in northern Thailand, where resources are limited and living conditions are more similar to this refuge than an urban community.

Results

Assembly of a multi-generational Asian American immigrant cohort

We recruited 514 healthy Hmong and Karen female individuals (aged 18-78, see Methods for full exclusion criteria) who either (1) were living in Thailand (HmongThai, KarenThai; n = 179), (2) were born in Southeast Asia and had moved to the U.S. (Hmong1st, Karen1st; n = 281), or (3) were born in the U.S. and whose parents were born in Southeast Asia (Hmong2nd; n = 54) (Figure 4.4.1A). We also recruited healthy Caucasian American female individuals to serve as U.S. controls (Controls; n = 36) (Figure 4.4.1A). We restricted the study population to females because the majority of recently arrived Hmong immigrants were projected to be female. Participants in each sample group were recruited into lean or overweight/obese body mass index (BMI) class stratifications (BMI < 25 or BMI ≥ 25, respectively), with the intent of obtaining similar sample sizes within each group (Table 4.1). Between February 2016 and March 2017, we

recruited and collected samples from eligible individuals throughout the Minneapolis - St. Paul metropolitan area in Minnesota, and at two locations in Thailand: a rural village in Chiang Mai province (Khun Chang Khian), and a refugee camp in Tak province (Mae La) (Figure 4.S1).

During face-to-face enrollment, bilingual-bicultural research team members collected migration and medical histories (Table 4.2), anthropometrics (weight, height, waist circumference), 24-hour dietary recalls, and a single stool sample. A single stool sample was collected for 16S rRNA and metagenomic profiling of the gut microbiome. Karen participants who identified themselves as having arrived in the U.S. within 2 months were invited to participate in a longitudinal sub-study, in which 24-hour dietary recalls and stool samples were collected monthly for 6 months (Figure 4.4.1A). As a result, we enrolled 19 individuals with longitudinal samples over their first 6 to 9 months of residency in the U.S.. This group included 6 individuals whose initial samples were collected in a refugee camp in Thailand prior to relocation. As a result of our recruitment efforts, we collected a total of 673 stool samples comprised of 531 single and 142 multiple time point collections. Because we stratified recruitment by a BMI threshold of 25, examining the ratio of obese (BMI $\geq$ 30) to overweight (BMI between 25 and 29.9) individuals provides a conservative estimate of the prevalence of obesity across groups. Consistent with the previously observed high rate of obesity in U.S. immigrants (see Introduction), we see that obesity prevalence increases after a decade in the U.S. (Figure 4.4.1B).

To understand whether or not the observed changes in the gut were driven by dietary intake, we collected 24-hour dietary recalls from all participants, and analyzed macronutrient content using the United States Department of Agriculture (USDA) SuperTracker food record system (Britten, 2013). A total of 224 unique foods were not found in the SuperTracker food database, hence additional information was supplemented from the more comprehensive USDA Food Composition Databases (United States Department of Agriculture Agricultural Research Service) and published literature. We

considered the relatedness of individual foods when assessing the similarity of dietary profiles across individuals. This approach relied on a hierarchical format of unique food codes that were derived from the USDA's Food Nutrient and Database for Dietary Studies (FNDDS). These hierarchical food codes allowed individual foods to be categorized into a phenetic tree format where more closely related foods are grouped together (Figure 4.4.1C). These groupings then allowed us to share statistical strength across closely related foods to complement dietary analysis of macronutrients, much in the way that phylogenetic beta diversity analysis complements taxonomy-based profiles of microbiomes. Foods reported by participants that were not found in any USDA database (n = 72, Table 4.3) were manually assigned new food codes and inserted into the hierarchical food taxonomy, allowing us to account for all foods reported by all participants. Using this hierarchical food tree, we observe a stark difference in the overall variety of foods eaten by Hmong in Thailand and second-generation Hmong, despite similar group sample sizes and age range (Figure 4.4.1C).

U.S. immigration is associated with loss of gut microbes

We performed amplicon-based sequencing of the 16S rRNA gene V4 region on 550 stool samples (one sample per participant). Principal coordinate analysis (PCoA) of unweighted UniFrac (Lozupone et al., 2011) distances reveal that Hmong and Karen ethnic groups harbor distinct microbial gut compositions regardless of country of residence, yet their microbiomes converge toward Caucasian American microbiomes after relocating to the U.S. The first two principal coordinate axes show that second-generation Hmong and Caucasian American microbiomes share nearly identical cluster centroids (Figure 4.4.2A), although Caucasian American microbiomes have lower inter-individual variation. We also find that both diversity and richness is highest in microbiomes from the groups in Thailand and decreases with generations in the U.S. (Figure 4.4.2B). As with other studies (Sze and Schloss, 2016; Turnbaugh and Gordon, 2009), we found that lower phylogenetic richness is associated with obesity across all of our groups, yet the median richness of obese individuals in Thailand is still higher than the median richness of any lean group in the U.S. (Figure 4.4.2B). These trends persist

after stratification by ethnicity (Figure 4.S2). These findings suggest that both obesity and residency in the U.S. are independently associated with loss of gut biodiversity. Furthermore, we observed a dramatic and systematic loss of native bacterial operational taxonomic units (OTUs) among first-generation Hmong (Figure 4.4.2C). Although 7 of the 10 most prevalent OTUs found in HmongThai are also found at similar levels in Hmong1st, others such as *otu1812* (*Faecalibacterium prausnitzii*) incur a 45% loss in prevalence (Fisher's exact test, FDR-corrected q = 3.05E-14) (Table 4.4). We found 28 OTUs (10.5% of all OTUs in 75% of HmongThai) that incurred more than a 50% loss in prevalence among first-generation Hmong, and more than half of them belong to the *Prevotella* genus (Table 4.4).

Bacteroides strains displace Prevotella strains across generations in the U.S.

The severe loss of overall biodiversity and native bacterial members in first-generation immigrants is caused by a profound taxonomic shift in the gut microbiome. We examined the Western-associated *Bacteroides* and non-Western-associated *Prevotella* and found a displacement of *Prevotella* with *Bacteroides* across generations in the U.S. (Figure 4.4.3A). Not surprisingly, the ratio of *Bacteroides* to *Prevotella* is lowest in Thailand-resident individuals, and highest in U.S.-born Caucasian Controls. The ratio of *Bacteroides* to *Prevotella* in first-generation Karen, first-generation Hmong, and second-generation Hmong increase in a stepwise fashion. This progression corresponds with the time that these groups have spent in the U.S.

Using deep shotgun metagenomics on 55 samples (mean 22,406,875 reads/sample) from Hmong in Thailand, newly arrived Karen, long-term resident Hmong (who lived in the U.S. for more than 30 years), and Controls, we profiled strain-level variation in *Bacteroides* and *Prevotella*. We aligned shotgun metagenomic sequences against a custom database that included 256 *Bacteroides* genomes and 153 *Prevotella* genomes. To minimize spurious genome alignments, we profiled only *Bacteroides* and *Prevotella* strains with a minimum genome coverage of 50% within at least one sample. We found that U.S. Controls have varied *Bacteroides* strain profiles, and those with *Prevotella* tend

to be limited to a single strain of *P. coprii* (Figure 4.4.3B). Conversely, Thailand-based individuals carry up to 4 strains of *Prevotella*, with low abundances and generally low genomic coverage of *Bacteroides* strains. There may be limited characterization of *Bacteroides* strains specific to Thailand residents available in the current reference genome databases, which could cause us to observe lower coverage of Bacteroides strains in those subjects. Long-term U.S.-resident Hmong displayed an intermediate profile, carrying a variety of *Bacteroides* strains and, in several individuals, multiple *Prevotella* strains. Our findings suggest that the increase in *Bacteroides* after moving to the U.S. is driven by both an expansion of pre-existing low-abundance strains, as there is diverse *Bacteroides* strain prevalence within the Thai-resident groups, and the acquisition of new U.S.-based strains shared with Control subjects.

U.S. immigrants lose enzymes associated with plant fiber degradation

We also profiled microbial functional pathways in our shotgun metagenomics samples (Abubucker et al., 2012) (Figure 4.S3A). In long-term-resident first-generation Hmong, we observed increases in relative abundances in sucrose degradation, glycerol degradation, glucose/xylose degradation, and glucose fermentation to lactate, suggesting that Hmong who have lived in the U.S. more than 30 years may consume more sugary foods. In Hmong in Thailand, we found an enrichment of pathways relating to the degradation of complex carbohydrates, which includes mannose biosynthesis, mannan degradation, and starch degradation (Flint et al., 2012). In order to better understand the potential substrates degraded by these pathways lost in U.S. immigrants, we assembled the deep shotgun metagenomic data into scaffolds (insert metaquast results), and annotated carbohydrate-degrading enzymes (CAZymes) (Lombard et al., 2014; Yin et al., 2012), and found that observed shifts in strain-level composition and functional pathways were accompanied by significant shifts in several types of CAZymes. We observed differential abundance of 58 CAZymes across the HmongThai, Hmong1st, and Control groups (Mann Whitney U, FDR $q < 0.05$, Figure 4.4.3C). These shifts included three beta-glucan-targeting glycoside hydrolases (GH17, GH64, GH87) that were almost completely lost from the Thailand-based group to the U.S.-based groups. This loss could

be associated with decreased ability of the microbiota to degrade certain dietary fibers. A loss of GH5 and GH26 glycoside hydrolases from HmongThai to Hmong1st and U.S. controls indicates a loss of cellulose, beta-mannan and possible xyloglucan degradative potential. Beta-mannan is produced by fungi as part of the cell wall, suggesting a lower load of fungal gut microbiota in post-immigration individuals (Engel et al., 2012). Cellulose and xyloglucan are plant cell-wall components, so the loss of glycoside hydrolases for degrading these is another indication that the microbiota of post-immigration individuals have lost some of their ability to degrade plant fibers (El Kaoutari et al., 2013).

Dietary acculturation partly explains microbiome acculturation

Across sample groups, we observed significant differences in the consumption of macronutrients commonly associated with a Western diet: sugars, fats, and protein. We find that consumption of sugars and fats are associated most significantly with residency in the U.S., and that protein consumption is highest among first- and second-generation Hmong (Figure 4.4A, Figure 4.S4). These findings suggest that new arrivals may have a higher preference towards high-sugar, high-fat foods, such as processed snacks, and that it takes longer to acculturate to eating a high-protein diet. Interestingly, total calorie consumption is similarly high among Karen in Thailand and U.S.-based Controls (Figure 4.4A).

Our use of a hierarchical food tree enabled approximate comparisons of common American foods to non-American foods, and as a result, to apply tree-based ecological analysis methods to the diet profiles of all subjects. PCoA of unweighted UniFrac (Lozupone et al., 2011) of interindividual dietary intake distances reveal distinct separation by sample group, and a gradient of increasing dietary acculturation along the first principal coordinate (Figure 4.4B). Shifts toward positive values of the first principal coordinate are driven by decreased consumption of rice, cooked and raw vegetables, and fish, and increased consumption of fruits, milk, coffee, breads, pastas, soft drinks and juices, processed meats, cookies, carrots, roasted beef products, and chicken (Table 4.6).

First- and second-generation Hmong shared similar food choices (Figure 4.4B) when considering hierarchical relationships of foods, which cannot be determined with macronutrients alone; these diet-based clusters are notably different from the groupings seen in the microbiome-based PCoA, where second-generation Hmong instead clustered closely with Controls.

A bipartite network of participants and their food choices show how most individual food choices are shared between groups (Figure 4.S5A). Sample groups are difficult to delineate because of the high degree of shared foods reported across all Hmong and Karen groups. Similar to PCoA based on food distances, we see strong overlap between KarenThai and Karen1st food choices, and observe highly individualized diets in Controls, where several participants consumed multiple foods not reported by anyone else in the study. Interestingly, the vast majority of diet records from Hmong and Karen included white rice (572 out of 630, 90.7%), compared to only 4 of the 36 Controls (11.1%) (Figure 4.S5B).

To understand the relationship between diet and microbiome compositions, we performed a permutation-based Procrustes analysis to compare distances between unweighted UniFrac diet and microbiome distance matrices and found strong similarity between variation in diets and microbiomes (P=0.001, n=999 permutations) (Figure 4.S6A). We validated this association by comparing per-sample Procrustes distances of the original distance matrices against per-sample Procrustes distances of permuted distance matrices, and found that the original matrices are more similar (P = 1e-10) (Figure 4.S6B). These Procrustes tests demonstrate that similarity of microbiome-based distances and food-based distances is significantly better than random chance. However, constrained ordination of the microbiome by the first 5 principal coordinates of the diet-based PCoA revealed that dietary variation alone explained only a fraction (16.8%) of the total microbiome variation (Figure 4.4C). Altogether, we find that although both microbiome and dietary acculturation increases with time in the U.S., diet is not the sole contributor to the observed gut microbiome changes in our cohort.

Gut biodiversity decreases according to duration of residence in the U.S.

In a PCoA with unweighted Unifrac microbiome-based distances, we find that time spent in the U.S. is significantly correlated with changes seen along the first principal coordinate (Figure 4.5A). Conversely, gut biodiversity, as measured by Faith's Phylogenetic Diversity, is negatively correlated with PC1 (Figure 4.5B). To understand the relationship between diversity and time in the U.S., we stratified our analysis by ethnic group to account for the distinct time frames of Hmong and Karen immigration (over 40 years versus 10 years). We found that gut biodiversity in first-generation Hmong significantly decreases with time in the U.S. (Figure 4.5C), while controlling for BMI. Further stratification by BMI class reveals similar trends of negative associations, but were not significant (data not shown). We also find a weak association between gut biodiversity and time spent in the U.S. in first-generation Karen (Figure 4.5C), which suggests that observable changes in biodiversity may take place after 10 years. We acknowledge that lifestyle differences between Hmong and Karen may also be contributing factors.

Prevotella displacement continues for more than one decade

Over time in the U.S., first-generation gut microbiome compositions diverge from their Thai counterparts and converge toward Caucasian Controls (Figure 4.6A). Loss of biodiversity impacts beta diversity, but our findings suggest that the contributions vary with ethnic group and timeframe. Instead, we find that the shifts in bacterial composition are largely governed by the displacement of *Prevotella* with *Bacteroides*. We observe a highly significant and strong association of time spent in the U.S. with the ratio of *Bacteroides* to *Prevotella* (Figure 4.6B), and these significant associations persist after stratification by ethnicity and within the shorter time frame of first-generation Karen (Figure 4.6B inset). These findings show that changes to the dominant members of the gut microbiome begin during the first decade of U.S. residence, and continue for longer than a decade.

Microbiome Westernization begins within 9 months after immigration

To understand whether changes in the gut microbiome can be detected immediately after relocation to the U.S., we examined the gut microbiomes of 19 newly arrived Karen in a longitudinal cohort. PCoA of the unweighted UniFrac distances of first and last month stool samples show that within six months, we can detect a significant shift in microbial composition along the first principal coordinate (Figure 4.S7). We also found that within this short time frame, all but one participant gained weight (Figure 4.7A), although weight gain after relocation is expected in Southeast Asian refugees (Careyva et al., 2015). We previously reported that protein consumption was similar between Karen in Thailand and first-generation Karen in the larger cross-sectional study (Figure 4.S4), but characterizing diets longitudinally allowed us to detect a subtle but significant increase in protein consumption after 6 months in the U.S. (Figure 4.7B). We also previously reported that Hmong in Thailand consume a limited variety of foods compared with second-generation Hmong (Figure 4.4.1C), which suggests that living in the U.S. may increase exposure to or encourage consumption of diverse foods. Instead, we found that longitudinal Karen participants reported eating fewer kinds of foods after 6 months (Figure 4.7C), which indicates that it may take a longer than half a year to acclimate to foods available in the U.S. Within this short time frame, once again we observe the displacement of *Prevotella* by *Bacteroides* (Figure 4.7D), indicating that microbiome westernization begins immediately after arrival to the U.S. Using deep shotgun metagenomics with 13 samples from 6 participants, we find that *Prevotella* and *Bacteroides* strain profiles remain largely stable over 6 months but can sometimes result in drastic changes (subject highlighted in blue, Figure 4.7D).

Our longitudinal sub-study includes six participants with baseline characterizations of their gut microbiomes prior to relocation to the U.S. (Figure 4.7F). While we found examples of disruption to the gut immediately after arrival (ID.273 and ID.304), we observed that physically relocating to the U.S. induces wide variation in gut microbial responses, including expansion of opportunistic pathogens (ID.305), gut disruption several months after arrival (ID.275), and stability (ID.274, ID.308) (Figure 4.7F).

Western diet and Western microbes lead to adverse health outcomes

In order to test whether the Westernized microbiome could induce deleterious changes in health, we performed fecal microbiota transplantation from study subjects into germ-free mice. We gavaged 4-7 week-old germ-free C57BL/6 mice (n = 38) with prepared stool samples from Thai- (T) or U.S.-based (U) donors (selected from HmongThai and Hmong2nd groups, matched on age and BMI) (Figure 4.6A). After humanization, mice were fed either a high-fiber (H) or low-fiber (L) diet and caged by diet-donor groups. At the end of 8 weeks, a subset of each group was sacrificed while remaining mice were co-housed within their diet group for an additional 2 weeks.

We performed 16s rRNA gene sequencing of collected mouse pellets and found clear separation of microbiomes by donor and diet at the end of the study (Figure 4.6B). The effect of cohousing varied depended on the diet: cohousing TL and UL resulted in an intermediate microbiome that resembled both groups, whereas cohousing TH and UH shifted both microbiomes towards the UH group.

We found that although low-fiber groups consistently consumed less chow throughout the study (Figure 4.S5), this behavior was due to their increased efficiency in metabolizing the low-fiber diet into energy (P=3.2e-05) (Figure 4.6C). As a result, the low-fiber groups exhibited increased adiposity (Figure 4.6D) and elevated blood glucose responses (P=0.013 and P=0.012, respectively) (Figure 4.6E). Examination of the ileal intraepithelial compartments revealed a donor-dependent response. We found an elevation of inflammatory cell subtypes, TCRγδ and CD8ααTCRαβ, in US-donor groups (P=0.035 and P=0.027, respectively) and observed lower levels of CD4TCRαβ in Thai-donor groups (P=0.016) (Figure 4.6F).

Discussion

This study represents the first large cohort study of the effects of migration from a non-Western country to a Western country on the gut microbiome. Leveraging both cross-sectional and longitudinal cohorts of immigrants and refugees, including pre-

immigration, first-generation immigrant, and second-generation immigrant individuals, allowed an unprecedented examination of microbiome resilience and response to migration to the U.S. In these cohorts, we observed that gut microbiome diversity, function, and strain composition are severely impacted by migration and that both short-term and long-term U.S. residence as well as being born in the U.S. shifts an individual's microbiome along an axis toward a Westernized state.

We found that U.S. immigration is associated with a loss of gut microbiome diversity. Diversity continues to decrease for at least a decade with time spent in the U.S., and is further decreased in second-generation individuals born in the U.S. We also found that U.S. immigrants undergo a marked loss of native gut microbiota strains, and begin exchanging dominant strains of *Prevotella* for dominant strains of *Bacteroides* within the first 9 months of arrival. This demonstrates that even a short period of residence in the U.S. is sufficient to induce pronounced increases, in most cases over 10-fold, in the ratio of *Bacteroides* to *Prevotella*. Our analysis using deep shotgun metagenomics demonstrated that this shift was largely due to changes in the relative abundance of extant strains in the immigrant gut microbiome rather than total novel strain acquisition. Beta diversity analysis showed that the trans-generational effects of immigration are large enough that, within one generation in the U.S., immigrant gut microbiomes become nearly indistinguishable from those of the Caucasian Controls. Metagenome assembly and functional annotation showed that the observed changes in bacterial strains were associated with dramatic post-immigration shifts in the profile of carbohydrate-active enzymes dominant in the gut microbiota, including a near-complete loss of certain beta-glucanases that may indicate loss of ability to break down specific dietary fibers.

In addition to studying immigrant microbiomes, we also performed extensive analysis and modeling of differences in dietary intake, as diet is known to be a strong driver of microbiome variation (Bokulich et al., 2016; David et al., 2014; Muegge et al., 2011). In the 24-hour diet recall data, we observed clear patterns of dietary acculturation through analysis of macronutrients as well as food choices. While we also observed a similar

trend of Westernization in microbiome and in dietary choices, we found that only a small amount of microbiome variation (16.8%) (Figure 4.S4C) in the microbiome is explained by diet.

Using humanized germ-free mice, we found that a Western-associated microbiome induces inflammation that may be deleterious to long-term health. Using flow cytometry to characterize immune cell populations in the gut intra-epithelial lymphocytes we found that the Westernized post-immigration microbiome stimulates increased inflammatory responses. This includes elevated TCRγδ in mice receiving post-immigration donor microbiomes. Elevated TCRγδ levels have been found to exhibit cytotoxic properties and produce inflammatory cytokines (Olivares-Villagómez and Van Kaer, 2018), and TCRγδ has been elevated in models of colitis, environmental enteropathy, and celiac disease (Abadie et al., 2012; Brown et al., 2015; Tsuchiya et al., 2003). CD4TCRαβ cells were elevated in Thai donor microbiome groups, which has been found to secrete IFNγ, a cytokine important for clearing pathogenic bacteria from the gut (Hess et al., 1996). Contradicting reported TCRγδ levels, we found that CD8ααTCRαβ, considered immunoregulatory and protective against colitis (Denning et al., 2007; Poussier et al., 2002), was significantly elevated in US donor microbiome groups. The levels of TCRγδ found at several orders of magnitude higher than its regulatory counterparts suggests a predominantly inflammatory intestinal environment. Although we do not observe differential inflammatory responses due to diet, we acknowledge that diet also has the potential to mediate the gut microbiome in order to induce low-grade inflammation (Cani et al., 2008, 2009a).

This study has several limitations. The fact that dietary acculturation only explains a small amount of microbiome variation suggests that immigration-induced microbiome changes are driven by a combination of diet and other, probably complex, factors associated with adjustment to life in the U.S. Most of these factors are challenging to examine in the context of this study. These include changes in exposure to stress, exercise, chlorinated municipal drinking water, antibiotics, and treatment for gut

parasites. There is likely to be an interacting web of altered exposures due to the dramatic shift in lifestyle following immigration to the U.S. that affect gut microbiome taxonomy, function, and diversity. In addition, although we have a large cross-sectional study population and a unique subset with longitudinal samples, our study design does not allow us to test what factors associated with U.S. immigration are causing loss of microbiome function and diversity, nor whether the changes in microbiome are contributing to the high incidence of obesity in U.S. immigrants.

Chapter 5: Concluding remarks and future work

In Chapter 2, we synthesized numerous complementary sources in this review, including microecological studies linking antibiotics and dysbiosis, mechanistic studies linking specific types of dysbiosis to specific disease outcomes, and reviews of epidemiological studies supporting antibiotics and increased disease risk. By synthesizing these independent literature reviews, we identified four major types of antibiotics-related dysbiosis, and we have presented a model for discussing and measuring pediatric dysbiosis in the context of several major diseases. Our findings indicate substantial existing evidence for a number of causal mechanisms by which the microbiome mediates antibiotic-related disease risk. The primary goal of continued research in pediatric dysbiosis will be to gain a mechanistic understanding how antibiotics usage by children may disrupt normal development of the gut microbiota, and at times consequently the immune system, potentially leading to increased risk of diseases like obesity, diabetes, allergies, asthma, and inflammatory bowel disease. Future work involving large, longitudinal cohorts of infants followed throughout life will be necessary in directly implicating microbial dysbiosis in mediating the link between childhood antibiotics and later development of disease.

In Chapter 3, we developed MLRepo, a repository of curated microbiome datasets made available for the computational community, and presented several case studies for how it can be a valuable resource. We expect that this repository will be of important use for machine-learning developers unfamiliar with microbiome data, educational purposes, and for scientists with limited access to high performance computing resources. We also hope that future datasets and tasks will be submitted by other researchers, and expect our repository to grow. Future work to transfer the repository from GitHub to a more robust database solution will be necessary.

In Chapter 4, we studied the impact of immigration on the gut microbiome by working with Hmong and Karen individuals pre-immigration, post-immigration, and who were born in the U.S. We demonstrated that U.S. immigration is associated with profound perturbations to the gut microbiome, including loss of native strain diversity, changes in

metabolic function, and shifts from *Prevotella* dominance to *Bacteroides* dominance. These changes begin immediately upon arrival and continue over decades of U.S. residence. The loss of diversity is compounded in obese individuals and in second-generation individuals born in the U.S. The microbiome has been shown to play a causal role in obesity and diabetes in animal models. Low gut microbiome diversity has been associated with a wide range of metabolic, infectious, and autoimmune diseases (Al-Ghalith et al., 2015; Chang et al., 2008; Gevers et al., 2014; Karlsson et al., 2013; Le Chatelier et al., 2013; Montassier et al., 2016; Turnbaugh et al., 2009b). These results demonstrate that further study is warranted to determine whether Westernization of the gut microbiome in immigrants may be contributing to their increased risk of metabolic diseases. In addition, the special circumstances that result in the relocation of refugees to the U.S. often means that these groups are especially vulnerable to socioeconomic disparities (Table 4.2), which may have additional direct or indirect impacts on health outcomes. Continued efforts to follow large, diverse, longitudinal immigrant and refugee cohorts, including dietary and therapeutic intervention studies, will be critical for determining how the microbiome may potentially be modulated to protect and improve immigrant metabolic health (Fu et al., 2016; Snijder et al., 2017).

Broader Impacts

Comprehensive assessment of microbiome structure in two minority ethnic groups

The main work in Chapter 4 focuses on two understudied at-risk ethnic groups who make up a large proportion of Minnesota's Asian population. As a community-based and community-participatory research project, this project has been formed with equal partnership of both community members and academic experts. Gut microbiome research in the U.S. has been primarily studied in Caucasian populations, and crowded-sourced projects such as the American Gut Project (McDonald et al., 2018) has self-selected for populations who are fluent in English, can navigate the Internet, have the means to donate $100, etc. This project represents an extremely unique opportunity for gut microbiome research to reach understudied and marginal groups, and for us to introduce and spread

knowledge of a relatively new and cutting-edge topic to these populations. Furthermore, although several studies have compared populations across countries (Yatsunenko et al., 2012), urban to rural gradients (Morton et al., 2015; Obregon-Tito et al., 2015), and short-term dietary interventions (David et al., 2014; O'Keefe et al., 2015), this research is, to our knowledge, the first to look at gut microbiomes of populations who have physically and permanently relocated between drastically different environments.

International project spanning migratory paths of the Hmong and Karen

Chapter 4 also transcends international borders as we worked with Hmong and Karen populations in the U.S. and Thailand. Thus, this work captures a comprehensive snapshot of how gut microbiomes differ across a rural village, refugee camp, and time-based integration into westernized life in Minnesota, while controlling for gender and ethnicity. The health of our target rural Hmong village has been documented over the last 15 years (Kunstadter, 2001), therefore the addition of gut microbiome analysis will be a powerful complement to this existing data. In addition, this study presents an unprecedented opportunity to work with residents of Mae La refugee camp, which is currently the largest refugee camp in Thailand and home to over 50,000 individuals of whom 90% are ethnic Karen (Banjong et al., 2003). Since 2005, nearly 100,000 refugees from Thailand have been resettled in the U.S. and other western countries (U.S. Department of State, Bureau of Population, Refugees, and Migration), hence emphasizing the relevance of our work for this population. The results of this study also has important implications for Thailand's Hmong-Thai and Karen-Thai residents, with potential to inform future work in understanding how non-communicable disease risk in these populations change with expansion of westernization and economic development, but also in response to in-country migration across the rural-urban gradient.

Cutting-edge tools enable novel insights into microbiome analyses

A combination of 16s rRNA amplicon sequencing, deep shotgun metagenomic sequencing, and targeted metabolomics will be applied to provide a comprehensive view of the changing gut microbiome over space and time. The amount of total data generated

will be on the order of hundreds of millions of sequences, requiring cutting-edge computational tools and techniques to identify signals and trends. The advanced computational methods developed for the human research study resulted in resources such as MLRepo that have potential to be important resources for computational scientists.

Potential for translatable results with impact to public health in Minnesota

Our results have the potential to transform nutritional guidance provided by clinicians and to be integrated permanently into the "new immigrant" core curriculum offered at partnering community organizations. Additionally, refugee health practices at the Minnesota Department of Health may be expanded to include gut microbiome analysis in medical examinations of new arrivals. Worksite wellness policies for Minnesota companies employing large numbers of Karen or Hmong employees may also be impacted to promote dietary fiber consumption. And importantly, prevention of obesity and hence diabetes, heart disease, and other obesity-related conditions in the broader U.S. immigrant population has tremendous potential to reduce the high economic burden of obesity, an estimated $147 to $210 billion a year (Finkelstein et al., 2009).

Tables

Chapter 2 Tables

INFECTIOUS DISEASES	
Antibiotics and Disease	(Ayres et al., 2012; Brandl et al., 2008; Buffie et al., 2012; Croswell et al., 2009; Noverr et al., 2004; Sekirov et al., 2008; Ubeda et al., 2010)
Antibiotics and Microbiome	(Ayres et al., 2012; Brandl et al., 2008; Buffie et al., 2012; Croswell et al., 2009; Donskey et al., 2000; Noverr et al., 2004; Sekirov et al., 2008; Ubeda et al., 2010)
Microbiome and Disease	(Abt et al., 2012; Asahara et al., 2001; Buffie et al., 2012; Mai et al., 2011b; Mazmanian et al., 2008; Shu et al., 2000; Ubeda et al., 2010; Wang et al., 2009)
ALLERGY, ATOPIC, AND AUTOIMMUNE DISEASES	
Antibiotics and Disease	(Droste et al., 2000; Farooqi and Hopkin, 1998; Johnson et al., 2005; McKeever et al., 2002; Ong et al., 2014; Russell et al., 2012; Stefka et al., 2014; Wickens et al., 1999)
Antibiotics and Microbiome	(Russell et al., 2012; Stefka et al., 2014)
Microbiome and Disease	(Abrahamsson et al., 2012; Atarashi et al., 2013; Bisgaard et al., 2011; Björkstén et al., 2001b; Brown et al., 2011; Cani et al., 2008; Giongo et al., 2011; Kalliomäki et al., 2001; Lee and Mazmanian, 2010; Lundell et al., 2014; Russell et al., 2012; Sellitto et al., 2012; Sjögren et al., 2009; Stefka et al., 2014; Valladares et al., 2010; Wen et al., 2008)
OBESITY	
Antibiotics and Disease	(Ajslev et al., 2011; Bailey et al., 2014; Cho et al., 2012c; Cox et al., 2014; Trasande et al., 2013)
Antibiotics and Microbiome	(Ajslev et al., 2011; Cho et al., 2012c; Cox et al., 2014)
Microbiome and Disease	(Ajslev et al., 2011; Cho et al., 2012c; Cox et al., 2014; Ley et al., 2006; Ridaura et al., 2013; Turnbaugh et al., 2009b)

Table 2.1. References synthesizing mechanistic and epidemiological evidence linking antibiotics, changes in the gut microbiome, and disease.

Chapter 3 Tables

Project Name	V Region	Target size	Num samples	Num subjects	Area	Description	Sequencing Technology	Study Design
Cho 2012	V3	177	95	47	Antibiotics	Mouse fecal and cecal samples, Control vs. 4 kinds of antibiotics	454	Cross-Sectional
Claesson 2012	V4	221	168	168	Age	Elderly and young adults	454	Cross-Sectional
David 2014	V4	282	235	11	Diet	Plant-based vs. Animal-based diet, Cross-over study	Illumina MiSeq	Longitudinal
Gevers 2014	V4	173	1321	668	IBD	Biopsies from IBD patients prior to treatment	Illumina MiSeq	Cross-Sectional
HMP 2012	V35	527	6407	242	Body Habitat, Gender	Up to 18 body sites across 242 healthy subjects at 1-2 time points	454	Cross-Sectional
Kostic 2012	V35	569	190	95	Colorectal Cancer	Adjacent Healthy vs. Tumor Colon Biopsy Tissues	454	Paired
Montassier 2016	V56	280	28	28	Bacteremia	Patients prior to chemotherapy who did or did not develop bacteremia	454	Cross-Sectional
Morgan 2012	V35	569	231	231	IBD	Healthy, Crohn's Disease, or Ulcerative Colitis patients	454	Cross-Sectional
Turnbaugh 2009	V2	230	281	154	Obesity	Monozygotic or dizygotic twin pairs concordant for BMI class, and their mothers	454	Cross-Sectional
Wu 2011	V12	244	95	10	Diet	Controlled HighFat or LowFat feeding on 10 subjects over 10 days	454	Longitudinal
Yatsunenko 2012	V4	282	531	531	Geography, Age, Gender	Humans of varying ages from the USA, Malawi, and Venezuela	Illumina MiSeq	Cross-Sectional
Ravel 2011	V12	240	396	396	Bacterial Vaginosis	Vaginal samples from four ethnic groups nugent scores for bacterial vaginosis	454	Cross-Sectional
Karlsson 2013	NA	NA	144	144	Diabetes	Patients with normal, impaired, or type 2 diabetes glucose tolerance categories	Illumina HiSeq	Cross-Sectional
Qin 2012	NA	NA	134	134	Diabetes	Healthy vs type 2 diabetes Chinese patients	Illumina HiSeq	Cross-Sectional
Qin 2014	NA	NA	130	130	Cirrhosis	Cirrhosis versus healthy	Illumina HiSeq	Cross-Sectional

Table 3.1. Dataset descriptions

Chapter 4 Tables

Sample Group	**BMI < 25**	**BMI ≥ 25**
KarenThai	45	39
HmongThai	42	53
Karen1st	77	67
Hmong1st	52	85
Hmong2nd	19	35
Controls	23	13

Table 4.1. Sample group recruitment stratified by BMI threshold of 25.

	KarenThai	HmongThai	Karen1st	Hmong1st	Hmong2nd	Control	P
N	84	95	144	137	54	36	
Age	35 (18-55)	43 (20-78)	35 (18-67)	39 (18-65)	25 (18-39)	34 (18-64)	*3.60E-16*
Waist-to-Height Ratio	0.52 (0.37-0.71)	0.61 (0.47-0.92)	0.57 (0.38-0.71)	0.61 (0.4-0.83)	0.61 (0.4-0.87)	0.55 (0.44-0.9)	*1.50E-18*
Years in US	NA	NA	3 (0.003-9.8)	20 (0.049-41)	NA	NA	*5.80E-40*
BMI Class							*5.00E-04*
Lean	45 (53.6)	42 (44.2)	77 (53.5)	52 (38)	19 (35.2)	23 (63.9)	
Overweight	30 (35.7)	37 (38.9)	51 (35.4)	54 (39.4)	16 (29.6)	4 (11.1)	
Obese	9 (10.7)	16 (16.8)	16 (11.1)	31 (22.6)	19 (35.2)	9 (25)	
Alcohol Use							*5.00E-04*
Never	83 (98.8)	84 (88.4)	118 (81.9)	113 (82.5)	27 (50)	5 (13.9)	
Daily	0 (0)	0 (0)	3 (2.08)	0 (0)	0 (0)	0 (0)	
Weekly	0 (0)	0 (0)	2 (1.39)	6 (4.38)	9 (16.7)	10 (27.8)	
Monthly	0 (0)	5 (5.26)	3 (2.08)	7 (5.11)	11 (20.4)	13 (36.1)	
< Monthly	0 (0)	0 (0)	12 (8.33)	10 (7.3)	6 (11.1)	7 (19.4)	
Quit	1 (1.19)	6 (6.32)	5 (3.47)	0 (0)	1 (1.85)	1 (2.78)	
Tobacco Use							*5.00E-04*
Never	73 (86.9)	92 (96.8)	130 (90.3)	135 (98.5)	48 (88.9)	28 (77.8)	
Daily	10 (11.9)	0 (0)	8 (5.56)	1 (0.73)	1 (1.85)	0 (0)	
< Monthly	1 (1.19)	1 (1.05)	1 (0.694)	0 (0)	3 (5.56)	2 (5.56)	
Quit	0 (0)	2 (2.11)	5 (3.47)	1 (0.73)	2 (3.7)	6 (16.7)	
Highest Education							*5.00E-04*
None	16 (19)	0 (0)	0 (0)	4 (2.92)	0 (0)	0 (0)	
ESL	0 (0)	0 (0)	96 (66.7)	14 (10.2)	0 (0)	0 (0)	
< HS	38 (45.2)	34 (35.8)	18 (12.5)	25 (18.2)	1 (1.85)	0 (0)	
HS	24 (28.6)	9 (9.47)	23 (16)	31 (22.6)	8 (14.8)	1 (2.78)	
College	2 (2.38)	4 (4.21)	0 (0)	41 (29.9)	38 (70.4)	10 (27.8)	
Graduate School	4 (4.76)	0 (0)	2 (1.39)	12 (8.76)	6 (11.1)	25 (69.4)	
Birth Location							*5.00E-04*
Refugee Camp	6 (7.14)	2 (2.11)	32 (22.2)	31 (22.6)	0 (0)	0 (0)	
Rural	77 (91.7)	93 (97.9)	110 (76.4)	101 (73.7)	1 (1.85)	1 (2.78)	
Urban	1 (1.19)	0 (0)	2 (1.39)	3 (2.19)	53 (98.1)	34 (94.4)	
Medical Assistance	NA	NA	119 (82.6)	60 (43.8)	15 (27.8)	2 (5.56)	*5.00E-04*
Public Housing	NA	NA	20 (13.9)	20 (14.6)	9 (16.7)	4 (11.1)	*0.92*
Children Receives Free Lunch	NA	NA	89 (61.8)	54 (39.4)	5 (9.26)	3 (8.33)	*5.00E-04*

Table 4.2. Sample Group Characteristics.
All values are represented as mean (min - max). HS = High School; ESL = English as a Second Language; < = less than. Note that all participants are female, for reasons already noted.

Acacia Leaves Cha om	M 150
Asia Mix	Milk Candy
Banana Flower	Naked Green Juice
Banana Trunk	Nature Valley Peanut Butter Cup
Banh Mi Vietnamese Pork Sandwich	Pacific Soup Sweet Potato Masala
Beijing Beef	Pad Kraprow
Birdy Thai Coffee	Pediasure
Chili Paste	Pork Skin
Djenkol Bean	Protein Powder
Dried Fish Soup	Pumpkin Leaves
Dried Fried Fish	Raising Canes 3 box combo no drink
Egg Noodles Yellow	Rambutan
Ei Kyar Kway	Rambutan canned
Ellse	Roselle Leaves
Exo Protein Bar	Schaut Thee Zay Byar
Fish Paste	Sesbania
Fish Soup	Shrimp Paste
Gourd	Skinny Cow Chocolate Bar
Green Max Yams and Multi Grain Cereal	Snake Loofah
Halawa	Spinach Smoothie
Hmong Sausage	Sweet Thai Chili Sauce
Hon Tsai Tai	Tapioca
Jack Fruit	Taro Leaf
Khao Poon	Thai Glass Noodle Salad
Kaw Naw	Thai Northern Sausage
Khao Soy Soup	Thai Papaya Salad
Khao Pia	Thai Tapioca Dessert with coconut
Larb Moo	Tomato Curry
Lead Tree	Veggie Fritters
Lead Tree Pod	Vietnamese Sausage
Lean 25 Smoothie	Voiz Cracker Milk
Leek and Potato Soup	Water Convolvulus Water Spinach
Sin Tone Ma Nwe	Wheat Powder and Sugar
Longan	Wing Bean
Loofah	Yakult
Luna Protein Bar	Zesty Chicken and Black Bean Salad Bowl Starbucks

Table 4.3. Foods that were individually researched, then entered as custom foods into SuperTracker and assigned a unique identifier for the food tree.

OTUID	q	diff-mean-prevalence	Taxonomy
220	6.84E-02	0.051	t__Prevotella copri DSM 18205
553	2.05E-01	0.029	Enterobacteriaceae
899	1.18E-01	0.036	Blautia faecis
921	3.85E-02	0.058	Hungatella effluvii
1175	7.26E-01	0.011	Romboutsia timonensis
58	5.05E-01	-0.011	Faecalibacterium prausnitzii
75	3.26E-01	0.034	Faecalibacterium prausnitzii
12	5.70E-01	0.023	Gemmiger formicilis
1611	2.46E-04	0.165	Clostridium
1812	3.05E-14	0.456	Faecalibacterium prausnitzii
267	2.76E-01	0.038	t__Prevotella copri DSM 18205
394	6.58E-01	-0.014	t__Bacteroides vulgatus ATCC 8482
543	7.51E-02	0.068	Enterobacterales
818	4.04E-01	0.031	Blautia luti
909	2.76E-01	0.038	Dorea formicigenerans
936	5.70E-01	0.023	Blautia
1276	1.81E-02	0.103	Faecalibacterium prausnitzii
1667	8.13E-01	0.013	[Eubacterium] hallii
1773	4.60E-01	0.035	Eubacterium
1845	9.16E-06	0.224	Faecalibacterium prausnitzii
1905	2.89E-09	0.337	Faecalibacterium prausnitzii
455	1.39E-03	0.148	t__Parabacteroides distasonis ATCC 8503
614	8.13E-01	0.013	Butyricicoccus
63	1.43E-06	0.254	Faecalibacterium prausnitzii
71	6.96E-03	0.118	Faecalibacterium prausnitzii
754	8.13E-01	0.013	Anaerostipes hadrus
806	6.26E-01	0.020	t__Blautia obeum ATCC 29174
822	2.86E-02	0.096	Blautia obeum
1643	1.64E-05	0.230	Clostridiales
1890	9.39E-06	0.238	Faecalibacterium prausnitzii
20	7.98E-04	0.169	Subdoligranulum variabile
832	8.02E-01	-0.013	Fusicatenibacter saccharivorans
884	4.00E-02	0.093	Lachnoclostridium
1200	7.38E-02	0.091	Intestinibacter bartlettii
1552	6.84E-01	0.029	Erysipelotrichaceae
1888	2.45E-08	0.337	Faecalibacterium prausnitzii
3761	2.10E-03	0.160	Blautia
881	6.84E-01	0.021	Lachnoclostridium
1458	2.47E-01	0.065	Clostridiales
1453	6.86E-04	0.190	Oscillospiraceae
3283	1.09E-03	0.182	Blautia
427	6.49E-01	-0.021	Bacteroides
576	3.35E-01	0.057	t__Haemophilus parainfluenzae ATCC 33392
828	1.77E-01	0.073	Fusicatenibacter
1809	5.48E-13	0.504	Faecalibacterium prausnitzii
1652	5.90E-01	0.031	Lactobacillus rogosae
1728	4.85E-01	-0.032	Roseburia faecis
1956	5.96E-08	0.354	Faecalibacterium
383	9.97E-03	0.149	Alistipes shahii
43	1.19E-07	0.338	Faecalibacterium prausnitzii
929	2.37E-02	0.125	Hungatella
1672	5.11E-01	-0.036	[Eubacterium] hallii
1709	8.23E-02	0.099	t__Roseburia intestinalis L1-82
1715	4.87E-01	0.044	t__Roseburia hominis A2-183
1846	1.21E-06	0.315	Faecalibacterium prausnitzii
2541	1.39E-08	0.386	t__Prevotella copri DSM 18205
285	6.20E-05	0.251	t__Prevotella copri DSM 18205
42	1.21E-06	0.315	Faecalibacterium prausnitzii
4334	5.96E-02	0.107	Blautia
738	2.08E-01	-0.052	Actinomyces odontolyticus
856	7.96E-04	0.203	Clostridiales
886	5.38E-03	0.171	Ruminococcus
895	2.18E-01	0.075	Lachnoclostridium
953	6.98E-01	-0.020	Bacteroides xylanolyticus
1283	3.04E-02	-0.080	Streptococcus
1288	3.78E-01	-0.048	Streptococcus

1752	2.29E-02	0.145	t__Ruminococcus faecis JCM 15917
930	1.28E-01	0.097	Hungatella
1277	1.41E-01	0.094	Faecalibacterium prausnitzii
3910	2.55E-01	0.078	Blautia
4372	7.98E-04	0.225	Faecalibacterium prausnitzii
743	5.43E-02	0.127	Tyzzerella
771	3.33E-01	0.070	Lachnoclostridium
1045	6.51E-01	0.034	Collinsella aerofaciens
1084	5.45E-01	0.042	Collinsella aerofaciens
1595	3.05E-02	0.150	Phascolarctobacterium succinatutens
1463	1.54E-01	0.100	Oscillibacter
1786	3.08E-02	0.141	Eubacterium
2346	4.46E-08	0.397	t__Prevotella copri DSM 18205
3924	2.70E-01	0.075	Blautia
435	2.04E-01	-0.065	t__Bacteroides xylanisolvens XB1A
534	1.54E-01	0.092	Desulfovibrio
9	1.71E-09	0.447	Gemmiger formicilis
928	8.23E-02	0.117	Hespellia
1863	3.11E-03	0.206	Faecalibacterium prausnitzii
1986	9.06E-09	0.432	Faecalibacterium prausnitzii
2571	5.00E-10	0.474	t__Prevotella copri DSM 18205
664	4.53E-01	0.056	Acutalibacter
954	5.54E-01	0.048	Bacteroides xylanolyticus
962	9.15E-01	0.014	Coprococcus catus
1252	1.37E-04	0.281	Gemmiger formicilis
1452	8.03E-05	0.290	Oscillibacter
1523	8.44E-03	0.188	Holdemanella biformis
23	6.26E-01	-0.040	Faecalibacterium
1922	2.84E-06	0.349	Faecalibacterium prausnitzii
1957	8.55E-07	0.374	Faecalibacterium
1971	1.86E-02	0.171	Faecalibacterium
2569	6.34E-11	0.510	t__Prevotella copri DSM 18205
411	6.26E-01	-0.040	Bacteroides uniformis
4326	8.44E-03	0.188	t__[Eubacterium rectale] ATCC 33656
940	3.77E-01	0.070	Blautia
95	8.03E-05	0.290	Faecalibacterium prausnitzii
1660	4.17E-02	0.152	t__[Eubacterium] eligens ATCC 27750
1725	6.45E-03	0.204	t__[Eubacterium rectale] ATCC 33656
1891	1.66E-04	0.281	Faecalibacterium prausnitzii
2337	8.84E-08	0.418	t__Prevotella copri DSM 18205
283	1.26E-08	0.452	t__Prevotella copri DSM 18205
3963	3.89E-01	0.067	Blautia
4830	3.38E-07	0.392	[Eubacterium] hallii
4880	3.29E-05	0.315	t__[Eubacterium rectale] ATCC 33656
503	1.86E-01	0.101	t__Parabacteroides merdae ATCC 43184
798	5.77E-02	0.144	Ruminococcus faecis
905	1.05E-01	0.118	Clostridiales
1206	2.04E-07	0.411	Terrisporobacter petrolearius
1402	3.24E-01	0.081	Veillonella
1688	3.32E-02	0.159	t__Senegalimassilia anaerobia JC110
1939	1.19E-07	0.419	Faecalibacterium prausnitzii
1951	1.19E-07	0.419	Faecalibacterium prausnitzii
1984	7.01E-14	0.619	Faecalibacterium prausnitzii
2325	3.68E-11	0.541	t__Prevotella copri DSM 18205
255	8.23E-02	0.133	t__Prevotella copri DSM 18205
3282	7.13E-11	0.532	Dorea longicatena
3760	4.38E-06	0.359	Dorea longicatena
4134	7.36E-03	0.203	Collinsella aerofaciens
4786	1.29E-05	0.341	Lactobacillus rogosae
773	6.59E-01	-0.031	Lachnoclostridium
918	6.84E-01	0.038	Blautia obeum
1033	2.66E-01	0.087	Collinsella aerofaciens
1086	5.53E-01	-0.045	Collinsella aerofaciens
1217	1.21E-02	0.201	Sutterella
1442	5.88E-04	0.271	Clostridium
2031	1.38E-04	0.307	Faecalibacterium prausnitzii
208	1.64E-06	0.386	t__Prevotella copri DSM 18205

2106	9.05E-06	0.359	Faecalibacterium prausnitzii
248	1.83E-08	0.465	t__Prevotella copri DSM 18205
2513	1.22E-07	0.429	t__Prevotella copri DSM 18205
2545	1.08E-09	0.508	t__Prevotella copri DSM 18205
2788	3.05E-14	0.640	t__Prevotella copri DSM 18205
3285	1.46E-04	0.298	Collinsella aerofaciens
3773	1.46E-04	0.298	Faecalibacterium prausnitzii
3929	4.88E-02	0.157	Blautia
3959	9.07E-04	0.263	Blautia
4112	1.22E-07	0.429	Collinsella aerofaciens
809	5.53E-03	0.219	[Clostridium] glycyrrhizinilyticum
917	1.18E-01	0.122	Clostridiales
1078	4.24E-01	0.067	Collinsella aerofaciens
1085	9.26E-01	-0.013	Collinsella aerofaciens
1849	3.35E-12	0.591	Faecalibacterium prausnitzii
227	2.16E-06	0.387	t__Prevotella copri DSM 18205
2349	1.21E-16	0.707	t__Prevotella copri DSM 18205
2462	4.25E-08	0.458	t__Prevotella copri DSM 18205
302	4.25E-08	0.458	t__Prevotella copri DSM 18205
400	4.48E-01	-0.058	Bacteroides dorei
437	5.57E-01	-0.049	Bacteroides ovatus
4371	3.16E-10	0.529	Faecalibacterium prausnitzii
69	4.55E-04	0.280	Faecalibacterium prausnitzii
938	9.26E-01	0.013	Blautia
1332	9.97E-02	0.135	Flintibacter
1466	5.59E-02	0.162	Oscillibacter
1813	1.28E-06	0.406	Faecalibacterium prausnitzii
1909	5.12E-11	0.568	Faecalibacterium prausnitzii
1987	9.78E-08	0.451	Faecalibacterium prausnitzii
202	2.46E-08	0.478	t__Prevotella copri DSM 18205
2121	1.77E-04	0.316	Faecalibacterium prausnitzii
2345	7.08E-14	0.649	t__Prevotella copri DSM 18205
2348	1.11E-15	0.694	t__Prevotella copri DSM 18205
2459	1.20E-05	0.370	t__Prevotella copri DSM 18205
2810	1.23E-17	0.739	t__Prevotella copri DSM 18205
4060	6.86E-09	0.496	Clostridiales
450	6.70E-03	0.226	t__Parabacteroides distasonis ATCC 8503
4496	2.01E-12	0.604	Clostridiales
524	1.30E-05	0.361	t__Prevotella copri DSM 18205
1178	5.01E-03	0.243	Romboutsia timonensis
1072	3.61E-01	0.088	Collinsella aerofaciens
1461	1.04E-01	0.142	Oscillospiraceae
1616	5.92E-02	0.161	Clostridium
1767	9.26E-01	-0.013	Blautia
1870	5.94E-11	0.571	Faecalibacterium prausnitzii
1970	5.68E-04	0.288	Faecalibacterium
201	8.87E-09	0.498	t__Prevotella copri DSM 18205
221	4.72E-06	0.389	t__Prevotella copri DSM 18205
2354	2.81E-06	0.398	t__Prevotella copri DSM 18205
2465	1.88E-07	0.453	t__Prevotella copri DSM 18205
2653	2.81E-06	0.398	t__Prevotella copri DSM 18205
2705	3.19E-11	0.580	t__Prevotella copri DSM 18205
3840	7.86E-02	0.151	Dorea longicatena
3876	4.40E-02	0.170	Dorea formicigenerans
4115	7.86E-02	0.151	Collinsella aerofaciens
666	9.26E-01	0.015	Ruminococcus bromii
758	7.10E-01	0.042	Ihubacter
1076	8.18E-01	0.029	Collinsella aerofaciens
1146	3.69E-01	0.085	Thermoactinomycetaceae
1548	6.29E-02	0.159	Turicibacter sanguinis
1597	2.81E-06	0.408	Phascolarctobacterium succinatutens
1766	3.02E-11	0.593	Coprococcus
1819	8.53E-09	0.510	Gemmiger formicilis
1835	4.88E-05	0.353	Faecalibacterium prausnitzii
1883	4.88E-06	0.399	Faecalibacterium prausnitzii
1999	8.23E-02	0.149	Faecalibacterium prausnitzii
2071	2.81E-06	0.408	Faecalibacterium prausnitzii

2338	1.71E-09	0.538	t__Prevotella copri DSM 18205
4146	4.48E-01	0.075	Collinsella aerofaciens
432	8.13E-01	-0.026	Bacteroides fragilis
4328	1.44E-03	0.279	t__Ruminococcus faecis JCM 15917
4911	1.38E-04	0.334	Faecalibacterium prausnitzii
52	1.19E-07	0.464	Faecalibacterium prausnitzii
5389	3.10E-08	0.491	Faecalibacterium prausnitzii
848	4.60E-02	0.177	Dorea longicatena
949	6.29E-02	0.159	Hungatella
1049	8.19E-01	0.025	Collinsella aerofaciens
1145	6.00E-03	0.241	Thermoactinomycetaceae
1123	1.53E-04	0.335	Blautia faecis
1262	6.00E-03	0.241	Parasutterella excrementihominis
1771	4.88E-02	0.175	t__[Eubacterium rectale] ATCC 33656
1848	3.02E-05	0.372	Faecalibacterium prausnitzii
1916	9.99E-06	0.391	Faecalibacterium prausnitzii
1928	6.64E-08	0.485	Faecalibacterium prausnitzii
1950	2.46E-04	0.325	Faecalibacterium prausnitzii
2009	1.81E-02	0.213	Faecalibacterium prausnitzii
3554	3.02E-05	0.372	t__[Eubacterium rectale] ATCC 33656
3950	1.13E-01	0.147	Blautia
4136	2.46E-04	0.325	Collinsella aerofaciens
4151	3.61E-02	0.185	Collinsella aerofaciens
4327	9.99E-06	0.391	Roseburia faecis
4439	2.93E-06	0.419	Faecalibacterium prausnitzii
6447	6.64E-08	0.485	Subdoligranulum variabile
14	4.55E-04	0.316	Gemmiger formicilis
1467	7.04E-04	0.307	Oscillibacter
1473	1.81E-03	0.278	Sporobacter
2081	9.84E-07	0.449	Faecalibacterium prausnitzii
219	5.02E-07	0.459	t__Prevotella copri DSM 18205
2317	2.48E-09	0.544	Blautia
2406	1.61E-12	0.649	t__Prevotella copri DSM 18205
3626	3.34E-06	0.421	Lactobacillus rogosae
3820	4.59E-01	0.079	Fusicatenibacter saccharivorans
3884	2.72E-02	0.202	Hungatella effluvii
4155	6.35E-01	0.050	Collinsella aerofaciens
4515	1.83E-08	0.516	Lachnospiraceae
744	1.63E-04	0.345	Tyzzerella
783	8.92E-02	0.155	t__[Eubacterium rectale] ATCC 33656
1486	1.42E-02	0.230	Sporobacter
1741	7.29E-01	0.037	t__[Eubacterium rectale] ATCC 33656
1958	3.13E-07	0.470	Faecalibacterium
224	7.53E-08	0.499	t__Prevotella copri DSM 18205
2441	4.20E-08	0.509	t__Prevotella copri DSM 18205
2497	3.13E-07	0.470	t__Prevotella copri DSM 18205
2542	1.69E-12	0.653	t__Prevotella copri DSM 18205
2543	3.33E-11	0.615	t__Prevotella copri DSM 18205
2628	1.34E-07	0.490	t__Prevotella copri DSM 18205
2656	4.40E-09	0.547	t__Prevotella copri DSM 18205
3307	2.16E-09	0.557	t__Prevotella copri DSM 18205
3631	2.05E-05	0.393	[Eubacterium] hallii
3790	7.85E-04	0.307	t__Blautia obeum ATCC 29174
3897	6.29E-05	0.374	Blautia
3905	2.01E-02	0.220	Blautia
4107	7.85E-04	0.307	Collinsella aerofaciens
4148	2.01E-02	0.220	Collinsella aerofaciens
423	8.18E-01	-0.031	t__Bacteroides stercoris ATCC 43183
4324	5.16E-04	0.316	t__Roseburia inulinivorans DSM 16841
4337	4.75E-04	0.326	Anaerosporobacter
772	1.42E-02	0.230	Lachnoclostridium
857	2.90E-04	0.335	Clostridiales
984	1.98E-06	0.441	Prevotella copri
985	7.00E-06	0.413	Prevotella

Table 4.4. OTU prevalences in HmongThai and Hmong1st. Related to Figure 2.

P1	Prevotella_stercorea_DSM_18206_Scfld0
P2	Prevotella_copri_strain_Indica_contig00001
P3	Prevotella_copri_DSM_18205_Scfld26
P4	Prevotellamassilia_timonensis_strain_Marseille-P2831
B1	Bacteroides_vulgatus.1cell.HGAP3_contig1
B2	Bacteroides_stercoris_ATCC_43183_Scfld_02_16
B3	Bacteroides_finegoldii_DSM_17565_Scfld32
B4	Bacteroides_uniformis_str._3978_T3_i_gbf3978T3i.contig.0
B5	Bacteroides_massiliensis_B84634_=_Timone_84634_=_DSM_17679_=_JCM_13223_strain_DSM_17679_aczJl-supercont1.1
B6	Bacteroides_dorei_CL02T12C06_supercont1.1
B7	Bacteroides_caccae_CL03T12C61_supercont1.1
B8	Bacteroides_caccae_strain_2789STDY5834946
B9	Bacteroides_intestinalis_DSM_17393_B_intestinalis-2.0.1_Cont607

Table 4.5. NCBI Genome IDs of Bacteroides and Prevotella strains. Related to Figure 3.

Food Item	q-value	p-value	ρ
Cooked cereals rice	7.29e-315	2.52e-316	-0.96
Fruits excluding berries	3.12e-28	2.15e-29	0.45
Milk fluid	2.82e-12	2.91e-13	0.30
Coffee	2.58e-07	3.71e-08	0.23
Other vegetables cooked	2.58e-07	4.45e-08	-0.23
White breads rolls	7.71e-07	1.60e-07	0.22
Mixtures mainly grain pasta or bread	2.15e-06	5.18e-07	0.21
Finfish	2.80e-06	7.72e-07	-0.21
Soft drinks carbonated	3.60e-05	1.12e-05	0.19
Other vegetables raw	7.46e-05	2.57e-05	-0.18
Citrus fruits	8.43e-05	3.20e-05	0.18
Frankfurters sausages lunchmeats meat spreads	8.70e-05	3.60e-05	0.18
Carrots	2.40e-04	1.07e-04	0.16
Chicken	1.14e-03	5.51e-04	0.15
Beef roasts stew meat corned beef beef brisket sandwich steaks	1.86e-03	9.61e-04	0.14
Fruit juices excluding citrus	2.18e-03	1.21e-03	0.14
Cookies	2.18e-03	1.28e-03	0.14

Table 4.6. Foods (summarized at level 3) that are significantly correlated with PC1 of the diet-based unweighted Unifrac PCOA ($q < 0.01$). Related to Figure 4.

Figures

Chapter 2 Figures

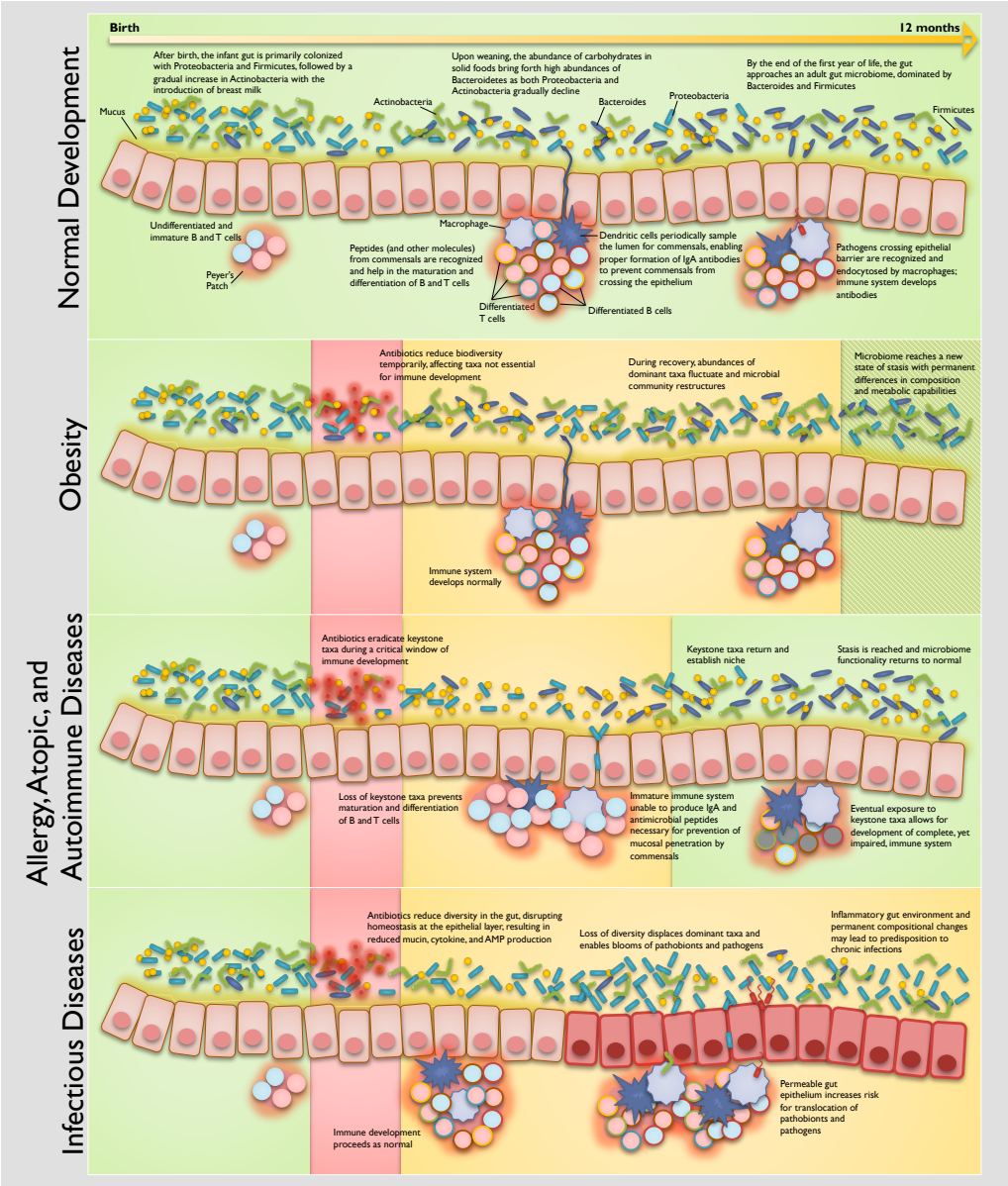


Figure 2.1. Disease model of host-microbiome development.
Disease classes are associated with cascading dysbiosis types, with important dependencies on host-microbiome development. Note that disease classes and dysbiosis types are not necessarily mutually exclusive. The proposed mechanisms presented are supported by extensive evidence in the literature, both from mechanistic studies and from epidemiological surveys. Due to the very large number of references, the citations represented in this figure can be found in Table 2.1.

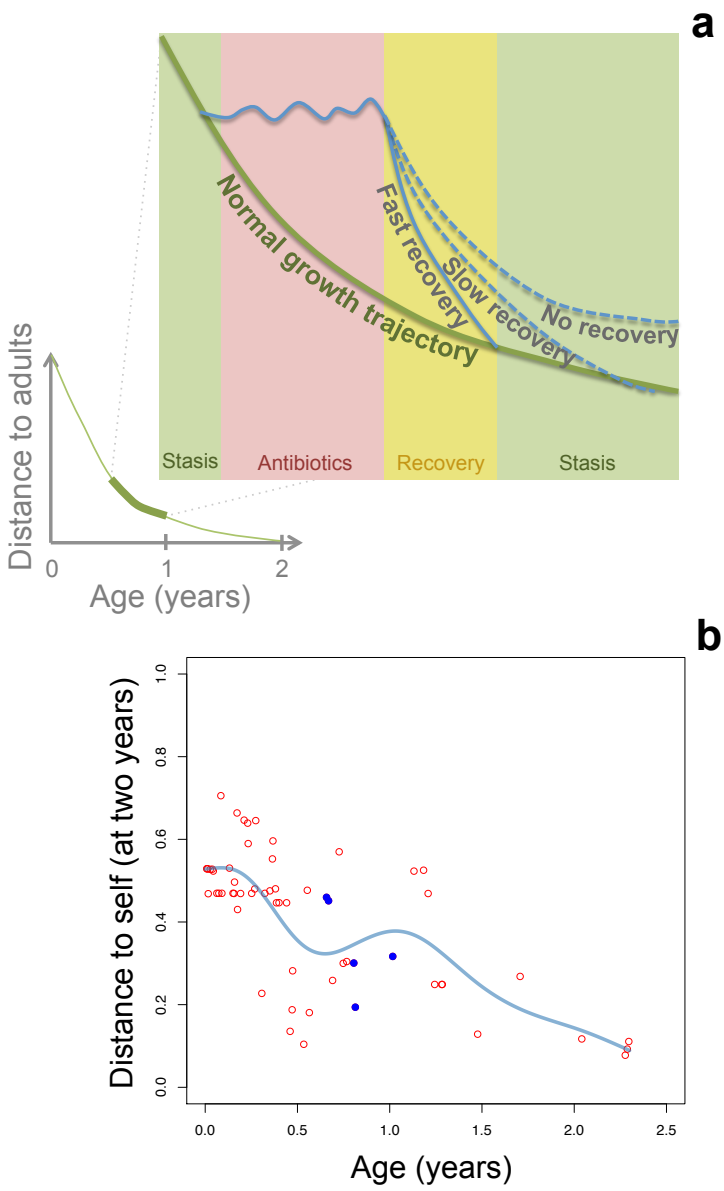


Figure 2.2. Trajectories for infant recovery after antibiotic exposure.
(A) Infant gut microbiomes develop rapidly and experience large changes during infancy before becoming indistinguishable from adult microbiomes by age 2. Dysbiosis in infants can displace (no recovery) or delay (slow recovery) development on the normal growth trajectory. (B) Samples were obtained from a single infant over time (Koenig et al., 2011), and microbiome distance (Bray-Curtis) to self at 2 years old was plotted over time. Fecal samples collected immediately after antibiotics are denoted in blue. A smoothing spline (in light blue) reveals a noticeable change in trajectory of development after use of antibiotics, mirroring the deviation in trajectory predicted in (A).

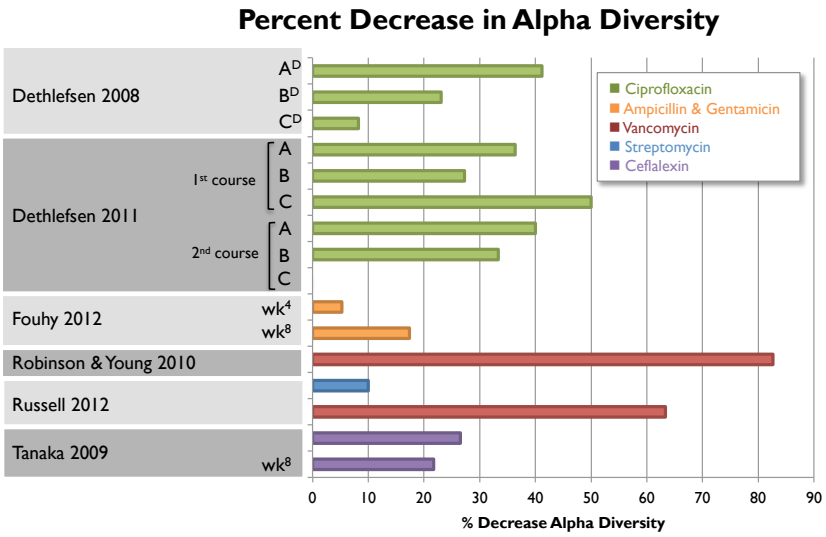


Figure 2.3. Percent decrease in gut microbiome biodiversity across studies with different antibiotic exposures.

All fecal samples were collected 1 week after antibiotic course was completed, except where noted by subscripts. The Dethlefsen 2011 study included three subjects (A, B, C) who received two courses 6 months apart. [D]Sample taken during antibiotic treatment; [4]Sample taken 4 weeks after antibiotic completion; [8]Sample taken 8 weeks after antibiotic completion (Dethlefsen and Relman, 2011; Dethlefsen et al., 2008; Fouhy et al., 2012; Robinson and Young, 2010; Russell et al., 2012; Tanaka et al., 2009).

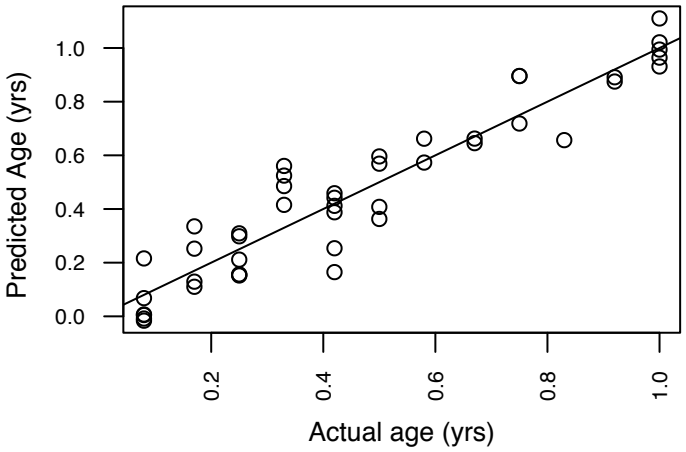


Figure 2.4. Predicted Microbiome Maturity Index (MMI).
The predictive MMI for a given child is compared to the true age of that child. The MMI was predicting using random forests regression algorithm trained on the microbiome compositions and true ages of all children except for one being predicted. True age was predicted to within ± 1.3 months (standard deviation of the predicted error), demonstrating the feasibility of modeling the maturation of the gut microbiota as a predictable process across individuals. Microbiome samples were obtained from children living in the USA (Yatsunenko et al., 2012).

Chapter 3 Figures

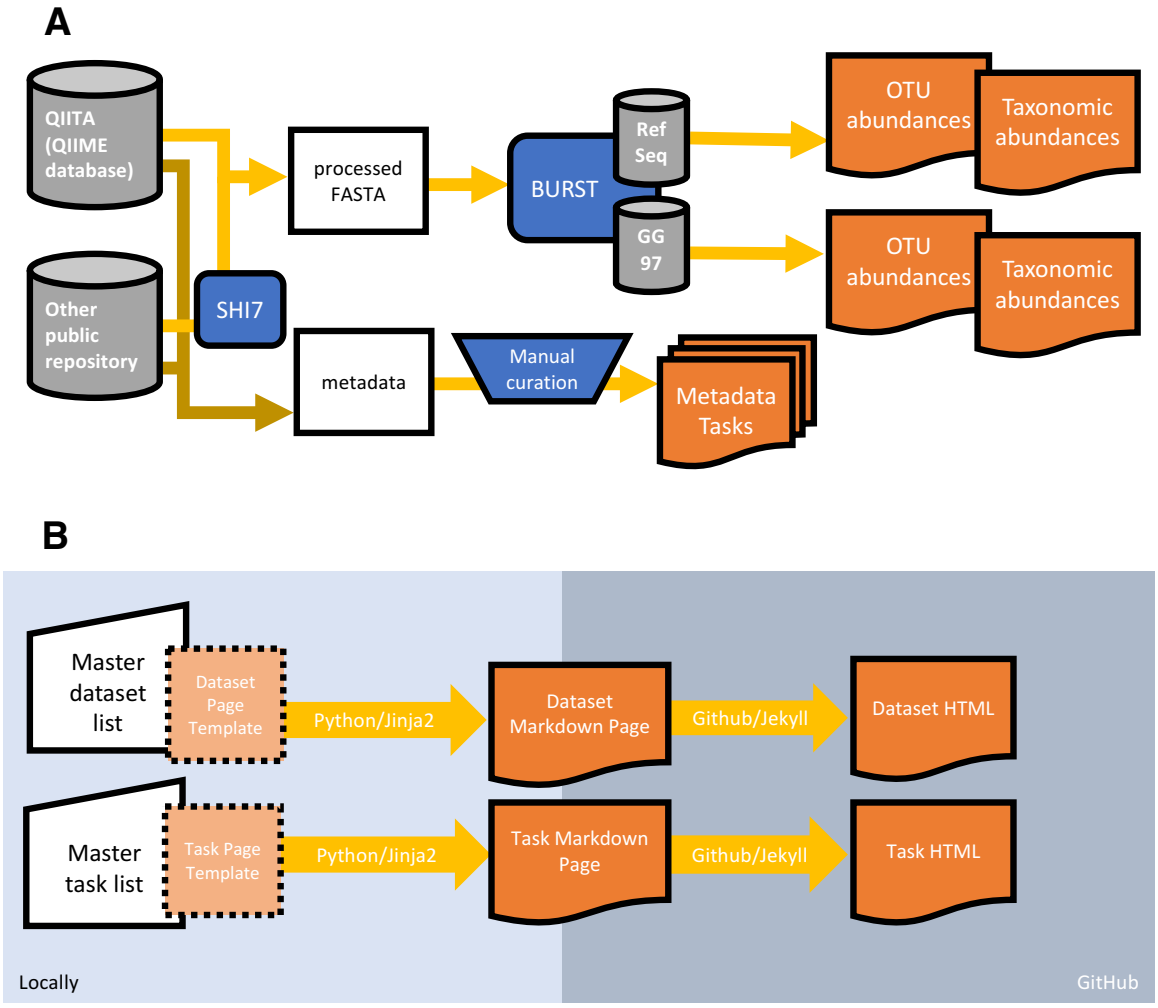


Figure 3.1. Workflow of data and website generation.

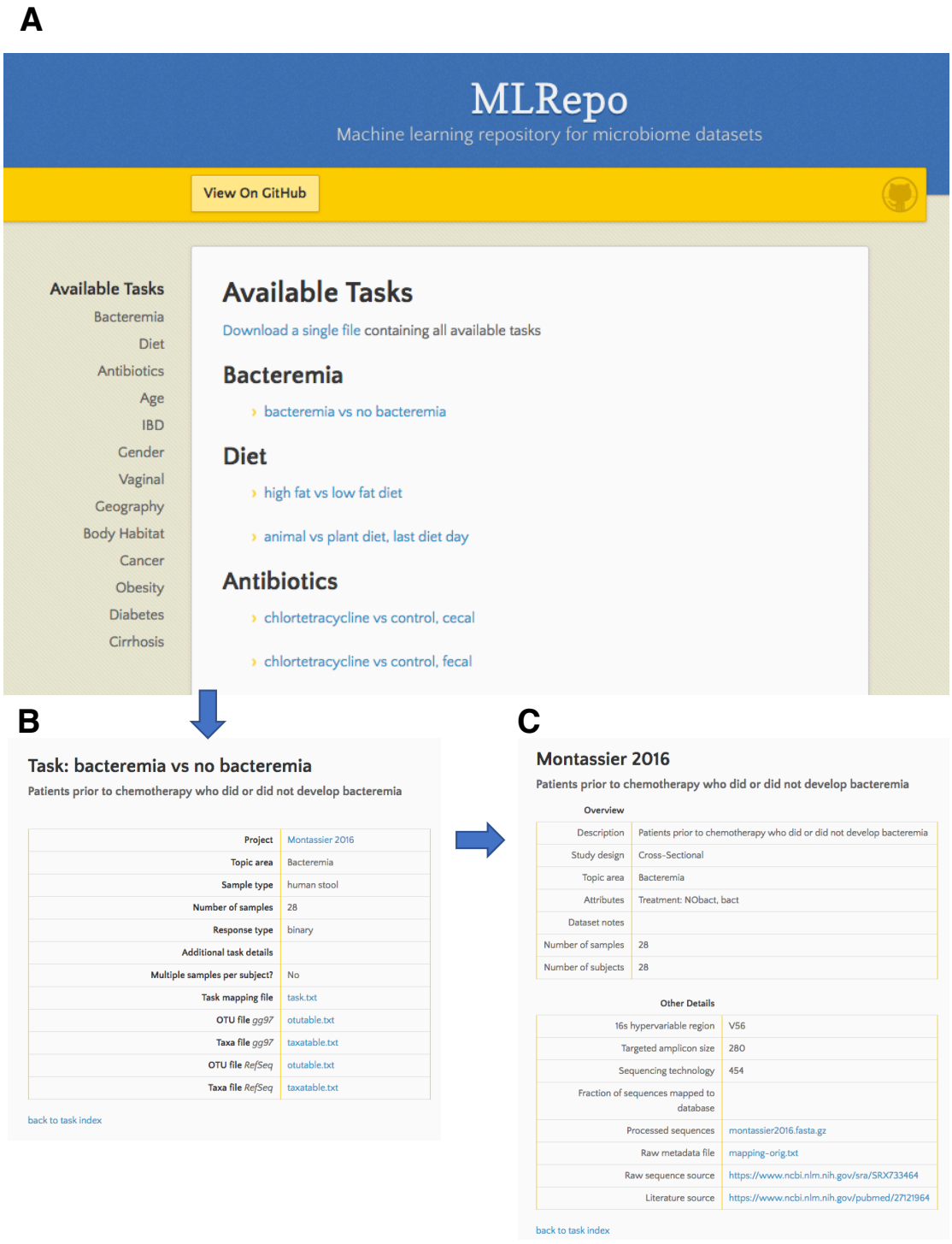
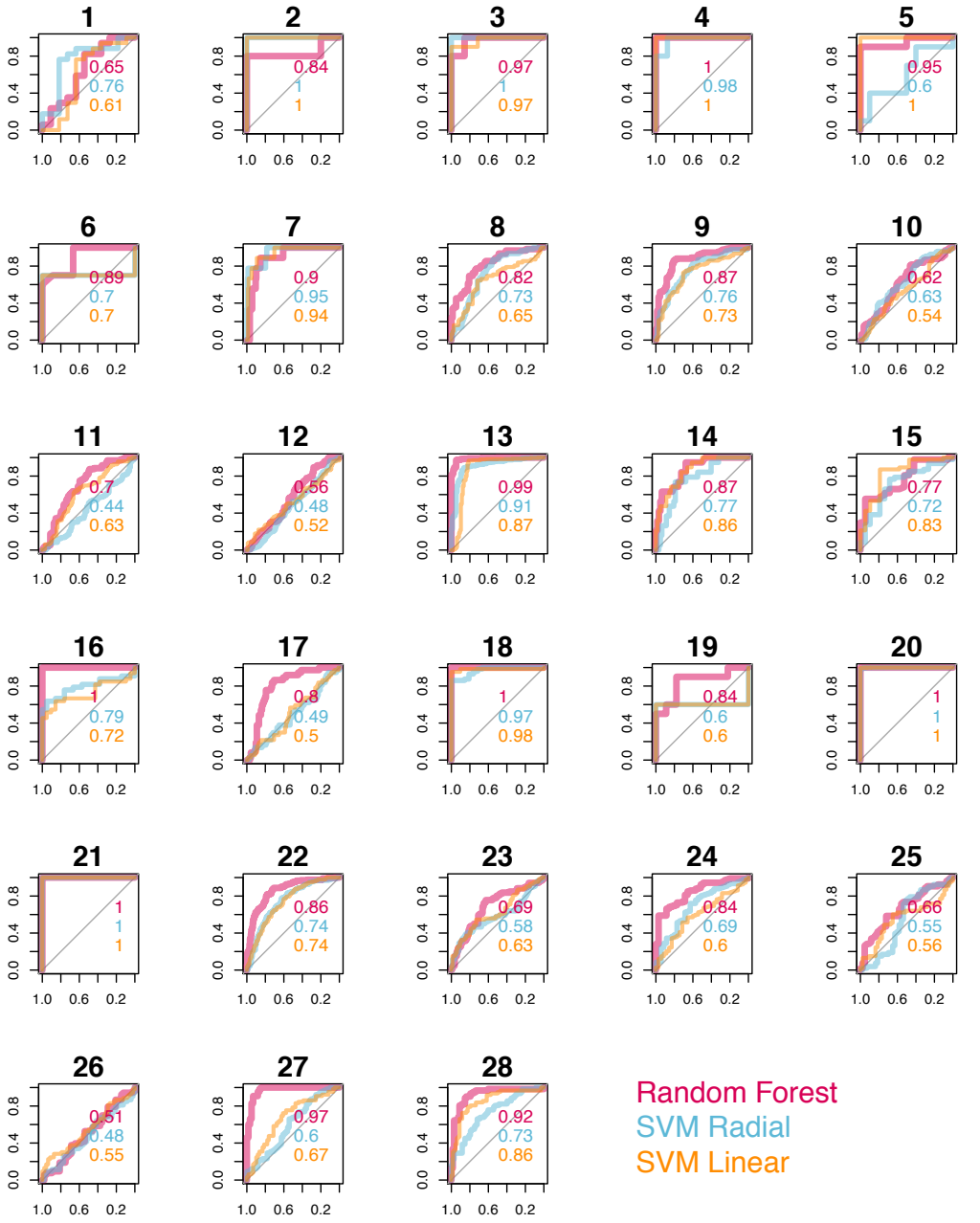
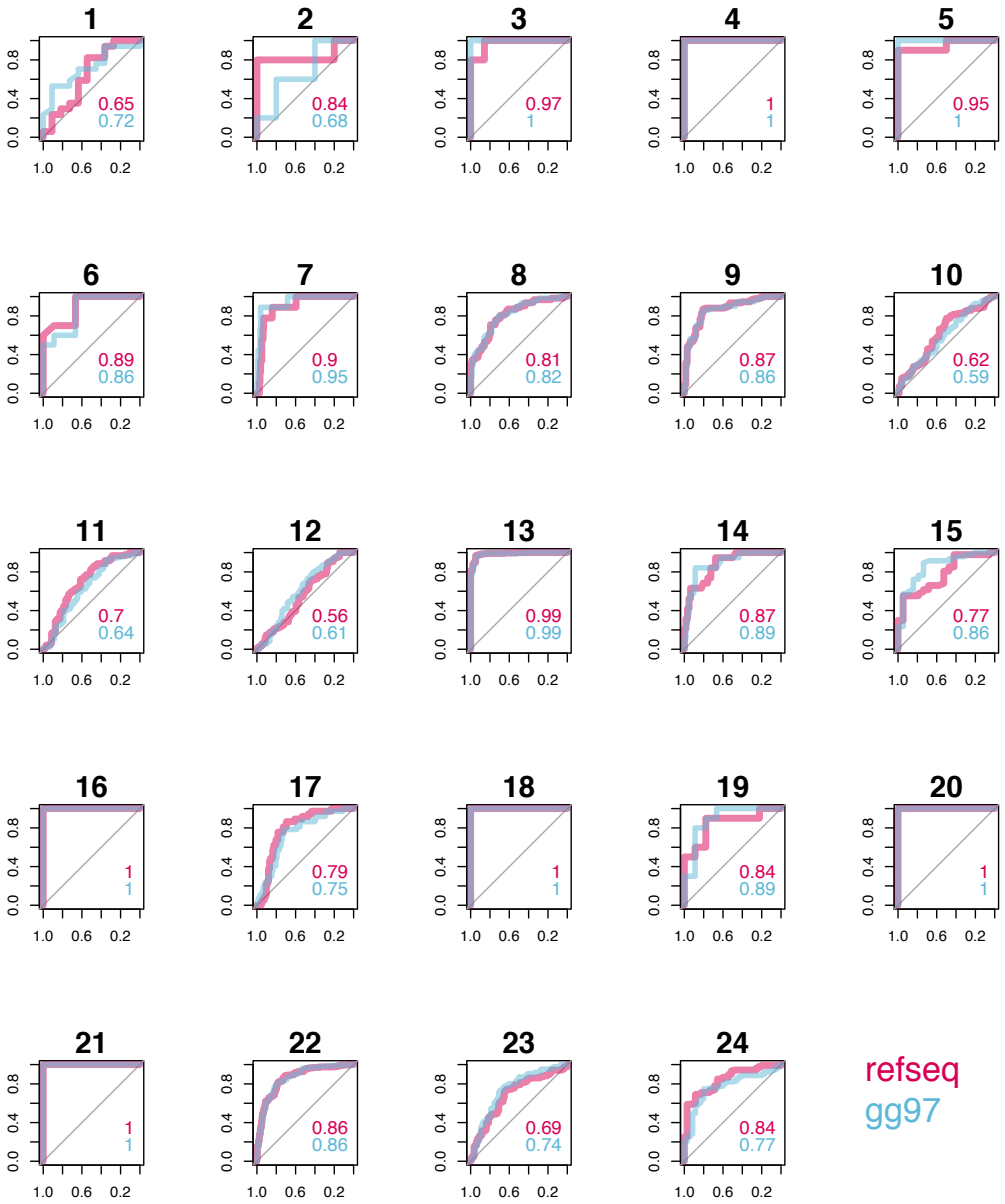


Figure 3.2. Website screenshots of MLRepo homepage, task, and dataset pages.



1 bacteremia vs no bacteremia
2 high fat vs low fat diet
3 chlortetracycline vs control, cecal
4 chlortetracycline vs control, fecal
5 penicillin vs vancomycin, cecal
6 penicillin vs vancomycin, fecal
7 elderly vs young
8 control vs cd, ileum
9 control vs cd, rectum
10 male vs female, stool
11 white vs black, vaginal
12 black vs hispanic, vaginal
13 low vs high nugent category
14 healthy vs cd, stool
15 healthy vs uc, stool
16 malawi vs venezuela, adults only
17 male vs female, usa
18 us vs malawi, adults only
19 animal vs plant diet, last diet day
20 gastrointestinal vs oral
21 stool vs tongue
22 subgingival vs supragingival plaque
23 healthy vs tumor biopsy, paired
24 lean vs obese, mz/dz/mom
25 normal vs diabetes glucose tolerance
26 impaired vs diabetes glucose tolerance
27 healthy vs type 2 diabetes
28 healthy vs cirrhosis

Figure 3.3. ROCs comparing random forest and SVM with different kernels.



1 bacteremia vs no bacteremia
2 high fat vs low fat diet
3 chlortetracycline vs control, cecal
4 chlortetracycline vs control, fecal
5 penicillin vs vancomycin, cecal
6 penicillin vs vancomycin, fecal
7 elderly vs young
8 control vs cd, ileum
9 control vs cd, rectum
10 male vs female, stool
11 white vs black, vaginal
12 black vs hispanic, vaginal
13 low vs high nugent category
14 healthy vs cd, stool
15 healthy vs uc, stool
16 malawi vs venezuela, adults only
17 male vs female, usa
18 us vs malawi, adults only
19 animal vs plant diet, last diet day
20 gastrointestinal vs oral
21 stool vs tongue
22 subgingival vs supragingival plaque
23 healthy vs tumor biopsy, paired
24 lean vs obese, mz/dz/mom

Figure 3.4. ROCs comparing NCBI RefSeq and Greengenes 97 databases.

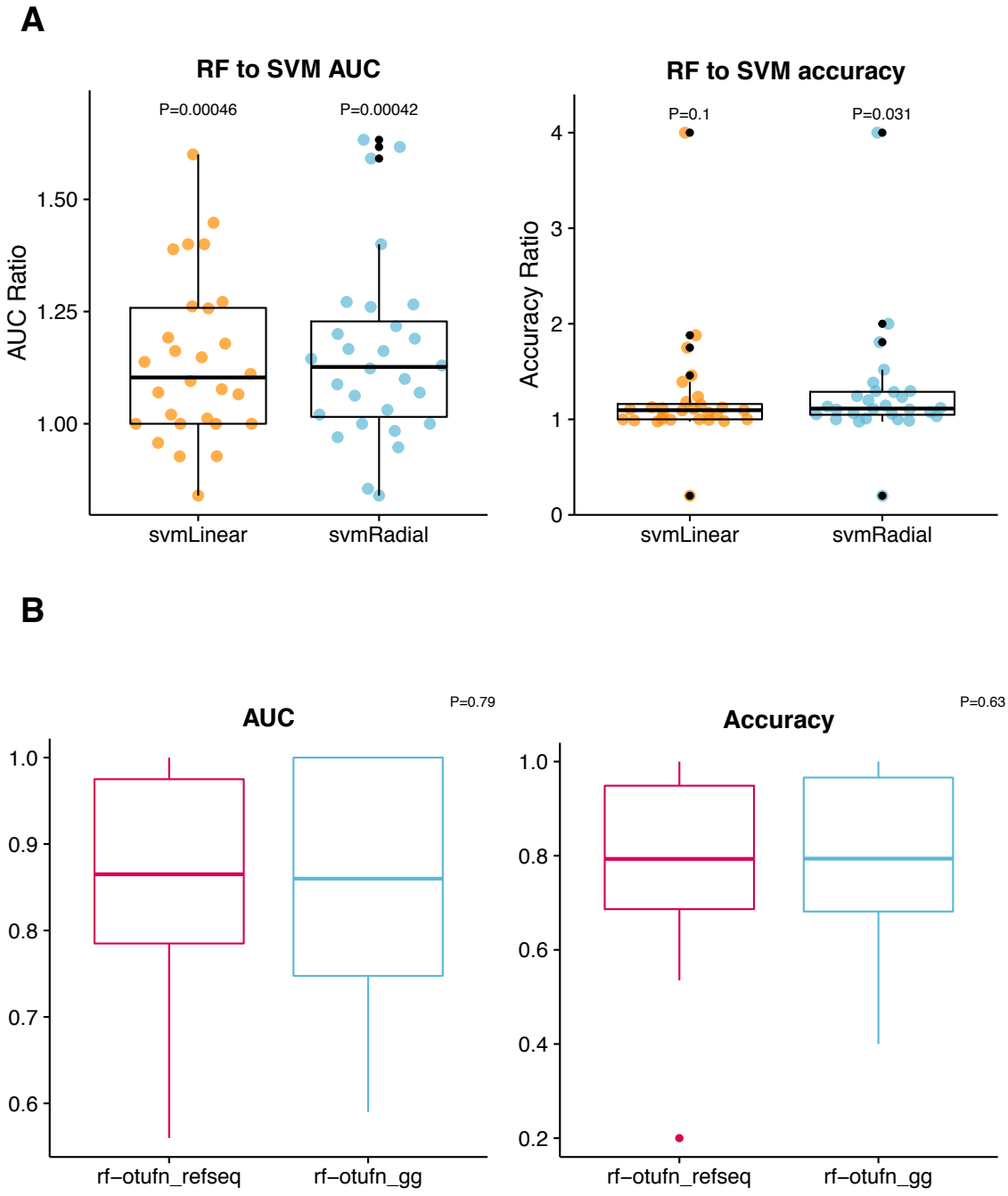


Figure 3.5. Summary statistics of framework and database comparisons.
(A) AUCs presented as ratio of RF AUC to SVM AUC. (B) AUCs presented as ratio of RefSeq database AUC to Greengenes database AUC.

Chapter 4 Figures

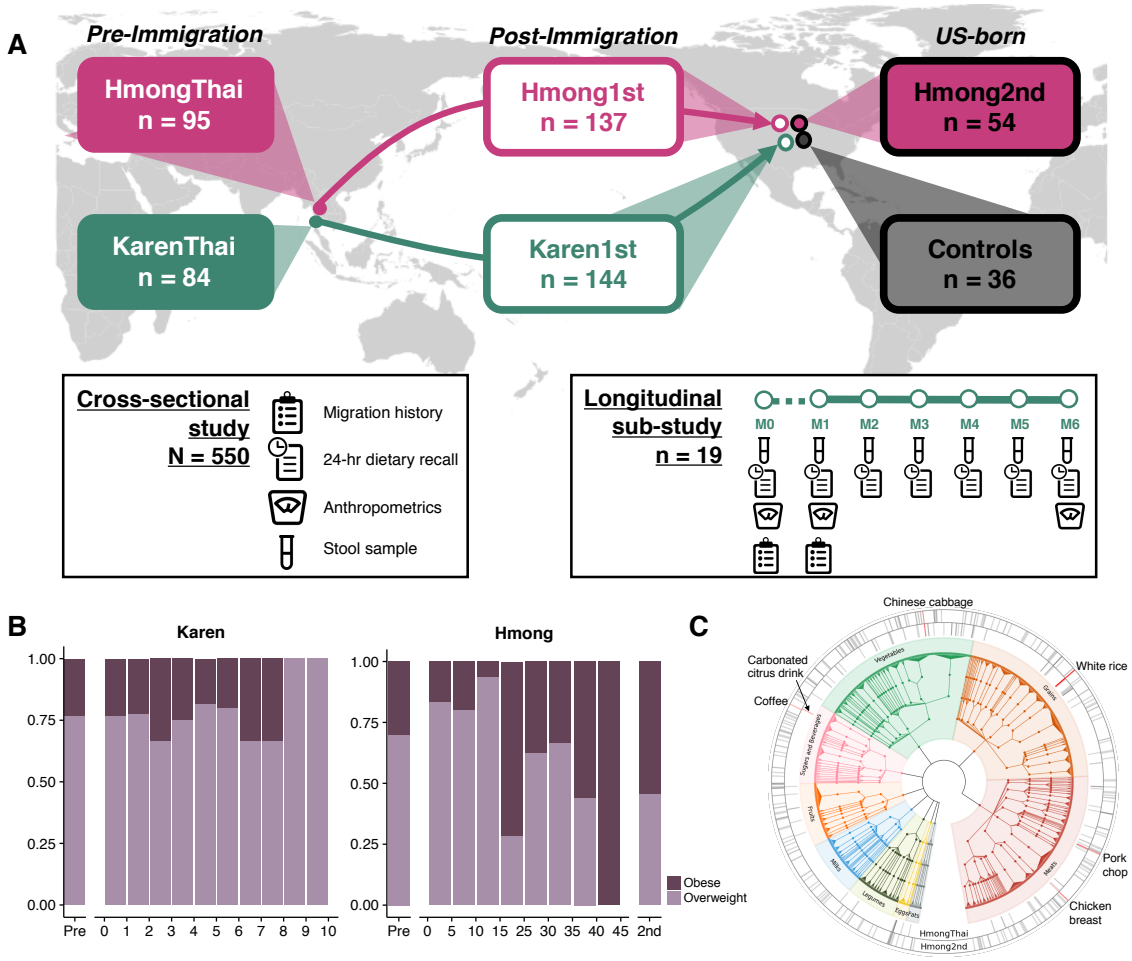


Figure 4.1. Assembly of a multi-generational Asian American cohort, while accounting for BMI and diet

(A) Experimental design for cross-sectional and longitudinal cohorts. (B) Ratios of overweight-to-obese individuals across sample groups and over time in the U.S., separated by ethnicity due to differences in time in years. Sample sizes are not evenly distributed across time in the U.S. (C) Hmong in Thailand (n = 43) and second-generation Hmong (n = 41) (ages 20-40) diet diversity, as seen across tree-based food items. Bars denote unique foods, with prevalence of foods reported averaged within HmongThai or Hmong2nd and displayed as a gradient. Items highlighted in red denote the most prevalent vegetables, sweets and beverages, grains, and meats reported within sample groups. Full descriptions of highlighted foods: Coffee, brewed, regular; Carbonated citrus fruit drink; Chinese cabbage or Bok Choy family, raw; Rice, white, no salt or fat added; Pork chop, broiled, baked, or grilled, lean only eaten; Chicken breast, roasted, skin not eaten.

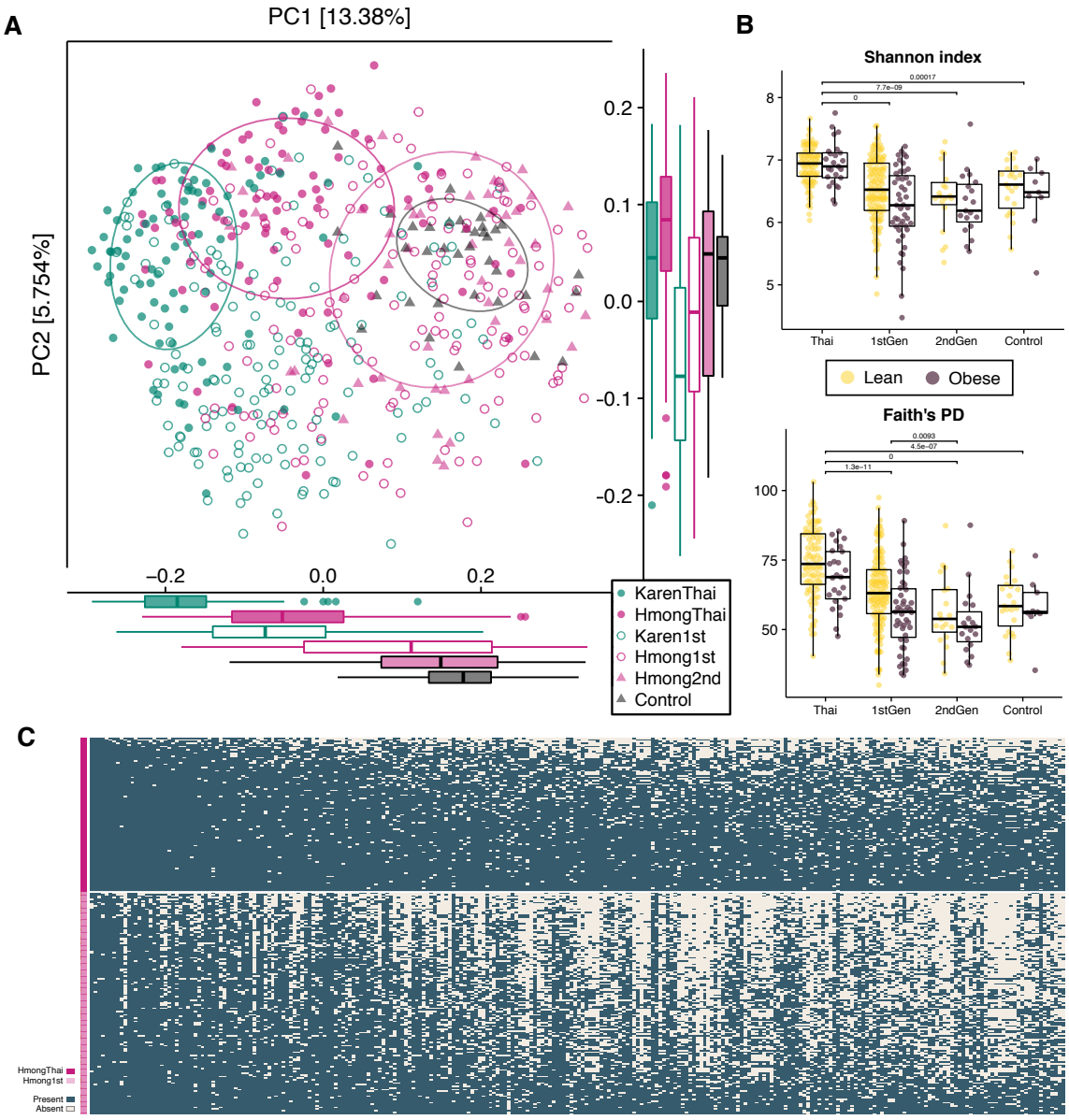


Figure 4.2. Loss of diversity and native bacterial taxa with time in the U.S.
(A) PCoA (unweighted UniFrac) of gut bacterial communities reveals that phylogenetic variation is strongly explained by sample group (ANOSIM R=0.25, P=0.001). 95% standard error ellipses are shown around HmongThai, KarenThai, Hmong2nd, and Controls. (B) Alpha diversity of obese and lean individuals across sample groups, in Shannon's Diversity index and Faith's Phylogenetic Distance (PD). P-values denote significantly different groups using pairwise tests of sample groups with pooled BMI classes (Tukey's HSD, $\alpha = 0.05$). Using a two-way ANOVA analysis with BMI class and sample group as covariates, we found that obesity is significantly lower across all groups ($P = 0.0044$). (C) Prevalence of operational taxonomic units (OTUs) in HmongThai and Hmong1st, sorted by prevalence in HmongThai and by richness within sample group. OTUs shown are found in > 75% of HmongThai samples (See Table 4.4).

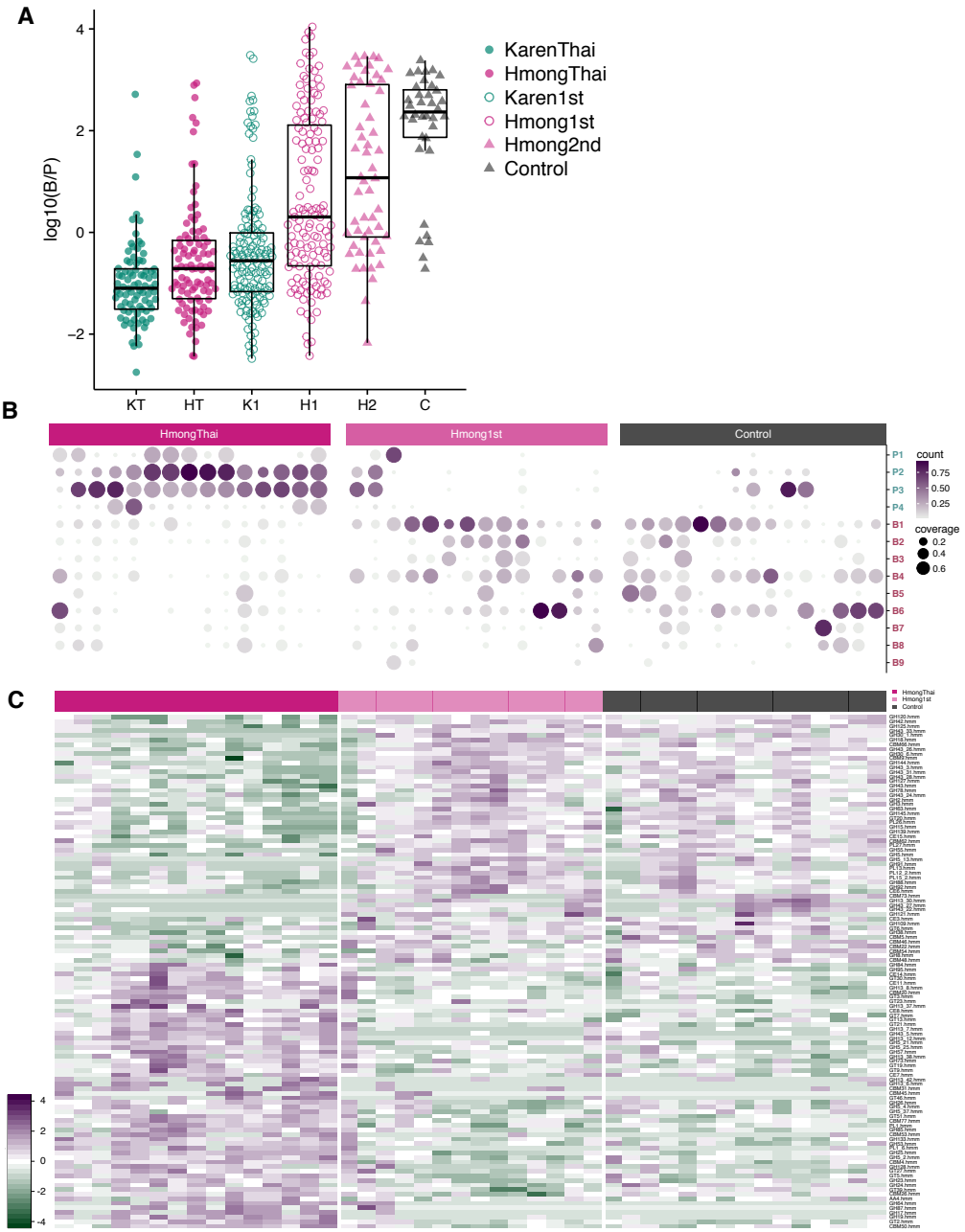


Figure 4.3. Bacteroides and Prevotella strain diversity and abundances

(A) Ratio of Bacteroides to Prevotella relative abundances, log transformed (B/P). Significant contributions from covariates that define the sample groups classes: Resident.Continent, P=3.4e−13; Birth.Continent, P=0.00085; Ethnicity, P=5.5e−12 (unbalanced two-way ANOVA). (KT=KarenThai; HT=HmongThai; K1=Karen1st; H1=Hmong1st; H2=Hmong2nd; C=Controls). (B) Bacteroides and Prevotella strain diversity in 44 samples across HmongThai, Hmong1st (who have lived in the U.S. for more than 30 years), and Controls. Strains were selected if coverage > 50% in at least one sample. Hierarchical clustering of strains and samples within group is based on relative abundances and coverage < 1% of a strain within person is considered not present (not plotted). See Table 4.5 for strain names. (C) CAZymes with significantly different relative abundances among HmongThai, Hmong1st (who have lived in the U.S. for more than 30 years), and Controls (Mann Whitney U test, FDR-corrected q < 0.05).

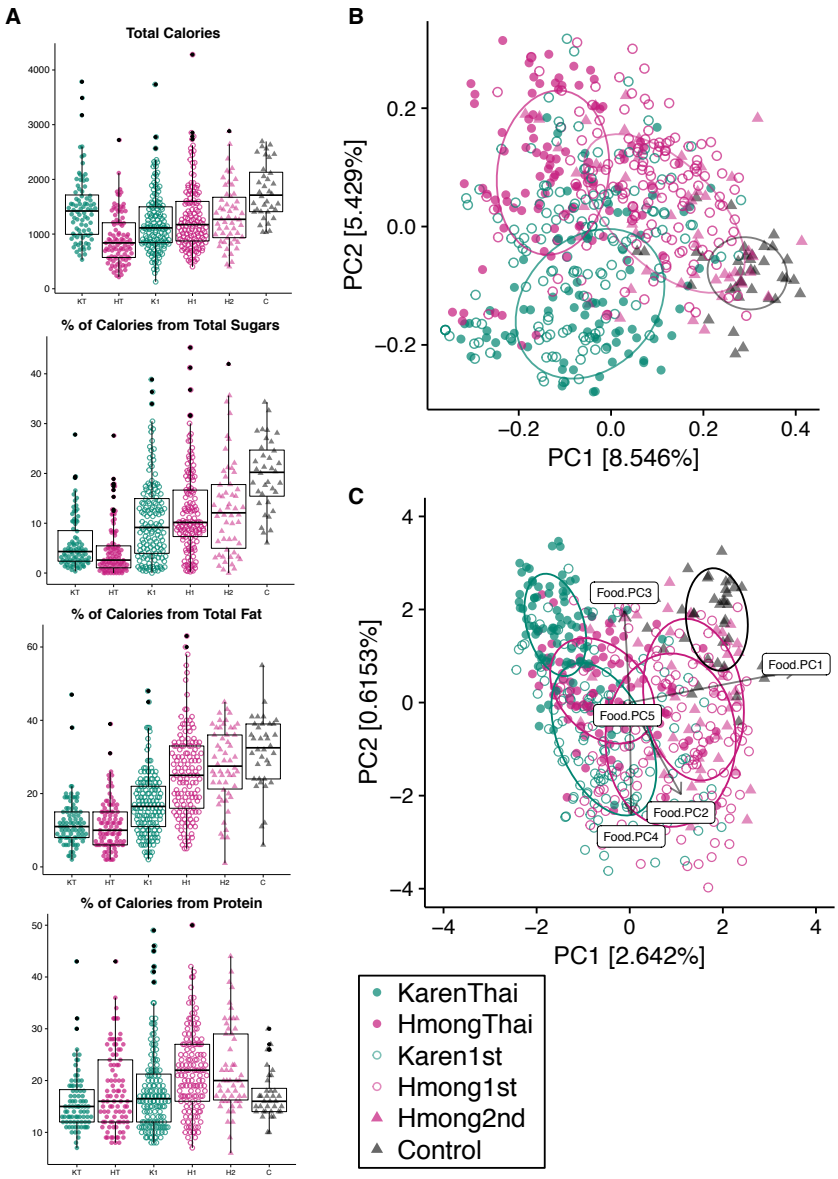


Figure 4.4. Dietary acculturation is detectable using novel food tree and partially explains microbiome variation

(A) Comparison of macronutrients consumption levels across sample groups. Ethnicity is significantly associated with calories (P=3.4e−05), sugars (P=0.00023), fat (P=1.3e−07), protein (P=3.2e−07), whereas current continent of residency is associated with sugar (P=1.3e−16), fat (P=7.1e−24), and protein consumption (P=5.7e−05), and birth continent is only associated with Fat consumption (P=0.0081) (unbalanced two-way ANOVA). (HT=HmongThai; KT=KarenThai; H1=Hmong1st; K1=Karen1st; H2=Hmong2nd; C=Controls). **(B)** PCoA of unweighted UniFrac diet-based distances reveal significant clustering by sample group (ANOSIM R=0.29, P=0.001), with Hmong2nd now clustering with Hmong1st instead of with Controls as reported with microbiome-based distances. Dietary acculturation can be seen along PC1, as it is significantly correlated with years spent in the U.S. (ρ=0.56, P=2.2e-16). **(C)** Redundancy analysis (RDA) of the unweighted Unifrac microbiome-distances constrained by the first 5 principal coordinates of the PCoA of unweighted Unifrac food-distances. The resulting RDA explains 16.8% of the total variation explained by PC1 and PC2 of the microbiome PCoA (Figure 4A).

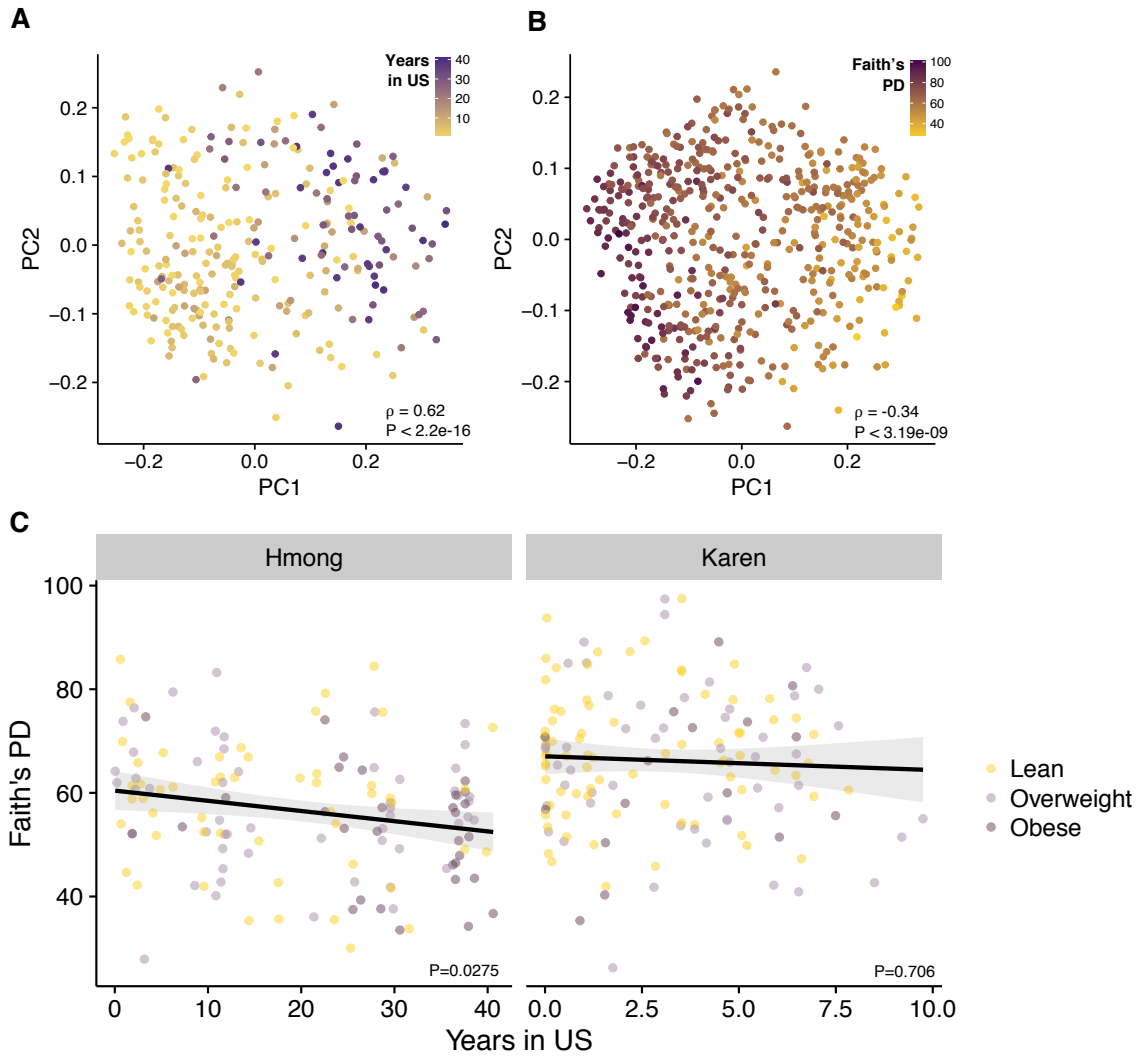


Figure 4.5. Gut biodiversity decreases with time spent in the U.S.
(A) Unweighted Unifrac PCoA of gut microbiomes of first-generation Hmong and Karen participants (N = 281), colored by years spent in the U.S. which ranges from 1 day to 40.6 years. PC1 is strongly correlated with the amount of time spent in the U.S. ($\rho = 0.62$, $p < 2.2e\text{-}16$). (B) Unweighted Unifrac PCoA of gut microbiomes of cross-sectional participants (N=550), colored by Faith's Phylogenetic Diversity. PC1 is negatively correlated with phylogenetic richness ($\rho = -0.34$, $p < 3.19e\text{-}09$). (C) In first-generation Hmong, diversity significantly decreases over time in the U.S. (multiple regression: Years in US $\beta = -0.18$, $P = 0.0275$; BMI $\beta = -0.05$, $P = 0.81$), but a significant association is not observed in first-generation Karen (Years in US $\beta = -0.17$, $P = 0.71$; BMI $\beta = -0.27$, $P = 0.28$). Interaction terms were not significantly associated with diversity, and were removed from the model.

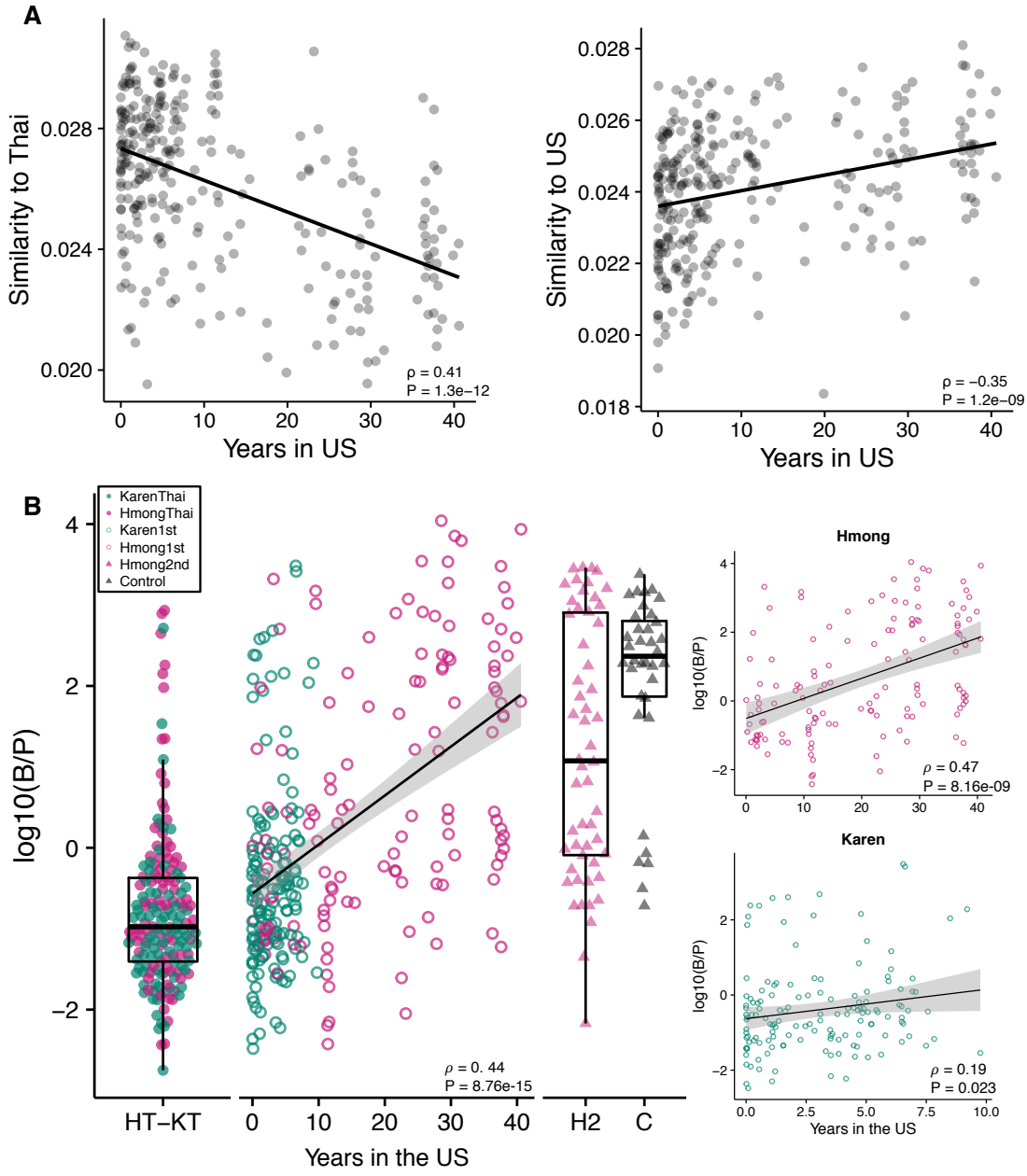


Figure 4.6. Prevotella displacement is observable within one decade in the U.S.
(A) Similarity (1 / Aitchison's distance) of microbiomes relative to Thai-based groups and to Controls. (B) Log ratio of *Bacteroides* to *Prevotella* of first-generation groups are significantly correlated to years spent in the U.S. (ρ = 0.44, P = 8.76e-15). Significantly correlated trends persist after stratification by ethnicity (Hmong ρ = 0.47, P = 8.16e-19; Karen ρ = 0.19, P = 0.023). (HT=HmongThai; KT=KarenThai; H2=Hmong2nd; C=Controls; 0-40=Years spent in the U.S. by Hmong1st and Karen1st).

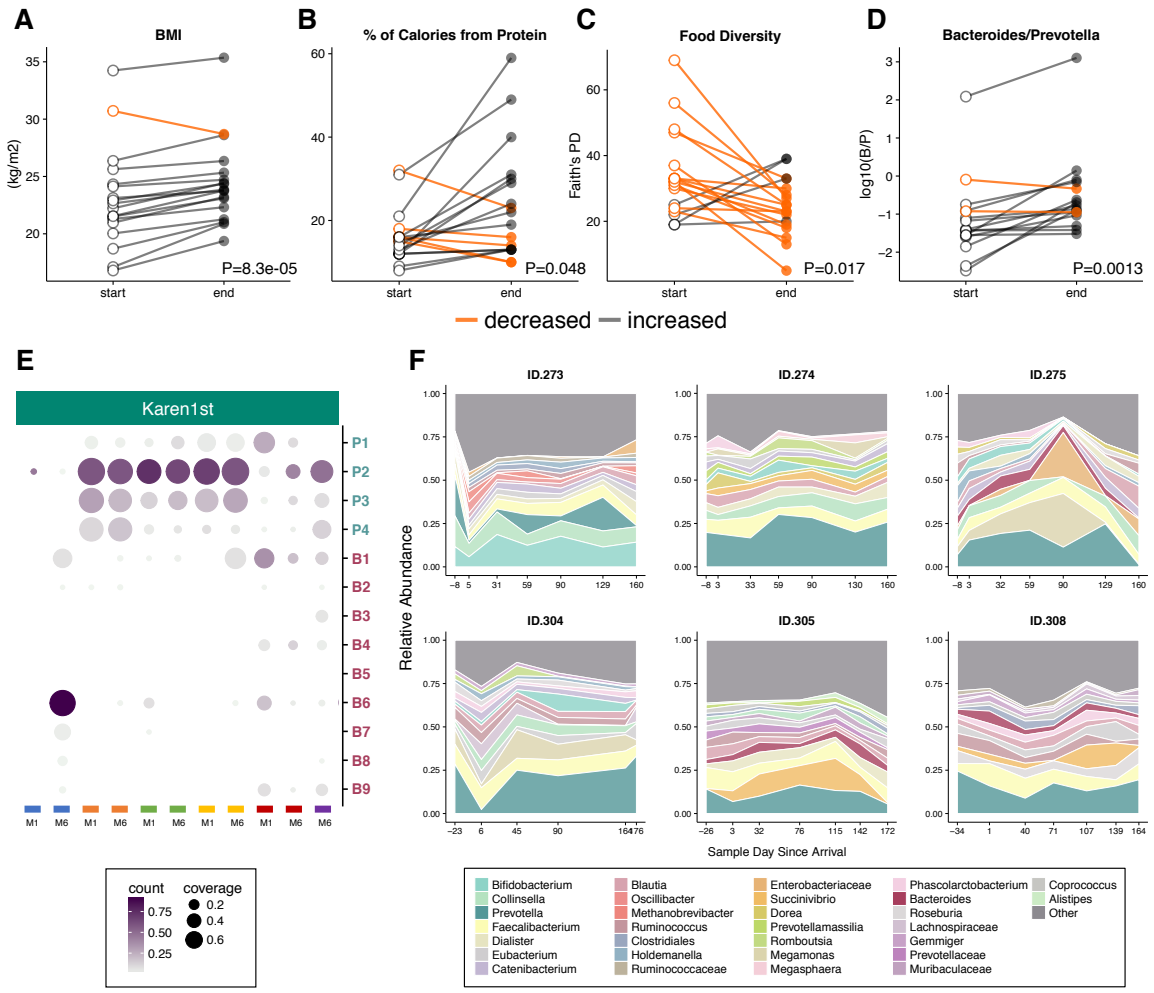


Figure 4.7. Longitudinal changes immediately pre- and post-arrival to the U.S.
(A) Comparison of per-participant changes between first and last months of the study in BMI, (B) protein consumption, (C) dietary diversity (Faith's PD), and (D) *Bacteroides* to *Prevotella* ratios (paired t-test, macronutrients adjusted for multiple comparisons using false discovery rate, $q < 0.05$). (E) *Bacteroides* and *Prevotella* strain profiles are mostly stable after 6 months. Samples (columns) from the same participant are denoted by color, and M1 and M6 correspond to Month 1 Sample and Month 6 Sample, respectively. Selected strains are identical to Figure 4B (at least 50% coverage per sample across N=55 samples, see Table 4.5). (F) Taxonomic area charts of relative abundances of dominant genera (other taxa not shown) in 6 individuals who began the longitudinal study while in a refugee camp in Thailand. First available samples were collected 6 to 34 days before departure, and second samples were collected 1 to 6 days after arrival to the U.S.

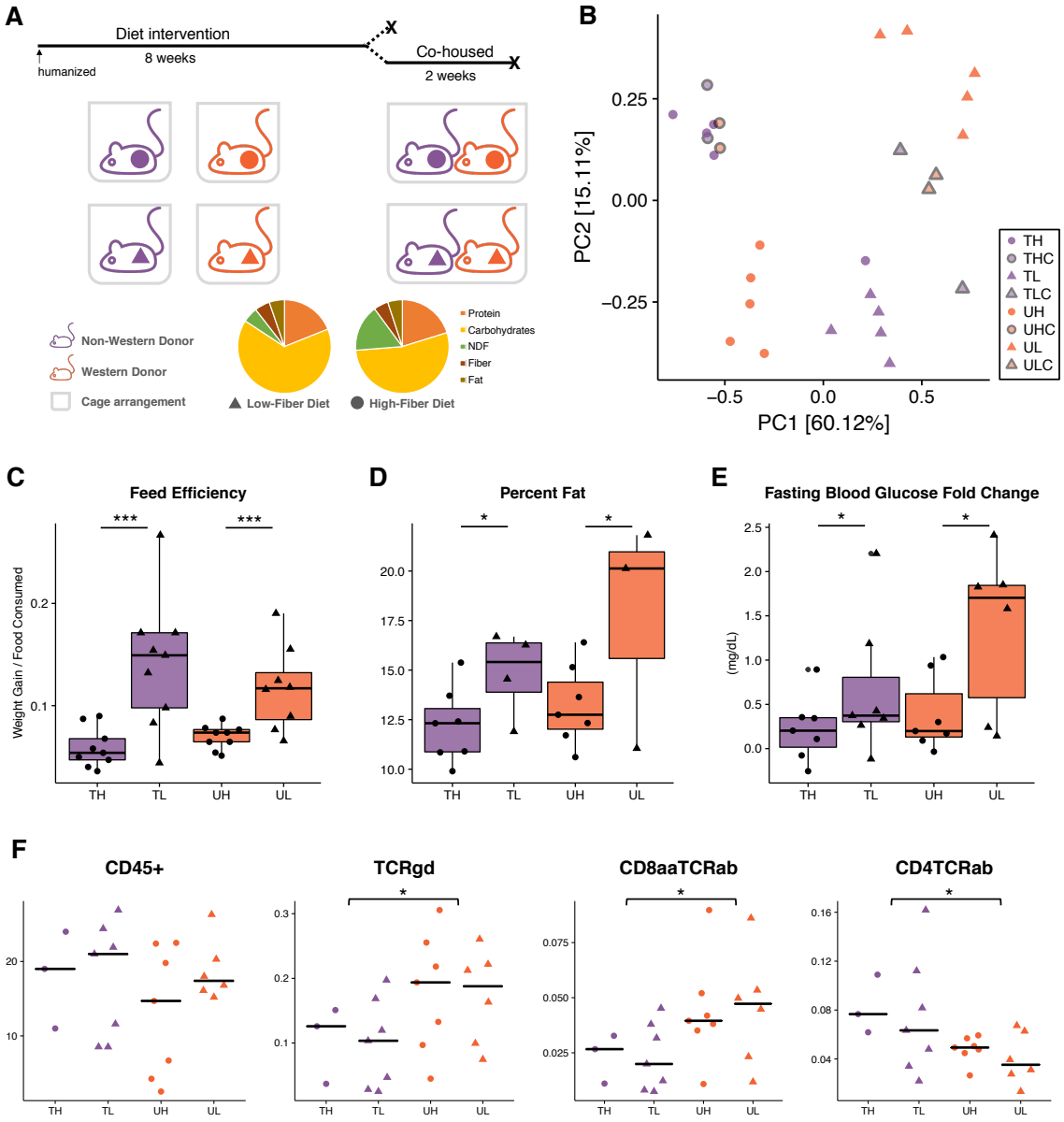


Figure 4.8. Western diet and Western microbiome induce deleterious responses in humanized mice

(A) Mouse experimental study design included Thai- (T) or U.S.-based (U) donors and high-fiber (H) or low-fiber (L) diet, resulting in four groups: TH=Thai-HighFiber; UH=US-HighFiber; TL=Thai-LowFiber; UL=US-LowFiber. (B) PCoA using unweighted Unifrac distances of mouse microbiomes at study endpoints (8 or 10 weeks). (C)-(E) Feed efficiency, percent fat, fasting blood glucose were compared between groups at the end of 8 weeks (two-way ANOVA). (F) Immune cell counts (cell population counts are normalized by counts of live CD45+ cells) in the small intestine intraepithelial lymphocytes (two-way ANOVA).

Chapter 4 Supplemental Figures

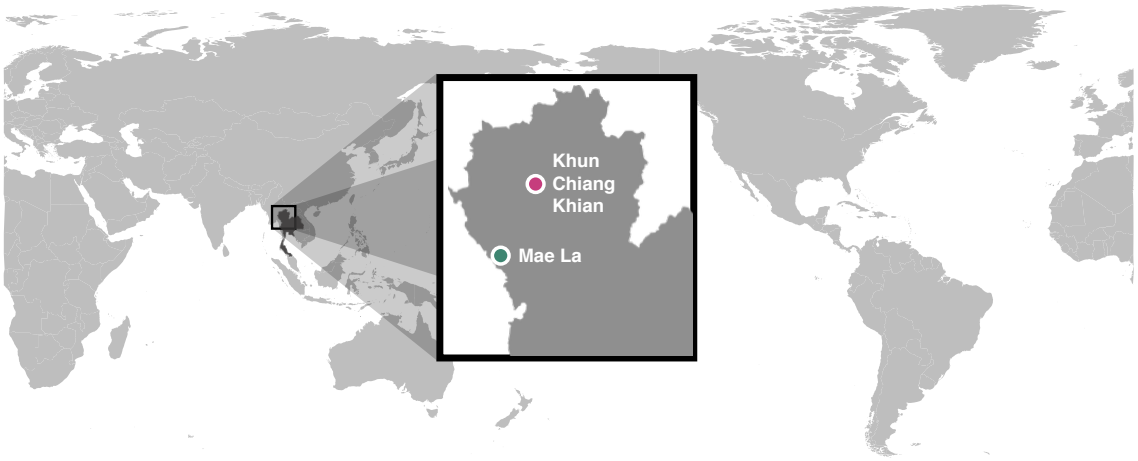


Figure 4.S1. Geographical locations of recruitment sites in Thailand. Related to Figure 4.1.
Khun Chang Khian in Chiang Mai province and Mae La camp in Tak Province.

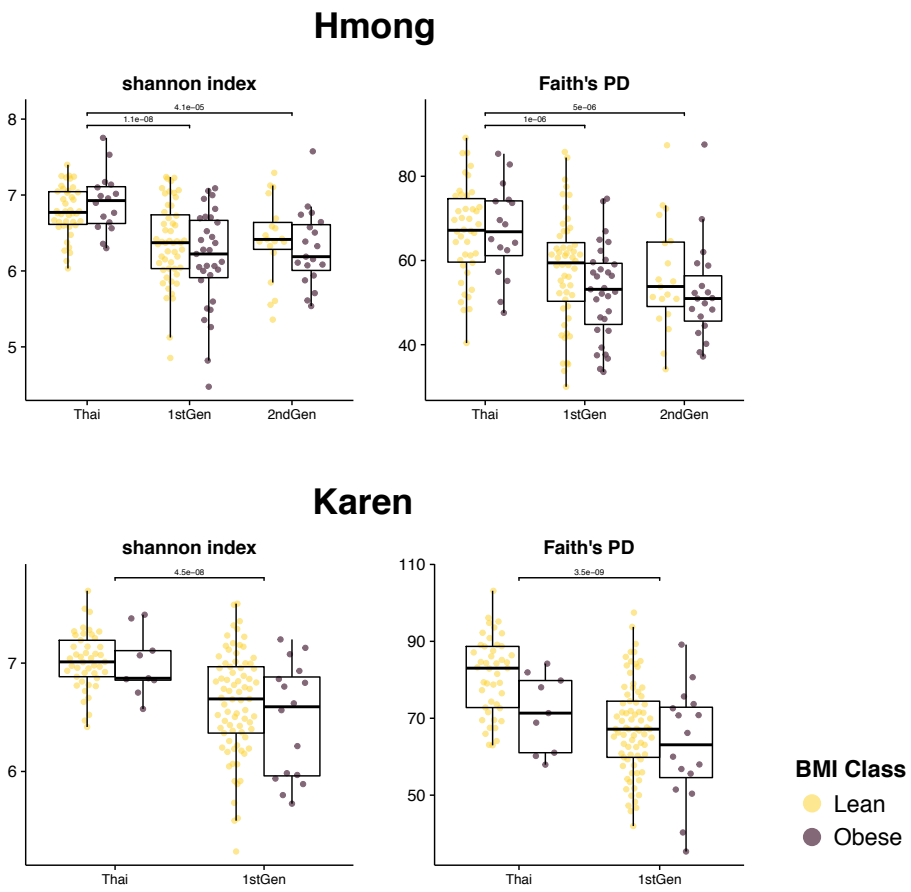


Figure 4.S2. Alpha diversity boxplots of obese and lean individuals, separated by ethnicity. Related to Figure 4.4.

Post-hoc analysis with Tukey's HSD test across sample groups.

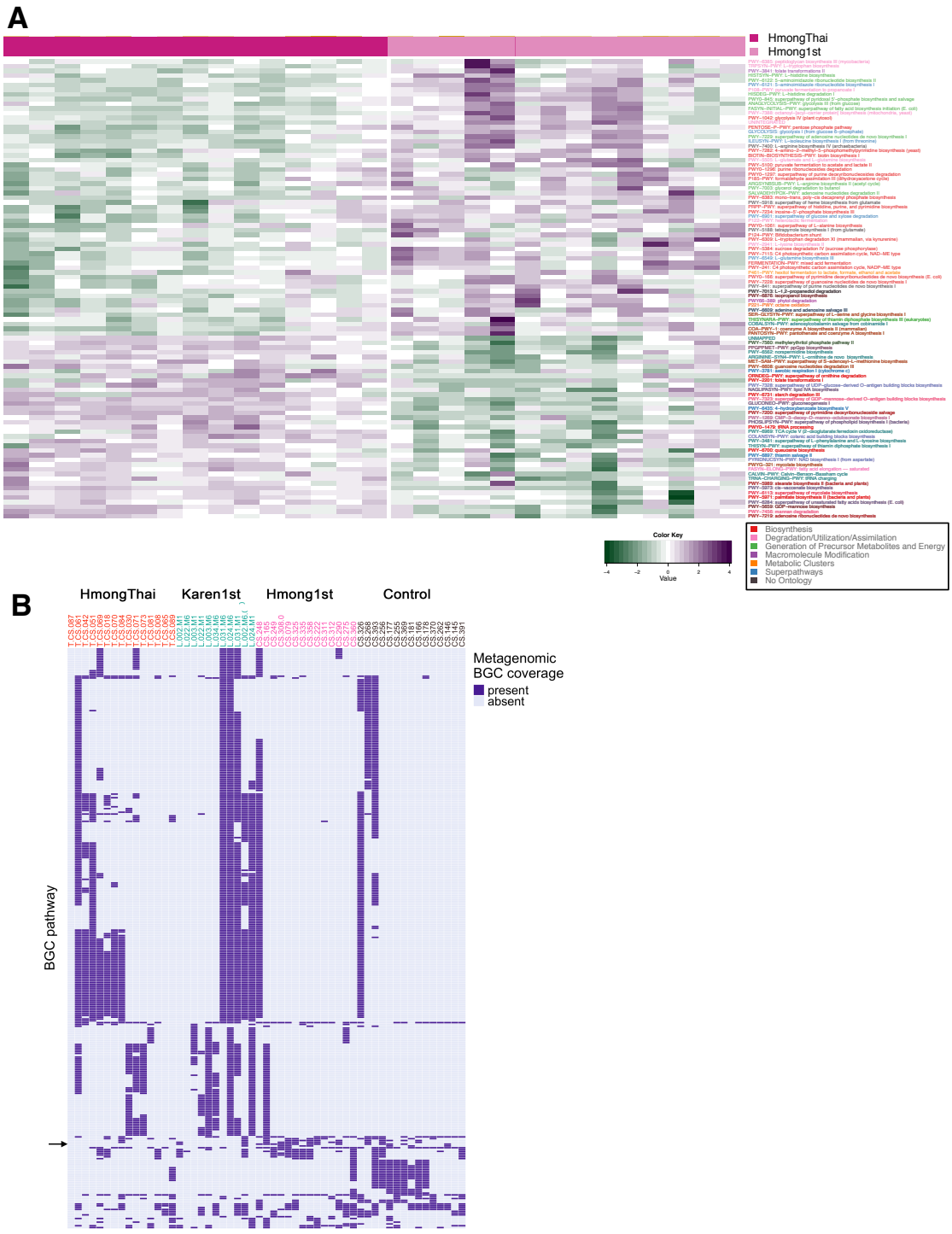


Figure 4.S3. Functional annotations. Related to Figure 4.4.
(A) Differentiated relative abundances of functional pathways between HmongThai and Hmong1st (asin-sqrt transformed abundances, ANOVA, FDR-corrected $q < 0.10$). (B) Prevalence of predicted biosynthetic gene clusters within sample groups. Comparison of groups with a looped Fisher's exact test reveals that a predicted bacteriocin cluster is enriched in Hmong1st relative to HmongThai ($P = 0.0002$, FDR-corrected $q < 0.10$).

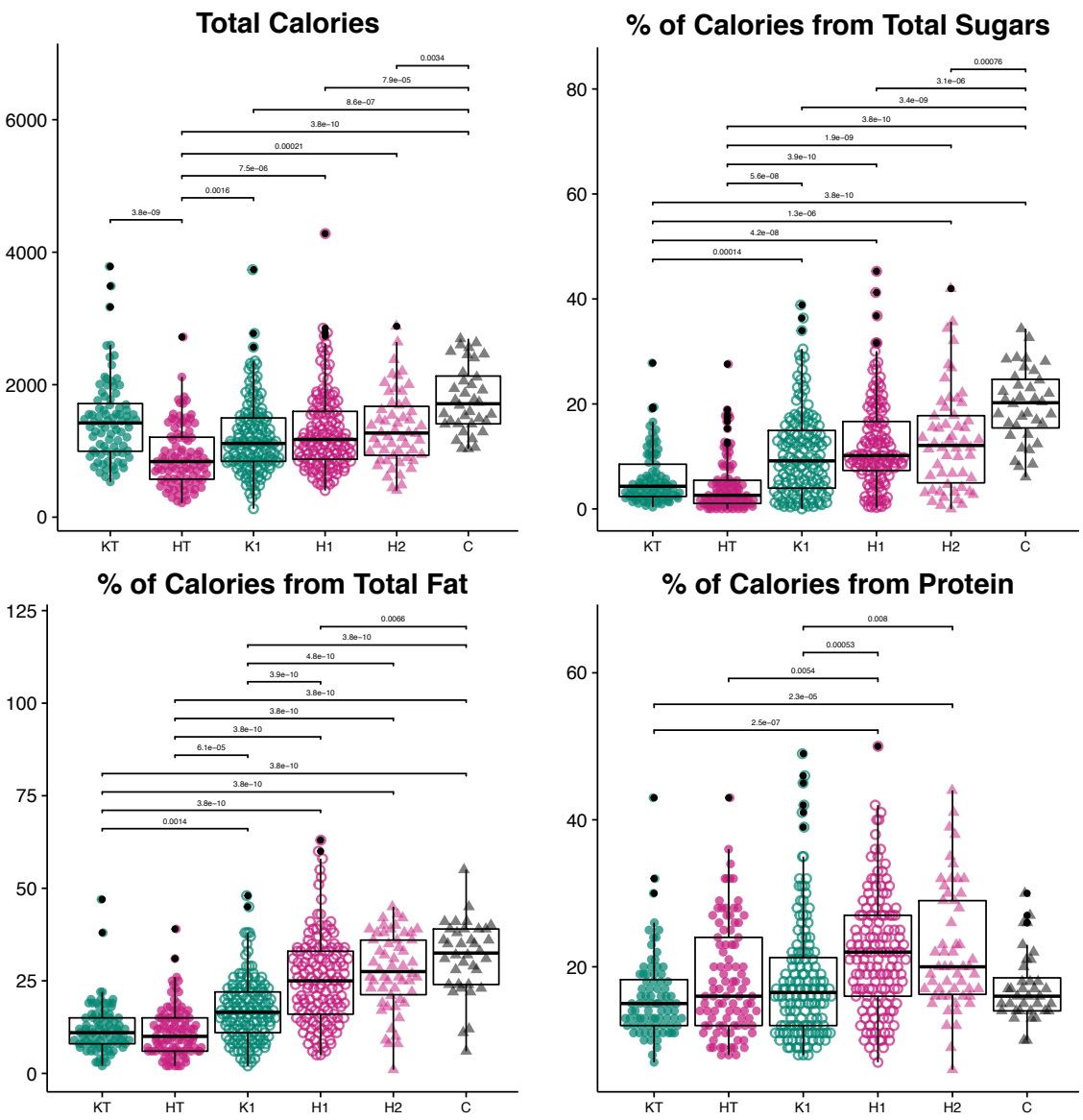


Figure 4.S4. Macronutrient pairwise comparisons. Related to Figure 4.4.
Pairwise comparisons with Tukeys' HSD, significant p-values < 0.05 are shown.

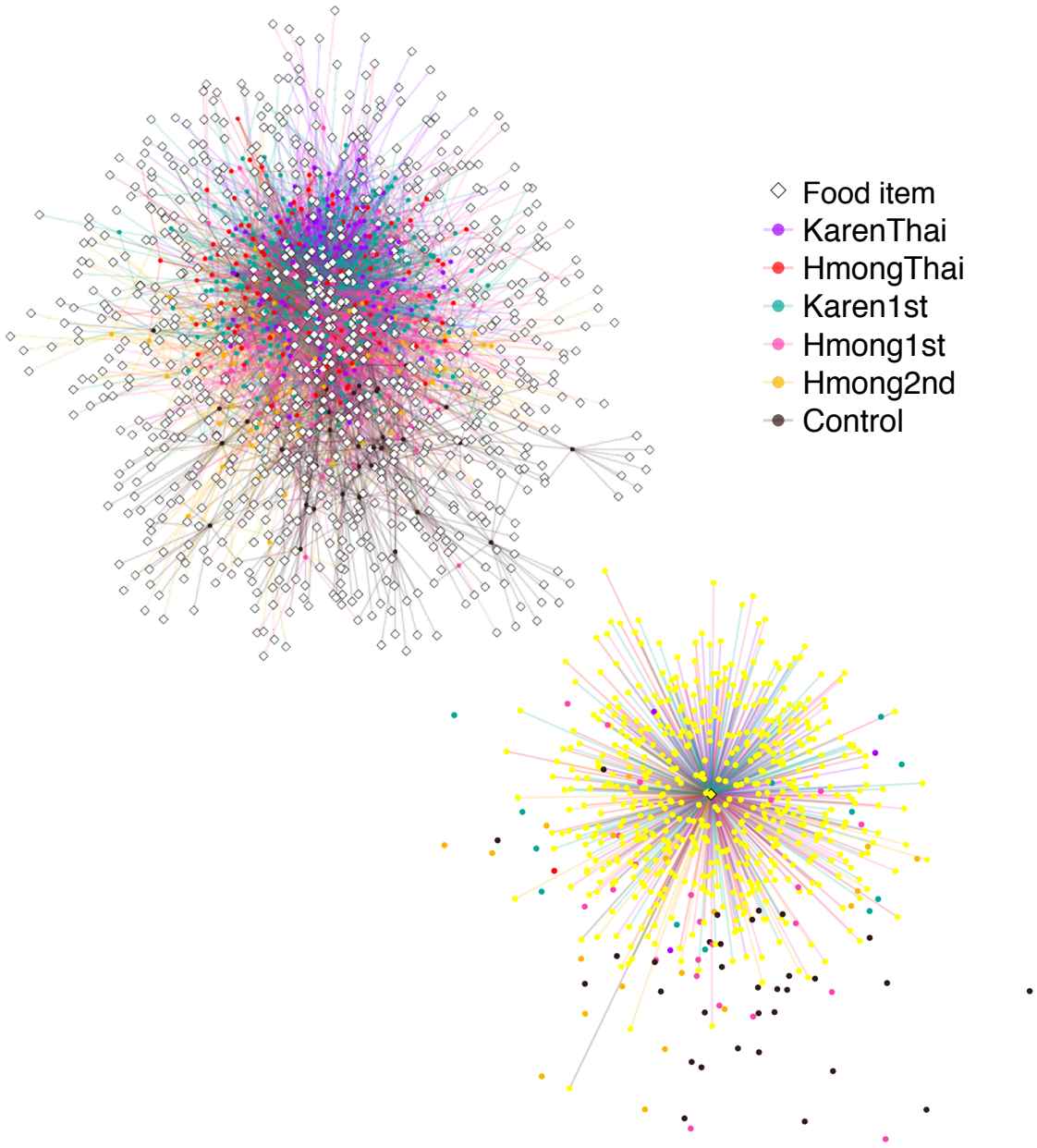


Figure 4.S5. Bipartite network of participant dietary records and food items. Related to Figure 4.4.

(A) Edges and participants are colored by sample group, and food items are shown as white-filled diamonds. (B) We highlight the high prevalence of rice consumption. Participants who consumed rice are denoted as yellow nodes and yellow edges connected to the centroid (rice), otherwise participants were colored by sample group.

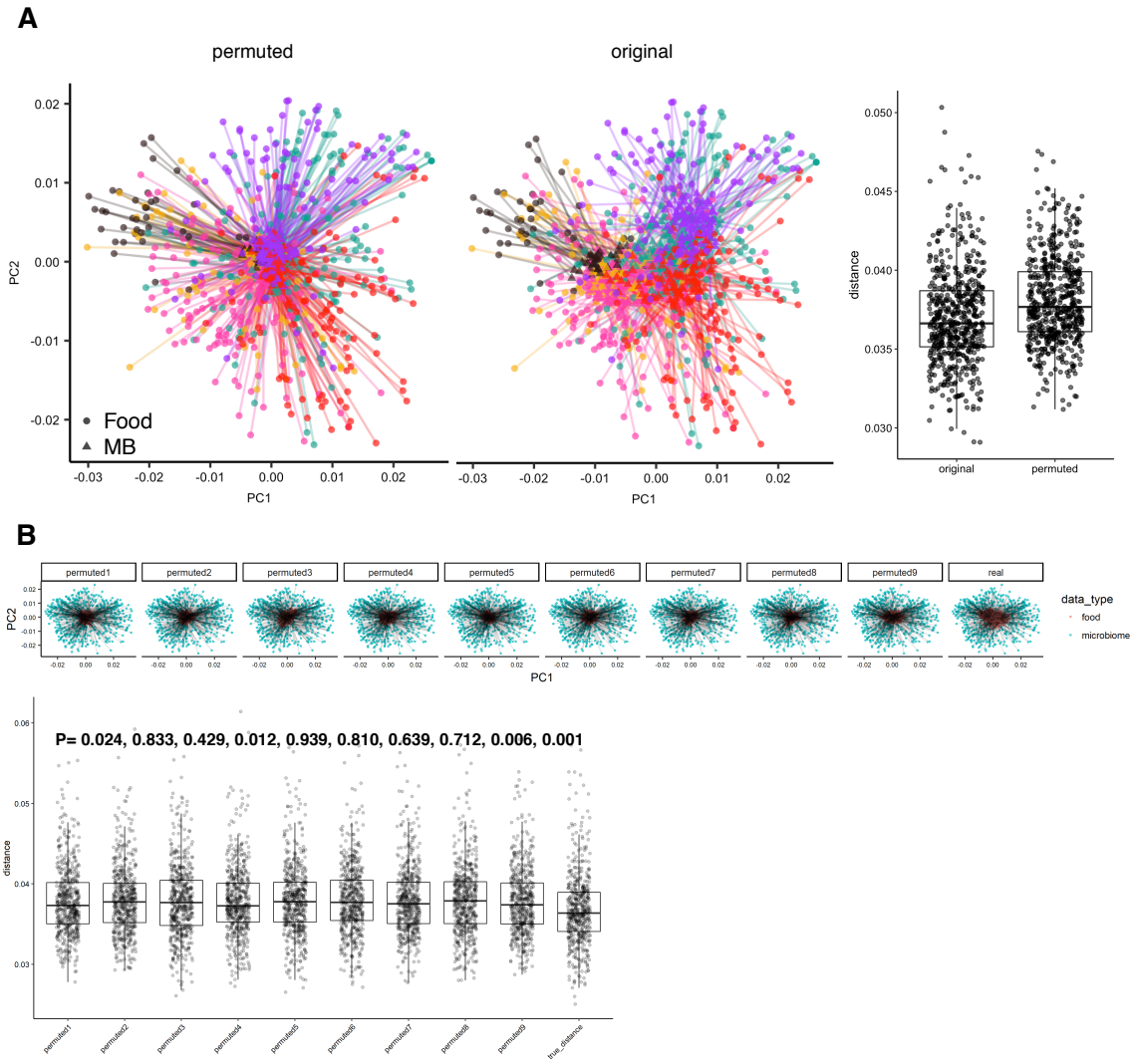


Figure 4.S6. Procrustes of diet and microbiome distances. Related to Figure 4.4.
(A) Procrustes permutation shows significant relatedness between individuals' food and microbiome profiles. Shown at left is the Procrustes PCoA for a representative permutation (median Procrustes sum of squares distance from 9 permutations) compared to the original data Procrustes PCoA, and at right are the individual multidimensional distances between each individuals' food and microbiome data after rotation. These points are significantly closer than expected by random chance (p = 1e-10, Mann Whitney U test). (B) All nine permutations of the Procrustes from panel A, including boxplots for the individual food-microbiome distances; p-values are generated from the *protest()* function in package "vegan" in R.

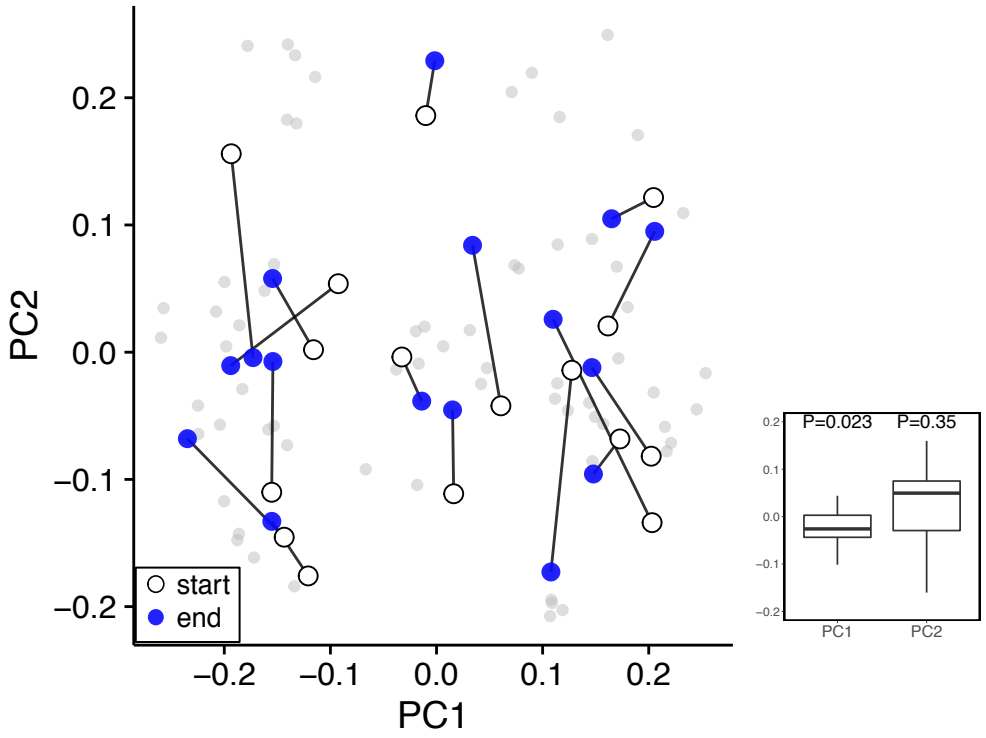


Figure 4.S7. PCoA of unweighted Unifrac distances of longitudinal samples. Related to Figure 4.7.

First and last month samples are highlighted and connected by participant, with all intermediate monthly samples in gray. Inset shows the within-individual changes along PC1 and PC2 from first to last months (one sample t-test with change in PC1 or PC2).

Bibliography

Aagaard, K., Riehle, K., Ma, J., Segata, N., Mistretta, T.-A., Coarfa, C., Raza, S., Rosenbaum, S., Van den Veyver, I., Milosavljevic, A., et al. (2012). A metagenomic approach to characterization of the vaginal microbiome signature in pregnancy. PLoS One *7*, e36466.

Aagaard, K., Ma, J., Antony, K.M., Ganu, R., Petrosino, J., and Versalovic, J. (2014). The placenta harbors a unique microbiome. Sci. Transl. Med. *6*, 237ra65.

Abadie, V., Discepolo, V., and Jabri, B. (2012). Intraepithelial lymphocytes in celiac disease immunopathology. Semin. Immunopathol. *34*, 551–566.

Abrahamsson, T.R., Jakobsson, H.E., Andersson, A.F., Björkstén, B., Engstrand, L., and Jenmalm, M.C. (2012). Low diversity of the gut microbiota in infants with atopic eczema. J. Allergy Clin. Immunol. *129*, 434–440, 440.e1–e2.

Abt, M.C., Osborne, L.C., Monticelli, L.A., Doering, T.A., Alenghat, T., Sonnenberg, G.F., Paley, M.A., Antenus, M., Williams, K.L., Erikson, J., et al. (2012). Commensal bacteria calibrate the activation threshold of innate antiviral immunity. Immunity *37*, 158–170.

Abubucker, S., Segata, N., Goll, J., Schubert, A.M., Izard, J., Cantarel, B.L., Rodriguez-Mueller, B., Zucker, J., Thiagarajan, M., Henrissat, B., et al. (2012). Metabolic reconstruction for metagenomic data and its application to the human microbiome. PLoS Comput. Biol. *8*, e1002358.

Adams, R.J., Heazlewood, S.P., Gilshenan, K.S., O'Brien, M., McGuckin, M.A., and Florin, T.H.J. (2008). IgG antibodies against common gut bacteria are more diagnostic for Crohn's disease than IgG against mannan or flagellin. Am. J. Gastroenterol. *103*, 386–396.

Adkins, B., Leclerc, C., and Marshall-Clarke, S. (2004). Neonatal adaptive immunity comes of age. Nat. Rev. Immunol. *4*, 553–564.

Ajslev, T.A., Andersen, C.S., Gamborg, M., Sørensen, T.I.A., and Jess, T. (2011). Childhood overweight after establishment of the gut microbiota: the role of delivery mode, pre-pregnancy weight and early administration of antibiotics. Int. J. Obes. *35*, 522–529.

Alexander, V.N., Northrup, V., and Bizzarro, M.J. (2011). Antibiotic exposure in the newborn intensive care unit and the risk of necrotizing enterocolitis. J. Pediatr. *159*, 392–397.

Al-Ghalith, G., and Knights, D. (2017). BURST enables optimal exhaustive DNA alignment for big data (Zenodo).

Al-Ghalith, G., and Knights, D. (2018). aKronyMer enables database-free metagenome comparison (Zenodo).

Al-Ghalith, G.A., Vangay, P., and Knights, D. (2015). The guts of obesity: progress and challenges in linking gut microbes to obesity. Discov. Med. *19*, 81–88.

Al-Ghalith, G.A., Hillmann, B., Ang, K., Shields-Cutler, R., and Knights, D. (2018). SHI7 Is a Self-Learning Pipeline for Multipurpose Short-Read DNA Quality Control. mSystems *3*.

Allen, M.L., Culhane-Pera, K.A., Call, K.T., and Pergament, S.L. (2011). Partners in research: curricula to prepare community and faculty for CBPR partnerships. CES4Health. Info.

Antecol, H., and Bedard, K. (2006). Unhealthy assimilation: why do immigrants converge to American health status levels? Demography *43*, 337–360.

Arcan, C., Larson, N., Bauer, K., Berge, J., Story, M., and Neumark-Sztainer, D. (2014). Dietary and weight-related behaviors and body mass index among Hispanic, Hmong, Somali, and white adolescents. J. Acad. Nutr. Diet. *114*, 375–383.

Arnold, I.C., Dehzad, N., Reuter, S., Martin, H., Becher, B., Taube, C., and Müller, A. (2011). Helicobacter pylori infection prevents allergic asthma in mouse models through the induction of regulatory T cells. J. Clin. Invest. *121*, 3088–3093.

Arnold, R.R., Brewer, M., and Gauthier, J.J. (1980). Bactericidal activity of human lactoferrin: sensitivity of a variety of microorganisms. Infect. Immun. *28*, 893–898.

Asahara, T., Nomoto, K., Shimizu, K., Watanuki, M., and Tanaka, R. (2001). Increased resistance of mice to Salmonella enterica serovar Typhimurium infection by synbiotic administration of Bifidobacteria and transgalactosylated oligosaccharides. J. Appl. Microbiol. *91*, 985–996.

Asnicar, F., Weingart, G., Tickle, T.L., Huttenhower, C., and Segata, N. (2015). Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ *3*, e1029.

Asuncion, A., and Newman, D. (2007). UCI machine learning repository.

Atarashi, K., Nishimura, J., Shima, T., Umesaki, Y., Yamamoto, M., Onoue, M., Yagita, H., Ishii, N., Evans, R., Honda, K., et al. (2008). ATP drives lamina propria TH17 cell differentiation. Nature *455*, 808–812.

Atarashi, K., Tanoue, T., Oshima, K., Suda, W., Nagano, Y., Nishikawa, H., Fukuda, S., Saito, T., Narushima, S., Hase, K., et al. (2013). Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. Nature *500*, 232–236.

Aujla, S.J., Dubin, P.J., and Kolls, J.K. (2007). Th17 cells and mucosal host defense. Semin. Immunol. *19*, 377–382.

Avery, G.B., Randolph, J.G., and Weaver, T. (1966). Gastric acidity in the first day of life. Pediatrics *37*, 1005–1007.

Ayres, J.S., Trinidad, N.J., and Vance, R.E. (2012). Lethal inflammasome activation by a multidrug-resistant pathobiont upon antibiotic disruption of the microbiota. Nat. Med. *18*, 799–806.

Bach, J.-F. (2002). The effect of infections on susceptibility to autoimmune and allergic diseases. N. Engl. J. Med. *347*, 911–920.

Bäckhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A., and Gordon, J.I. (2005). Host-bacterial mutualism in the human intestine. Science *307*, 1915–1920.

Bailey, L.C., Forrest, C.B., Zhang, P., Richards, T.M., Livshits, A., and DeRusso, P.A. (2014). Association of antibiotics in infancy with early childhood obesity. JAMA Pediatr. *168*, 1063–1069.

Banjong, O., Menefee, A., Sranacharoenpong, K., Chittchang, U., Eg-kantrong, P., Boonpraderm, A., and Tamachotipong, S. (2003). Dietary assessment of refugees living in camps: a case study of Mae La Camp, Thailand. Food Nutr. Bull. *24*, 360–367.

Barcenas, C.H., Wilkinson, A.V., Strom, S.S., Cao, Y., Saunders, K.C., Mahabir, S., Hernández-Valero, M.A., Forman, M.R., Spitz, M.R., and Bondy, M.L. (2007). Birthplace, years of residence in the United States, and obesity among Mexican-American adults. Obesity *15*, 1043–1052.

Bates, L.M., Acevedo-Garcia, D., Alegría, M., and Krieger, N. (2008). Immigration and generational trends in body mass index and obesity in the United States: results of the National Latino and Asian American Survey, 2002-2003. Am. J. Public Health *98*, 70–77.

Bezirtzoglou, E. (1997). The intestinal microflora during the first weeks of life. Anaerobe *3*, 173–177.

Biasucci, G., Rubini, M., Riboni, S., Morelli, L., Bessi, E., and Retetangos, C. (2010). Mode of delivery affects the bacterial community in the newborn gut. Early Hum. Dev. *86 Suppl 1*, 13–15.

Biedermann, L., and Rogler, G. (2015). The intestinal microbiota: its role in health and disease. Eur. J. Pediatr. *174*, 151–167.

Bisgaard, H., Li, N., Bonnelykke, K., Chawes, B.L.K., Skov, T., Paludan-Müller, G., Stokholm, J., Smith, B., and Krogfelt, K.A. (2011). Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age. J. Allergy Clin. Immunol. *128*, 646–652.e1–e5.

Björkstén, B., Sepp, E., Julge, K., Voor, T., and Mikelsaar, M. (2001a). Allergy development and the intestinal microflora during the first year of life. J. Allergy Clin. Immunol. *108*, 516–520.

Blaser, M.J., and Falkow, S. (2009). What are the consequences of the disappearing human microbiota? Nat. Rev. Microbiol. *7*, 887–894.

Blin, K., Wolf, T., Chevrette, M.G., Lu, X., Schwalen, C.J., Kautsar, S.A., Suarez Duran, H.G., de los Santos, E.L.C., Kim, H.U., Nave, M., et al. (2017). antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Res. *45*, W36–W41.

Bode, L. (2012). Human milk oligosaccharides: every baby needs a sugar mama. Glycobiology *22*, 1147–1162.

Bokulich, N.A., Rideout, J.R., Mercurio, W.G., Shiffer, A., Wolfe, B., Maurice, C.F., Dutton, R.J., Turnbaugh, P.J., Knight, R., and Caporaso, J.G. (2016a). mockrobiota: a Public Resource for Microbiome Bioinformatics Benchmarking. mSystems *1*.

Bokulich, N.A., Chung, J., Battaglia, T., Henderson, N., Jay, M., Li, H., D Lieber, A., Wu, F., Perez-Perez, G.I., Chen, Y., et al. (2016b). Antibiotics, birth mode, and diet shape microbiome maturation during early life. Sci. Transl. Med. *8*, 343ra82.

Brandl, K., Plitas, G., Mihu, C.N., Ubeda, C., Jia, T., Fleisher, M., Schnabl, B., DeMatteo, R.P., and Pamer, E.G. (2008). Vancomycin-resistant enterococci exploit antibiotic-induced innate immune deficits. Nature *455*, 804–807.

Breiman, L. (2001). Random Forests. Mach. Learn. *45*, 5–32.

Britten, P. (2013). SuperTracker Incorporates Food Composition Data into Innovative Online Consumer Tool. Procedia Food Science *2*, 172–179.

Brown, C.T., Davis-Richardson, A.G., Giongo, A., Gano, K.A., Crabb, D.B., Mukherjee, N., Casella, G., Drew, J.C., Ilonen, J., Knip, M., et al. (2011). Gut microbiome metagenomics analysis suggests a functional model for the development of autoimmunity for type 1 diabetes. PLoS One *6*, e25792.

Brown, E.M., Wlodarska, M., Willing, B.P., Vonaesch, P., Han, J., Reynolds, L.A., Arrieta, M.-C., Uhrig, M., Scholz, R., Partida, O., et al. (2015). Diet and specific microbial exposure trigger features of environmental enteropathy in a novel murine model. Nat. Commun. *6*, 7806.

Brugman, S., Klatter, F.A., Visser, J.T.J., Wildeboer-Veloo, A.C.M., Harmsen, H.J.M., Rozing, J., and Bos, N.A. (2006). Antibiotic treatment partially protects against type 1 diabetes in the Bio-Breeding diabetes-prone rat. Is the gut flora involved in the development of type 1 diabetes? Diabetologia *49*, 2105–2108.

Buffie, C.G., Jarchum, I., Equinda, M., Lipuma, L., Gobourne, A., Viale, A., Ubeda, C., Xavier, J., and Pamer, E.G. (2012). Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to Clostridium difficile-induced colitis. Infect. Immun. *80*, 62–73.

Bureau of Population, Refugees and Migration (2004). Long Wait is Over: Hmong from Wat Tham Krabok Begin Arriving in U.S. U.S. Refugee Admissions Program News *2*.

Cabrera-Rubio, R., Collado, M.C., Laitinen, K., Salminen, S., Isolauri, E., and Mira, A. (2012). The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. Am. J. Clin. Nutr. *96*, 544–551.

Cairney, J., and Ostbye, T. (1999). Time since immigration and excess body weight. Can. J. Public Health *90*, 120–124.

Cani, P.D., Bibiloni, R., Knauf, C., Waget, A., Neyrinck, A.M., Delzenne, N.M., and Burcelin, R. (2008). Changes in Gut Microbiota Control Metabolic Endotoxemia-Induced Inflammation in High-Fat Diet–Induced Obesity and Diabetes in Mice. Diabetes *57*, 1470–1481.

Cani, P.D., Possemiers, S., Van de Wiele, T., Guiot, Y., Everard, A., Rottier, O., Geurts, L., Naslain, D., Neyrinck, A., Lambert, D.M., et al. (2009a). Changes in gut microbiota control inflammation in obese mice through a mechanism involving GLP-2-driven improvement of gut permeability. Gut *58*, 1091–1103.

Cani, P.D., Lecourt, E., Dewulf, E.M., Sohet, F.M., Pachikian, B.D., Naslain, D., De Backer, F., Neyrinck, A.M., and Delzenne, N.M. (2009b). Gut microbiota fermentation of prebiotics increases satietogenic and incretin gut peptide production with consequences for appetite sensation and glucose response after a meal. Am. J. Clin. Nutr. *90*, 1236–1243.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods *7*, 335–336.

Caporaso, J.G., Lauber, C.L., Costello, E.K., Berg-Lyons, D., Gonzalez, A., Stombaugh, J., Knights, D., Gajer, P., Ravel, J., Fierer, N., et al. (2011). Moving pictures of the human microbiome. Genome Biol. *12*, R50.

Careyva, B., LaNoue, M., Bangura, M., de la Paz, A., Gee, A., Patel, N., and Mills, G. (2015). The effect of living in the United States on body mass index in refugee patients. J. Health Care Poor Underserved *26*, 421–430.

Caspi, R., Foerster, H., Fulcher, C.A., Kaipa, P., Krummenacker, M., Latendresse, M., Paley, S., Rhee, S.Y., Shearer, A.G., Tissier, C., et al. (2008). The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res. *36*, D623–D631.

Center For Disease Control (2014). National Health and Nutrition Examination Survey (NHANES) anthropometry procedures manual; 2009.

Cerqueira, M.T., Fry, M.M., and Connor, W.E. (1979). The food and nutrient intakes of the Tarahumara Indians of Mexico. Am. J. Clin. Nutr. *32*, 905–915.

Chai, G., Governale, L., McMahon, A.W., Trinidad, J.P., Staffa, J., and Murphy, D. (2012). Trends of Outpatient Prescription Drug Utilization in US Children, 2002–2010. Pediatrics *130*, 23–31.

Chang, J.Y., Antonopoulos, D.A., Kalra, A., Tonelli, A., Khalife, W.T., Schmidt, T.M., and Young, V.B. (2008). Decreased Diversity of the Fecal Microbiome in Recurrent Clostridium difficile—Associated Diarrhea. J. Infect. Dis. *197*, 435–438.

Chassaing, B., Koren, O., Goodrich, J.K., Poole, A.C., Srinivasan, S., Ley, R.E., and Gewirtz, A.T. (2015). Dietary emulsifiers impact the mouse gut microbiota promoting colitis and metabolic syndrome. Nature *519*, 92–96.

Cherrier, M., and Eberl, G. (2012). The development of LTi cells. Curr. Opin. Immunol. *24*, 178–183.

Cho, I., Yamanishi, S., Cox, L., Methé, B.A., Zavadil, J., Li, K., Gao, Z., Mahana, D., Raju, K., Teitler, I., et al. (2012a). Antibiotics in early life alter the murine colonic microbiome and adiposity. Nature *488*, 621–626.

Clemente, J.C., Pehrsson, E.C., Blaser, M.J., Sandhu, K., Gao, Z., Wang, B., Magris, M., Hidalgo, G., Contreras, M., Noya-Alarcón, Ó., et al. (2015). The microbiome of uncontacted Amerindians. Sci Adv *1*.

Cortes, C., and Vapnik, V. (1995). Support-vector networks. Mach. Learn. *20*, 273–297.

Cotten, C.M., Taylor, S., Stoll, B., Goldberg, R.N., Hansen, N.I., Sánchez, P.J., Ambalavanan, N., Benjamin, D.K., Jr, and NICHD Neonatal Research Network (2009).

Prolonged duration of initial empirical antibiotic treatment is associated with increased rates of necrotizing enterocolitis and death for extremely low birth weight infants. Pediatrics *123*, 58–66.

Cox, L.M., Yamanishi, S., Sohn, J., Alekseyenko, A.V., Leung, J.M., Cho, I., Kim, S.G., Li, H., Gao, Z., Mahana, D., et al. (2014). Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. Cell *158*, 705–721.

Croswell, A., Amir, E., Teggatz, P., Barman, M., and Salzman, N.H. (2009). Prolonged impact of antibiotics on intestinal microbial ecology and susceptibility to enteric Salmonella infection. Infect. Immun. *77*, 2741–2753.

Cummins, A.G., and Thompson, F.M. (2002). Effect of breast milk and weaning on epithelial growth of the small intestine in humans. Gut *51*, 748–754.

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

Dawson-Hahn, E., Pak-Gorstein, S., Matheson, J., Zhou, C., Yun, K., Scott, K., Payton, C., Stein, E., Holland, A., Grow, H.M., et al. (2016). Growth Trajectories of Refugee and Nonrefugee Children in the United States. Pediatrics *138*.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., and Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc. Natl. Acad. Sci. U. S. A. *107*, 14691–14696.

Denning, T.L., Granger, S., Mucida, D., Graddy, R., Leclercq, G., Zhang, W., Honey, K., Rasmussen, J.P., Cheroutre, H., Rudensky, A.Y., et al. (2007). Mouse TCRαβ+CD8αα intraepithelial lymphocytes express genes that down-regulate their antigen reactivity and suppress immune responses. The Journal of Immunology *178*, 6654–6654.

Department of Economic and Social Affairs, Population Division (2017). International Migration Report 2017 (United Nations).

Deshmukh, H.S., Liu, Y., Menkiti, O.R., Mei, J., Dai, N., O'Leary, C.E., Oliver, P.M., Kolls, J.K., Weiser, J.N., and Worthen, G.S. (2014). The microbiota regulates neutrophil homeostasis and host resistance to Escherichia coli K1 sepsis in neonatal mice. Nat. Med. *20*, 524–530.

Dethlefsen, L., and Relman, D.A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc. Natl. Acad. Sci. U. S. A. *108 Suppl 1*, 4554–4561.

Dethlefsen, L., Huse, S., Sogin, M.L., and Relman, D.A. (2008). The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. PLoS Biol. *6*, e280.

De Vadder, F., Kovatcheva-Datchary, P., Goncalves, D., Vinera, J., Zitoun, C., Duchampt, A., Bäckhed, F., and Mithieux, G. (2014). Microbiota-generated metabolites promote metabolic benefits via gut-brain neural circuits. Cell *156*, 84–96.

Dominguez-Bello, M.G., Costello, E.K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., and Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc. Natl. Acad. Sci. U. S. A. *107*, 11971–11975.

Dong, P., Yang, Y., and Wang, W.-P. (2010). The role of intestinal bifidobacteria on immune system development in young rats. Early Hum. Dev. *86*, 51–58.

Donskey, C.J., Chowdhry, T.K., Hecker, M.T., Hoyen, C.K., Hanrahan, J.A., Hujer, A.M., Hutton-Thomas, R.A., Whalen, C.C., Bonomo, R.A., and Rice, L.B. (2000). Effect of antibiotic therapy on the density of vancomycin-resistant enterococci in the stool of colonized patients. N. Engl. J. Med. *343*, 1925–1932.

Droste, J.H., Wieringa, M.H., Weyler, J.J., Nelen, V.J., Vermeire, P.A., and Van Bever, H.P. (2000). Does the use of antibiotics in early childhood increase the risk of asthma and allergic disease? Clin. Exp. Allergy *30*, 1547–1553.

El Kaoutari, A., Armougom, F., Gordon, J.I., Raoult, D., and Henrissat, B. (2013). The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nat. Rev. Microbiol. *11*, 497–504.

Engel, J., Schmalhorst, P.S., and Routier, F.H. (2012). Biosynthesis of the fungal cell wall polysaccharide galactomannan requires intraluminal GDP-mannose. J. Biol. Chem. *287*, 44418–44424.

Faith, D.P. (1992). Conservation evaluation and phylogenetic diversity. Biol. Conserv. *61*, 1–10.

Faith, J.J., Guruge, J.L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A.L., Clemente, J.C., Knight, R., Heath, A.C., Leibel, R.L., et al. (2013). The long-term stability of the human gut microbiota. Science *341*, 1237439.

Falk, P.G., Hooper, L.V., Midtvedt, T., and Gordon, J.I. (1998). Creating and maintaining the gastrointestinal ecosystem: what we know and need to know from gnotobiology. Microbiol. Mol. Biol. Rev. *62*, 1157–1170.

Farooqi, I.S., and Hopkin, J.M. (1998). Early childhood infection and atopic disorder. Thorax *53*, 927–932.

Favier, C.F., de Vos, W.M., and Akkermans, A.D.L. (2003). Development of bacterial and bifidobacterial communities in feces of newborn babies. Anaerobe *9*, 219–229.

Febinia, C.A. (2017). The Gut Microbiota of Bali among the World Populations: Connecting Diet, Urbanisation, and Obesity. University of Sydney.

Fierro, J.L., Prasad, P.A., Localio, A.R., Grundmeier, R.W., Wasserman, R.C., Zaoutis, T.E., and Gerber, J.S. (2014). Variability in the diagnosis and treatment of group a streptococcal pharyngitis by primary care pediatricians. Infect. Control Hosp. Epidemiol. *35 Suppl 3*, S79–S85.

Finkelstein, E.A., Trogdon, J.G., Cohen, J.W., and Dietz, W. (2009). Annual medical spending attributable to obesity: payer-and service-specific estimates. Health Aff. *28*, w822–w831.

Flint, H.J., Scott, K.P., Duncan, S.H., Louis, P., and Forano, E. (2012). Microbial degradation of complex carbohydrates in the gut. Gut Microbes *3*, 289–306.

Forchielli, M.L., and Walker, W.A. (2005). The role of gut-associated lymphoid tissues and mucosal defence. Br. J. Nutr. *93 Suppl 1*, S41–S48.

Forster, S.C., Browne, H.P., Kumar, N., Hunt, M., Denise, H., Mitchell, A., Finn, R.D., and Lawley, T.D. (2016). HPMCD: the database of human microbial communities from metagenomic datasets and microbial reference genomes. Nucleic Acids Res. *44*, D604–D609.

Fouhy, F., Guinane, C.M., Hussey, S., Wall, R., Ryan, C.A., Dempsey, E.M., Murphy, B., Ross, R.P., Fitzgerald, G.F., Stanton, C., et al. (2012). High-throughput sequencing reveals the incomplete, short-term recovery of infant gut microbiota following parenteral antibiotic treatment with ampicillin and gentamicin. Antimicrob. Agents Chemother. *56*, 5811–5820.

Franzen, L., and Smith, C. (2009). Acculturation and environmental change impacts dietary habits among adult Hmong. Appetite *52*, 173–183.

Fu, B.C., Randolph, T.W., Lim, U., Monroe, K.R., Cheng, I., Wilkens, L.R., Le Marchand, L., Hullar, M.A.J., and Lampe, J.W. (2016). Characterization of the gut microbiome in epidemiologic studies: the multiethnic cohort experience. Ann. Epidemiol. *26*, 373–379.

Funkhouser, L.J., and Bordenstein, S.R. (2013). Mom knows best: the universality of maternal microbial transmission. PLoS Biol. *11*, e1001631.

Furey, T.S., Cristianini, N., Duffy, N., Bednarski, D.W., Schummer, M., and Haussler, D. (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics *16*, 906–914.

Gerber, J.S., Prasad, P.A., and Localio, A.R. (2014). Variation in antibiotic prescribing across a pediatric primary care network. Journal of the.

Gevers, D., Kugathasan, S., Denson, L.A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S.J., Yassour, M., et al. (2014). The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe *15*, 382–392.

Giongo, A., Gano, K.A., Crabb, D.B., Mukherjee, N., Novelo, L.L., Casella, G., Drew, J.C., Ilonen, J., Knip, M., Hyöty, H., et al. (2011). Toward defining the autoimmune microbiome for type 1 diabetes. ISME J. *5*, 82–91.

Goel, M.S., McCarthy, E.P., Phillips, R.S., and Wee, C.C. (2004). Obesity among US immigrant subgroups by duration of residence. JAMA *292*, 2860–2867.

Gohl, D.M., Vangay, P., Garbe, J., MacLean, A., Hauge, A., Becker, A., Gould, T.J., Clayton, J.B., Johnson, T.J., Hunter, R., et al. (2016). Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. Nat. Biotechnol. *34*, 942–949.

Gomez, A., Petrzelkova, K.J., Burns, M.B., Yeoman, C.J., Amato, K.R., Vlckova, K., Modry, D., Todd, A., Jost Robinson, C.A., Remis, M.J., et al. (2016). Gut Microbiome of Coexisting BaAka Pygmies and Bantu Reflects Gradients of Traditional Subsistence Patterns. Cell Rep. *14*, 2142–2153.

Gonzales, R., Malone, D.C., Maselli, J.H., and Sande, M.A. (2001). Excessive antibiotic use for acute respiratory infections in the United States. Clin. Infect. Dis. *33*, 757–762.

Goossens, H., Ferech, M., Vander Stichele, R., Elseviers, M., and ESAC Project Group (2005). Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. Lancet *365*, 579–587.

Gruber, J., Swartz, A., and Others (2004). Markdown.

Haas, H., Falcone, F.H., Holland, M.J., Schramm, G., Haisch, K., Gibbs, B.F., Bufe, A., and Schlaak, M. (1999). Early interleukin-4: its role in the switch towards a Th2 response and IgE-mediated allergy. Int. Arch. Allergy Immunol. *119*, 86–94.

Hansen, C.H.F., Nielsen, D.S., Kverka, M., Zakostelska, Z., Klimesova, K., Hudcovic, T., Tlaskalova-Hogenova, H., and Hansen, A.K. (2012). Patterns of early gut colonization shape future immune responses of the host. PLoS One *7*, e34043.

Heney, J.H., Dimock, C.C., Friedman, J.F., and Lewis, C. (2014). Pediatric refugees in Rhode Island: increases in BMI percentile, overweight, and obesity following resettlement. R. I. Med. J. *98*, 43–47.

Hersh, A.L., Jackson, M.A., Hicks, L.A., and American Academy of Pediatrics Committee on Infectious Diseases (2013). Principles of judicious antibiotic prescribing for upper respiratory tract infections in pediatrics. Pediatrics *132*, 1146–1154.

Hervey, K., Vargas, D., Klesges, L., Fischer, P.R., Trippel, S., and Juhn, Y.J. (2009). Overweight among refugee children after arrival in the United States. J. Health Care Poor Underserved *20*, 246–256.

Hess, J., Ladel, C., Miko, D., and Kaufmann, S.H. (1996). Salmonella typhimurium aroA-infection in gene-targeted immunodeficient mice: major role of CD4+ TCR-alpha beta cells and IFN-gamma in bacterial clearance independent of intracellular location. The Journal of Immunology *156*, 3321–3326.

Hibberd, C.M., Brooke, O.G., Carter, N.D., Haug, M., and Harzer, G. (1982). Variation in the composition of breast milk during the first 5 weeks of lactation: implications for the feeding of preterm infants. Arch. Dis. Child. *57*, 658–662.

Hicks, L.A., Taylor, T.H., and Hunkler, R.J. (2013). U.S. Outpatient Antibiotic Prescribing, 2010. N. Engl. J. Med. *368*, 1461–1462.

Hildebrandt, M.A., Hoffmann, C., Sherrill-Mix, S.A., Keilbaugh, S.A., Hamady, M., Chen, Y.-Y., Knight, R., Ahima, R.S., Bushman, F., and Wu, G.D. (2009). High-fat diet determines the composition of the murine gut microbiome independently of obesity. Gastroenterology *137*, 1716–1724.e1–e2.

Hill, D.A., Siracusa, M.C., Abt, M.C., Kim, B.S., Kobuley, D., Kubo, M., Kambayashi, T., Larosa, D.F., Renner, E.D., Orange, J.S., et al. (2012). Commensal bacteria-derived signals regulate basophil hematopoiesis and allergic inflammation. Nat. Med. *18*, 538–546.

Himes, J.H., Story, M., Czaplinski, K., and Dahlberg-Luby, E. (1992). Indications of early obesity in low-income Hmong children. Am. J. Dis. Child. *146*, 67–69.

Horton, D.B., Scott, F.I., Haynes, K., Putt, M.E., Rose, C.D., Lewis, J.D., and Strom, B.L. (2014). Antibiotic Exposure and the Development of Juvenile Idiopathic Arthritis: A Population-Based Case-Control Study.

Howarth, N.C., Saltzman, E., and Roberts, S.B. (2001). Dietary fiber and weight regulation. Nutr. Rev. *59*, 129–139.

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

Human Microbiome Project Consortium (2012b). A framework for human microbiome research. Nature *486*, 215–221.

Hunt, K.M., Foster, J.A., Forney, L.J., Schütte, U.M.E., Beck, D.L., Abdo, Z., Fox, L.K., Williams, J.E., McGuire, M.K., and McGuire, M.A. (2011). Characterization of the diversity and temporal stability of bacterial communities in human milk. PLoS One *6*, e21313.

Hunter, S., Corbett, M., Denise, H., Fraser, M., Gonzalez-Beltran, A., Hunter, C., Jones, P., Leinonen, R., McAnulla, C., Maguire, E., et al. (2014). EBI metagenomics—a new resource for the analysis and archiving of metagenomic data. Nucleic Acids Res. *42*, D600–D606.

Hviid, A., Svanström, H., and Frisch, M. (2011). Antibiotic use and inflammatory bowel diseases in childhood. Gut *60*, 49–54.

Ivanov, I.I., Atarashi, K., Manel, N., Brodie, E.L., Shima, T., Karaoz, U., Wei, D., Goldfarb, K.C., Santee, C.A., Lynch, S.V., et al. (2009). Induction of intestinal Th17 cells by segmented filamentous bacteria. Cell *139*, 485–498.

Johnson, C.C., Ownby, D.R., Alford, S.H., Havstad, S.L., Williams, L.K., Zoratti, E.M., Peterson, E.L., and Joseph, C.L.M. (2005). Antibiotic exposure in early infancy and risk for childhood atopy. J. Allergy Clin. Immunol. *115*, 1218–1224.

Jordan, M.I., and Mitchell, T.M. (2015). Machine learning: Trends, perspectives, and prospects. Science *349*, 255–260.

Jujo, K., Renz, H., Abe, J., Gelfand, E.W., and Leung, D.Y. (1992). Decreased interferon gamma and increased interleukin-4 production in atopic dermatitis promotes IgE synthesis. J. Allergy Clin. Immunol. *90*, 323–331.

Kalliomäki, M., Kirjavainen, P., Eerola, E., Kero, P., Salminen, S., and Isolauri, E. (2001). Distinct patterns of neonatal gut microflora in infants in whom atopy was and was not developing. J. Allergy Clin. Immunol. *107*, 129–134.

Kao, C.-Y., Chen, Y., Thai, P., Wachi, S., Huang, F., Kim, C., Harper, R.W., and Wu, R. (2004). IL-17 Markedly Up-Regulates β-Defensin-2 Expression in Human Airway Epithelium via JAK and NF-κB Signaling Pathways. The Journal of Immunology *173*, 3482–3491.

Kaplan, M.S., Huguet, N., Newsom, J.T., and McFarland, B.H. (2004). The association between length of residence and obesity among Hispanic immigrants. Am. J. Prev. Med. *27*, 323–326.

Karlsson, F.H., Tremaroli, V., Nookaew, I., Bergström, G., Behre, C.J., Fagerberg, B., Nielsen, J., and Bäckhed, F. (2013). Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature *498*, 99–103.

Kelly, E.J., Newell, S.J., Brownlee, K.G., Primrose, J.N., and Dear, P.R. (1993). Gastric acid secretion in preterm infants. Early Hum. Dev. *35*, 215–220.

Kennedy, S., McDonald, J.T., Biddle, N., and Others (2006). The healthy immigrant effect and immigrant selection: evidence from four countries.

Knights, D., Parfrey, L.W., Zaneveld, J., Lozupone, C., and Knight, R. (2011a). Human-associated microbial signatures: examining their predictive value. Cell Host Microbe *10*, 292–296.

Knights, D., Costello, E.K., and Knight, R. (2011b). Supervised classification of human microbiota. FEMS Microbiol. Rev. *35*, 343–359.

Kodaman, N., Pazos, A., Schneider, B.G., Piazuelo, M.B., Mera, R., Sobota, R.S., Sicinschi, L.A., Shaffer, C.L., Romero-Gallo, J., de Sablet, T., et al. (2014). Human and Helicobacter pylori coevolution shapes the risk of gastric disease. Proc. Natl. Acad. Sci. U. S. A. *111*, 1455–1460.

Koenig, J.E., Spor, A., Scalfone, N., Fricker, A.D., Stombaugh, J., Knight, R., Angenent, L.T., and Ley, R.E. (2011). Succession of microbial consortia in the developing infant gut microbiome. Proc. Natl. Acad. Sci. U. S. A. *108 Suppl 1*, 4578–4585.

Kostic, A.D., Gevers, D., Pedamallu, C.S., Michaud, M., Duke, F., Earl, A.M., Ojesina, A.I., Jung, J., Bass, A.J., Tabernero, J., et al. (2012). Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Res. *22*, 292–298.

Koumpilova, M. (2015). Minnesota prepares to receive more refugees in 2016. Star Tribune.

Kronman, M.P., Zhou, C., and Mangione-Smith, R. (2014). Bacterial prevalence and antimicrobial prescribing trends for acute respiratory tract infections. Pediatrics *134*, e956–e965.

Kuhn, M., and Others (2008). Caret package. J. Stat. Softw. *28*, 1–26.

Kunstadter, P. (2001). Health implications of globalization at the village level: the good, the bad, and the ugly: some results of comparative research in Thailand and the US. Presentation at Woodrow Wilson School of International Studies. Princeton, New Jersey: Princeton University.

Kwok, L.-Y., Zhang, J., Guo, Z., Gesudu, Q., Zheng, Y., Qiao, J., Huo, D., and Zhang, H. (2014). Characterization of fecal microbiota across seven Chinese ethnic groups by quantitative polymerase chain reaction. PLoS One *9*, e93631.

Lanzavecchia, A., and Sallusto, F. (2001). Regulation of T cell immunity by dendritic cells. Cell *106*, 263–266.

Lauderdale, D.S., and Rathouz, P.J. (2000). Body mass index in a US national sample of Asian Americans: effects of nativity, years since immigration and socioeconomic status. Int. J. Obes. Relat. Metab. Disord. *24*, 1188–1194.

Lawley, T.D., Bouley, D.M., Hoy, Y.E., Gerke, C., Relman, D.A., and Monack, D.M. (2008). Host transmission of Salmonella enterica serovar Typhimurium is controlled by virulence factors and indigenous intestinal microbiota. Infect. Immun. *76*, 403–416.

Lebenthal, A., and Lebenthal, E. (1999). The ontogeny of the small intestinal epithelium. JPEN J. Parenter. Enteral Nutr. *23*, S3–S6.

Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J.-M., Kennedy, S., et al. (2013). Richness of human gut microbiome correlates with metabolic markers. Nature *500*, 541–546.

Lee, Y.K., and Mazmanian, S.K. (2010). Has the microbiota played a critical role in the evolution of the adaptive immune system? Science *330*, 1768–1773.

Leinonen, R., Sugawara, H., Shumway, M., and International Nucleotide Sequence Database Collaboration (2011). The sequence read archive. Nucleic Acids Res. *39*, D19–D21.

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

Ley, R.E., Turnbaugh, P.J., Klein, S., and Gordon, J.I. (2006). Microbial ecology: human gut microbes associated with obesity. Nature *444*, 1022–1023.

Lindeberg, S., and Lundh, B. (1993). Apparent absence of stroke and ischaemic heart disease in a traditional Melanesian island: a clinical study in Kitava. J. Intern. Med. *233*, 269–275.

Liu, S., Willett, W.C., Manson, J.E., Hu, F.B., Rosner, B., and Colditz, G. (2003). Relation between changes in intakes of dietary fiber and grain products and changes in weight and development of obesity among middle-aged women. Am. J. Clin. Nutr. *78*, 920–927.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. *42*, D490–D495.

Louis, N.A., and Lin, P.W. (2009). The Intestinal Immune Barrier. Neoreviews *10*, e180–e190.

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

Lundell, A.-C., Johansen, S., Adlerberth, I., Wold, A.E., Hesselmar, B., and Rudin, A. (2014). High proportion of CD5+ B cells in infants predicts development of allergic disease. J. Immunol. *193*, 510–518.

Maachi, M., Piéroni, L., Bruckert, E., Jardel, C., Fellahi, S., Hainque, B., Capeau, J., and Bastard, J.-P. (2004). Systemic low-grade inflammation is related to both circulating and adipose tissue TNFα, leptin and IL-6 levels in obese women. Int. J. Obes. *28*, 993–997.

Macatonia, S.E., Hosken, N.A., Litton, M., Vieira, P., Hsieh, C.S., Culpepper, J.A., Wysocka, M., Trinchieri, G., Murphy, K.M., and O'Garra, A. (1995). Dendritic cells produce IL-12 and direct the development of Th1 cells from naive CD4+ T cells. J. Immunol. *154*, 5071–5079.

Macpherson, A.J., and Harris, N.L. (2004). Interactions between commensal intestinal bacteria and the immune system. Nat. Rev. Immunol. *4*, 478–485.

Maheshwari, A., and Zemlin, M. (2009). Ontogeny of the intestinal immune system. Hematology Meeting Reports (formerly Haematologica Reports) *2*.

Mai, V., Young, C.M., Ukhanova, M., Wang, X., Sun, Y., Casella, G., Theriaque, D., Li, N., Sharma, R., Hudak, M., et al. (2011a). Fecal microbiota in premature infants prior to necrotizing enterocolitis. PLoS One *6*, e20647.

Mårild, K., Ye, W., Lebwohl, B., Green, P.H.R., Blaser, M.J., Card, T., and Ludvigsson, J.F. (2013). Antibiotic exposure and the development of coeliac disease: a nationwide case--control study. BMC Gastroenterol. *13*, 109.

Matamoros, S., Gras-Leguen, C., Le Vacon, F., Potel, G., and de La Cochetiere, M.-F. (2013). Development of intestinal microbiota in infants and its impact on health. Trends Microbiol. *21*, 167–173.

Maynard, C.L., Elson, C.O., Hatton, R.D., and Weaver, C.T. (2012). Reciprocal interactions of the intestinal microbiota and immune system. Nature *489*, 231–241.

Mazmanian, S.K., Round, J.L., and Kasper, D.L. (2008). A microbial symbiosis factor prevents intestinal inflammatory disease. Nature *453*, 620–625.

McCaig, L.F., Besser, R.E., and Hughes, J.M. (2003). Antimicrobial drug prescription in ambulatory care settings, United States, 1992-2000. Emerg. Infect. Dis. *9*, 432–437.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., and Hugenholtz, P. (2012). An improved Greengenes

taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. *6*, 610–618.

McDonald, D., Hyde, E., Debelius, J.W., Morton, J.T., Gonzalez, A., Ackermann, G., Aksenov, A.A., Behsaz, B., Brennan, C., Chen, Y., et al. (2018). American Gut: an Open Platform for Citizen Science Microbiome Research. mSystems *3*.

McKeever, T.M., Lewis, S.A., Smith, C., Collins, J., Heatlie, H., Frischer, M., and Hubbard, R. (2002). Early exposure to infections and antibiotics and the incidence of allergic disease: a birth cohort study with the West Midlands General Practice Research Database. J. Allergy Clin. Immunol. *109*, 43–50.

McTigue, K.M., Garrett, J.M., and Popkin, B.M. (2002). The natural history of the development of obesity in a cohort of young U.S. adults between 1981 and 1998. Ann. Intern. Med. *136*, 857–864.

Ménard, D. (2004). Functional development of the human gastrointestinal tract: hormone-and growth factor-mediated regulatory mechanisms. Canadian Journal of Gastroenterology= Journal Canadien de Gastroenterologie *18*, 39–44.

Minnesota Department of Health MDH Primary Refugee Arrival Data.

Moles, L., Manzano, S., Fernández, L., Montilla, A., Corzo, N., Ares, S., Rodríguez, J.M., and Espinosa-Martos, I. (2015). Bacteriological, biochemical, and immunological properties of colostrum and mature milk from mothers of extremely preterm infants. J. Pediatr. Gastroenterol. Nutr. *60*, 120–126.

Montassier, E., Al-Ghalith, G.A., Ward, T., Corvec, S., Gastinne, T., Potel, G., Moreau, P., de la Cochetiere, M.F., Batard, E., and Knights, D. (2016). Pretreatment gut microbiome predicts chemotherapy-related bloodstream infection. Genome Med. *8*, 49.

Montgomery, R.K., Mulberg, A.E., and Grand, R.J. (1999). Development of the human gastrointestinal tract: twenty years of progress. Gastroenterology *116*, 702–731.

Morello, C. (2016). Obama administration to expand number of refugees admitted to U.S. Washington Post.

Morton, E.R., Lynch, J., Froment, A., Lafosse, S., Heyer, E., Przeworski, M., Blekhman, R., and Ségurel, L. (2015). Variation in Rural African Gut Microbiota Is Strongly Correlated with Colonization by Entamoeba and Subsistence. PLoS Genet. *11*, e1005658.

Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., González, A., Fontana, L., Henrissat, B., Knight, R., and Gordon, J.I. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science *332*, 970–974.

Mulasi-Pokhriyal, U., Smith, C., and Franzen-Castle, L. (2012). Investigating dietary acculturation and intake among US-born and Thailand/Laos-born Hmong-American children aged 9–18 years. Public Health Nutr. *15*, 176–185.

Nash, D.R., Harman, J., Wald, E.R., and Kelleher, K.J. (2002). Antibiotic prescribing by primary care physicians for children with upper respiratory tract infections. Arch. Pediatr. Adolesc. Med. *156*, 1114–1119.

Nichols, J.H., Bezkorovainy, A., and Landau, W. (1974). Human colostral whey M-l glycoproteins and their L. bifidus var. Penn. growth promoting activities. Life Sci. *14*, 967–976.

Noverr, M.C., Noggle, R.M., Toews, G.B., and Huffnagle, G.B. (2004). Role of antibiotics and fungal microbiota in driving pulmonary allergic responses. Infect. Immun. *72*, 4996–5003.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. Genome Res. *27*, 824–834.

Nyquist, A.C., Gonzales, R., Steiner, J.F., and Sande, M.A. (1998). Antibiotic prescribing for children with colds, upper respiratory tract infections, and bronchitis. JAMA *279*, 875–877.

Obregon-Tito, A.J., Tito, R.Y., Metcalf, J., Sankaranarayanan, K., Clemente, J.C., Ursell, L.K., Zech Xu, Z., Van Treuren, W., Knight, R., Gaffney, P.M., et al. (2015). Subsistence strategies in traditional societies distinguish gut microbiomes. Nat. Commun. *6*, 6505.

O'Keefe, S.J.D., Li, J.V., Lahti, L., Ou, J., Carbonero, F., Mohammed, K., Posma, J.M., Kinross, J., Wahl, E., Ruder, E., et al. (2015). Fat, fibre and cancer risk in African Americans and rural Africans. Nat. Commun. *6*, 6342.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44*, D733–D745.

Olivares-Villagómez, D., and Van Kaer, L. (2018). Intestinal Intraepithelial Lymphocytes: Sentinels of the Mucosal Barrier. Trends Immunol. *39*, 264–275.

Olszak, T., An, D., Zeissig, S., Vera, M.P., Richter, J., Franke, A., Glickman, J.N., Siebert, R., Baron, R.M., Kasper, D.L., et al. (2012). Microbial exposure during early life has persistent effects on natural killer T cell function. Science *336*, 489–493.

Ong, M.-S., Umetsu, D.T., and Mandl, K.D. (2014). Consequences of antibiotics and infections in infancy: bugs, drugs, and wheezing. Ann. Allergy Asthma Immunol. *112*, 441–445.e1.

Paine, R.T., Tegner, M.J., and Johnson, E.A. (1998). Compounded Perturbations Yield Ecological Surprises. Ecosystems *1*, 535–545.

Palmer, C., Bik, E.M., DiGiulio, D.B., Relman, D.A., and Brown, P.O. (2007). Development of the human infant intestinal microbiota. PLoS Biol. *5*, e177.

Parnell, J.A., and Reimer, R.A. (2012). Prebiotic fibres dose-dependently increase satiety hormones and alter Bacteroidetes and Firmicutes in lean and obese JCR:LA-cp rats. Br. J. Nutr. *107*, 601–613.

Pasolli, E., Truong, D.T., Malik, F., Waldron, L., and Segata, N. (2016). Machine Learning Meta-analysis of Large Metagenomic Datasets: Tools and Biological Insights. PLoS Comput. Biol. *12*, e1004977.

Passel, J.S., and Cohn, D. (2008). US population projections: 2005-2050.

Pawlowsky-Glahn, V., and Buccianti, A. (2011). Compositional Data Analysis: Theory and Applications (John Wiley & Sons).

Penders, J., Thijs, C., Vink, C., Stelma, F.F., Snijders, B., Kummeling, I., van den Brandt, P.A., and Stobberingh, E.E. (2006). Factors influencing the composition of the intestinal microbiota in early infancy. Pediatrics *118*, 511–521.

Penders, J., Gerhold, K., Stobberingh, E.E., Thijs, C., Zimmermann, K., Lau, S., and Hamelmann, E. (2013). Establishment of the intestinal microbiota and its role for atopic dermatitis in early childhood. J. Allergy Clin. Immunol. *132*, 601–607.e8.

Pfeifer, M.E., and Thao, B.K. (2013). State of the Hmong American Community (Hmong National Development).

Polak-Charcon, S., Shoham, J., and Ben-Shaul, Y. (1980). Tight junctions in epithelial cells of human fetal hindgut, normal colon, and colon adenocarcinoma. J. Natl. Cancer Inst. *65*, 53–62.

Poussier, P., Ning, T., Banerjee, D., and Julius, M. (2002). A unique subset of self-specific intraintestinal T cells maintains gut integrity. J. Exp. Med. *195*, 1491–1497.

Prescott, S.L., Macaubas, C., Smallacombe, T., Holt, B.J., Sly, P.D., and Holt, P.G. (1999). Development of allergen-specific T-cell memory in atopic and normal children. Lancet *353*, 196–200.

Qiita Development Team QIITA.

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature *490*, 55–60.

Rashidi, A., Ebadi, M., Shields-Cutler, R.R., DeFor, T.E., Al-Ghalith, G.A., Ferrieri, P., Young, J.-A.H., Dunny, G.M., Knights, D., and Weisdorf, D.J. (2018). Pretransplant Gut Colonization with Intrinsically Vancomycin-Resistant Enterococci (E. gallinarum and E. casseliflavus) and Outcomes of Allogeneic Hematopoietic Cell Transplantation. Biol. Blood Marrow Transplant. *24*, 1260–1263.

Rautava, S., Ruuskanen, O., Ouwehand, A., Salminen, S., and Isolauri, E. (2004). The Hygiene Hypothesis of Atopic Disease—An Extended Version. J. Pediatr. Gastroenterol. Nutr. *38*, 378.

Ridaura, V.K., Faith, J.J., Rey, F.E., Cheng, J., Duncan, A.E., Kau, A.L., Griffin, N.W., Lombard, V., Henrissat, B., Bain, J.R., et al. (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science *341*, 1241214.

Robinson, C.J., and Young, V.B. (2010). Antibiotic administration alters the community structure of the gastrointestinal micobiota. Gut Microbes *1*, 279–284.

Rogier, E.W., Frantz, A.L., Bruno, M.E.C., Wedlund, L., Cohen, D.A., Stromberg, A.J., and Kaetzel, C.S. (2014). Secretory antibodies in breast milk promote long-term intestinal homeostasis by regulating the gut microbiota and host gene expression. Proc. Natl. Acad. Sci. U. S. A. *111*, 3074–3079.

Ronacher, A. (2017). Jinja2.

van Rossum, G., and Drake, F.L. (2011). The Python Language Reference Manual (Network Theory Ltd.).

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P.I., Godneva, A., Kalka, I.N., Bar, N., et al. (2018). Environment dominates over host genetics in shaping human gut microbiota. Nature *555*, 210–215.

Round, J.L., and Mazmanian, S.K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. Nat. Rev. Immunol. *9*, 313–323.

Rumbo, M., and Schiffrin, E.J. (2005). Ontogeny of intestinal epithelium immune functions: developmental and environmental regulation. Cell. Mol. Life Sci. *62*, 1288–1296.

Russell, S.L., Gold, M.J., Hartmann, M., Willing, B.P., Thorson, L., Wlodarska, M., Gill, N., Blanchet, M.-R., Mohn, W.W., McNagny, K.M., et al. (2012). Early life antibiotic-driven changes in microbiota enhance susceptibility to allergic asthma. EMBO Rep. *13*, 440–447.

Rutz, S., Eidenschenk, C., and Ouyang, W. (2013). IL-22, not simply a Th17 cytokine. Immunol. Rev. *252*, 116–132.

Sabat, R., Ouyang, W., and Wolk, K. (2014). Therapeutic opportunities of the IL-22-IL-22R1 system. Nat. Rev. Drug Discov. *13*, 21–38.

Sanos, S.L., Bui, V.L., Mortha, A., Oberle, K., Heners, C., Johner, C., and Diefenbach, A. (2009). RORgammat and commensal microflora are required for the differentiation of mucosal interleukin 22-producing NKp46+ cells. Nat. Immunol. *10*, 83–91.

Santacruz, A., Marcos, A., Wärnberg, J., Martí, A., Martin-Matillas, M., Campoy, C., Moreno, L.A., Veiga, O., Redondo-Figuero, C., Garagorri, J.M., et al. (2009). Interplay between weight loss and gut microbiota composition in overweight adolescents. Obesity *17*, 1906–1915.

Satoh-Takayama, N., Vosshenrich, C.A.J., Lesjean-Pottier, S., Sawa, S., Lochner, M., Rattis, F., Mention, J.-J., Thiam, K., Cerf-Bensussan, N., Mandelboim, O., et al. (2008). Microbial flora drives interleukin 22 production in intestinal NKp46+ cells that provide innate mucosal immune defense. Immunity *29*, 958–970.

Sawa, S., Cherrier, M., Lochner, M., Satoh-Takayama, N., Fehling, H.J., Langa, F., Di Santo, J.P., and Eberl, G. (2010). Lineage Relationship Analysis of RORγt+ Innate Lymphoid Cells. Science *330*, 665–669.

Schafer, M.J., White, T.A., Evans, G., Tonne, J.M., Verzosa, G.C., Stout, M.B., Mazula, D.L., Palmer, A.K., Baker, D.J., Jensen, M.D., et al. (2016). Exercise Prevents Diet-Induced Cellular Senescence in Adipose Tissue. Diabetes *65*, 1606–1615.

Schnorr, S.L., Candela, M., Rampelli, S., Centanni, M., Consolandi, C., Basaglia, G., Turroni, S., Biagi, E., Peano, C., Severgnini, M., et al. (2014). Gut microbiome of the Hadza hunter-gatherers. Nat. Commun. *5*, 3654.

Schwiertz, A., Taras, D., Schäfer, K., Beijer, S., Bos, N.A., Donus, C., and Hardt, P.D. (2010). Microbiota and SCFA in lean and overweight healthy subjects. Obesity *18*, 190–195.

Seelig, M.S. (1966). The role of antibiotics in the pathogenesis of Candida infections. Am. J. Med. *40*, 887–917.

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

Sekirov, I., Tam, N.M., Jogova, M., Robertson, M.L., Li, Y., Lupp, C., and Finlay, B.B. (2008). Antibiotic-induced perturbations of the intestinal microbiota alter host susceptibility to enteric infection. Infect. Immun. *76*, 4726–4736.

Sela, D.A., Chapman, J., Adeuya, A., Kim, J.H., Chen, F., Whitehead, T.R., Lapidus, A., Rokhsar, D.S., Lebrilla, C.B., German, J.B., et al. (2008). The genome sequence of

Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome. Proc. Natl. Acad. Sci. U. S. A. *105*, 18964–18969.

Sellitto, M., Bai, G., Serena, G., Fricke, W.F., Sturgeon, C., Gajer, P., White, J.R., Koenig, S.S.K., Sakamoto, J., Boothe, D., et al. (2012). Proof of concept of microbiome-metabolome analysis and delayed gluten exposure on celiac disease autoimmunity in genetically at-risk infants. PLoS One *7*, e33387.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Sharp, J.L. (1954). The growth of Candida albicans during antibiotic therapy. Lancet *266*, 390–392.

Shehab, N., Patel, P.R., Srinivasan, A., and Budnitz, D.S. (2008). Emergency department visits for antibiotic-associated adverse events. Clin. Infect. Dis. *47*, 735–743.

Shields-Cutler, R.R., Hillmann, B., Al-Ghalith, G., and Knights, D. (2018). Predicted secondary metabolite profiles for microbiome datasets.

Shipp, M.A., Ross, K.N., Tamayo, P., Weng, A.P., Kutok, J.L., Aguiar, R.C.T., Gaasenbeek, M., Angelo, M., Reich, M., Pinkus, G.S., et al. (2002). Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat. Med. *8*, 68–74.

Shu, Q., Lin, H., Rutherfurd, K.J., Fenwick, S.G., Prasad, J., Gopal, P.K., and Gill, H.S. (2000). Dietary Bifidobacterium lactis (HN019) enhances resistance to oral Salmonella typhimurium infection in mice. Microbiol. Immunol. *44*, 213–222.

Sjögren, Y.M., Jenmalm, M.C., Böttcher, M.F., Björkstén, B., and Sverremark-Ekström, E. (2009). Altered early infant gut microbiota in children developing allergy up to 5 years of age. Clin. Exp. Allergy *39*, 518–526.

Slavin, J.L. (2005). Dietary fiber and body weight. Nutrition *21*, 411–418.

Smith, C., and Franzen-Castle, L. (2012). Dietary acculturation and body composition predict American Hmong children's blood pressure. Am. J. Hum. Biol. *24*, 666–674.

Smith, M.I., Yatsunenko, T., Manary, M.J., Trehan, I., Mkakosya, R., Cheng, J., Kau, A.L., Rich, S.S., Concannon, P., Mychaleckyj, J.C., et al. (2013). Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. Science *339*, 548–554.

Smits, S.A., Leach, J., Sonnenburg, E.D., Gonzalez, C.G., Lichtman, J.S., Reid, G., Knight, R., Manjurano, A., Changalucha, J., Elias, J.E., et al. (2017). Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. Science *357*, 802–806.

Snijder, M.B., Galenkamp, H., Prins, M., Derks, E.M., Peters, R.J.G., Zwinderman, A.H., and Stronks, K. (2017). Cohort profile: the Healthy Life in an Urban Setting (HELIUS) study in Amsterdam, The Netherlands. BMJ Open *7*, e017873.

Sonnenberg, G.F., Monticelli, L.A., Alenghat, T., Fung, T.C., Hutnick, N.A., Kunisawa, J., Shibata, N., Grunberg, S., Sinha, R., Zahm, A.M., et al. (2012). Innate lymphoid cells promote anatomical containment of lymphoid-resident commensal bacteria. Science *336*, 1321–1325.

Sonnenburg, E.D., Smits, S.A., Tikhonov, M., Higginbottom, S.K., Wingreen, N.S., and Sonnenburg, J.L. (2016). Diet-induced extinctions in the gut microbiota compound over generations. Nature *529*, 212–215.

Speek, A.J., Speek-Saichua, S., and Schreurs, W.H.P. (1991). Determination of macronutrient and micronutrient levels in thai foods: An evaluation of the Thai Food Composition Table. Food Chem. *40*, 251–262.

Stefka, A.T., Feehley, T., Tripathi, P., Qiu, J., McCoy, K., Mazmanian, S.K., Tjota, M.Y., Seo, G.-Y., Cao, S., Theriault, B.R., et al. (2014). Commensal bacteria protect against food allergen sensitization. Proc. Natl. Acad. Sci. U. S. A. *111*, 13145–13150.

Subramanian, S., Huq, S., Yatsunenko, T., Haque, R., Mahfuz, M., Alam, M.A., Benezra, A., DeStefano, J., Meier, M.F., Muegge, B.D., et al. (2014). Persistent gut microbiota immaturity in malnourished Bangladeshi children. Nature *510*, 417–421.

Suez, J., Korem, T., Zeevi, D., Zilberman-Schapira, G., Thaiss, C.A., Maza, O., Israeli, D., Zmora, N., Gilad, S., Weinberger, A., et al. (2014). Artificial sweeteners induce glucose intolerance by altering the gut microbiota. Nature *514*, 181–186.

Suzek, B.E., Wang, Y., Huang, H., McGarvey, P.B., Wu, C.H., and UniProt Consortium (2015). UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics *31*, 926–932.

Sze, M.A., and Schloss, P.D. (2016). Looking for a Signal in the Noise: Revisiting Obesity and the Microbiome. MBio *7*.

Tanaka, S., Kobayashi, T., Songjinda, P., Tateyama, A., Tsubouchi, M., Kiyohara, C., Shirakawa, T., Sonomoto, K., and Nakayama, J. (2009). Influence of antibiotic exposure in the early postnatal period on the development of intestinal microbiota. FEMS Immunol. Med. Microbiol. *56*, 80–87.

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Ciufo, S., and Li, W. (2013). The NCBI handbook. National Center for Biotechnology Information.

Team, R.C., and Others (2013). R: A language and environment for statistical computing.

Thompson, E.A., Beura, L.K., Nelson, C.E., Anderson, K.G., and Vezys, V. (2016). Shortened Intervals during Heterologous Boosting Preserve Memory CD8 T Cell Function but Compromise Longevity. J. Immunol. *196*, 3054–3063.

Tippett, K.S., Enns, C.W., and Moshfegh, A.J. (1999). Food consumption surveys in the US Department of Agriculture. Nutr. Today *34*, 33–46.

Torack, R.M. (1957). Fungus infections associated with antibiotic and steroid therapy. Am. J. Med. *22*, 872–882.

Tormo-Badia, N., Håkansson, Å., Vasudevan, K., Molin, G., Ahrné, S., and Cilio, C.M. (2014). Antibiotic Treatment of Pregnant Non-Obese Diabetic Mice Leads to Altered Gut Microbiota and Intestinal Immunological Changes in the Offspring. Scand. J. Immunol. *80*, 250–260.

Trasande, L., Blustein, J., Liu, M., Corwin, E., Cox, L.M., and Blaser, M.J. (2013). Infant antibiotic exposures and early-life body mass. Int. J. Obes. *37*, 16–23.

Tsuchiya, T., Fukuda, S., Hamada, H., Nakamura, A., Kohama, Y., Ishikawa, H., Tsujikawa, K., and Yamamoto, H. (2003). Role of gamma delta T cells in the inflammatory response of experimental colitis mice. J. Immunol. *171*, 5507–5513.

Tsuji, N.M., and Kosaka, A. (2008). Oral tolerance: intestinal homeostasis and antigen-specific regulatory T cells. Trends Immunol. *29*, 532–540.

Turnbaugh, P.J., and Gordon, J.I. (2009). The core gut microbiome, energy balance and obesity. J. Physiol. *587*, 4153–4158.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. Nature *444*, 1027–1031.

Turnbaugh, P.J., Ridaura, V.K., Faith, J.J., Rey, F.E., Knight, R., and Gordon, J.I. (2009a). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci. Transl. Med. *1*, 6ra14.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., et al. (2009b). A core gut microbiome in obese and lean twins. Nature *457*, 480–484.

Ubeda, C., Taur, Y., Jenq, R.R., Equinda, M.J., Son, T., Samstein, M., Viale, A., Socci, N.D., van den Brink, M.R.M., Kamboj, M., et al. (2010). Vancomycin-resistant Enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. J. Clin. Invest. *120*, 4332–4341.

United States Department of Agriculture Agricultural Research Service USDA Food Composition Databases.

U.S. Department of State, Bureau of Population, Refugees, and Migration Worldwide Refugee Admissions Processing System (U.S. Department of State).

Vaarala, O., Atkinson, M.A., and Neu, J. (2008). The “perfect storm” for type 1 diabetes: the complex interplay between intestinal microbiota, gut permeability, and mucosal immunity. Diabetes *57*, 2555–2562.

Vaishampayan, P.A., Kuehl, J.V., Froula, J.L., Morgan, J.L., Ochman, H., and Francino, M.P. (2010). Comparative metagenomics and population dynamics of the gut microbiota in mother and infant. Genome Biol. Evol. *2*, 53–66.

Valladares, R., Sankar, D., Li, N., Williams, E., Lai, K.-K., Abdelgeliel, A.S., Gonzalez, C.F., Wasserfall, C.H., Larkin, J., Schatz, D., et al. (2010). Lactobacillus johnsonii N6.2 mitigates the development of type 1 diabetes in BB-DP rats. PLoS One *5*, e10507.

Van Der Velden, V., Laan, M.P., Baert, M., De Waal Malefyt, R., Neijens, H.J., and Savelkoul, H. (2001). Selective development of a strong Th2 cytokine profile in high-risk children who develop atopy: risk factors and regulatory role of IFN-γ, IL-4 and IL-10. Clinical & Experimental Allergy *31*, 997–1006.

Vonarbourg, C., Mortha, A., Bui, V.L., Hernandez, P.P., Kiss, E.A., Hoyler, T., Flach, M., Bengsch, B., Thimme, R., Hölscher, C., et al. (2010). Regulated expression of nuclear receptor RORγt confers distinct functional fates to NK cell receptor-expressing RORγt(+) innate lymphocytes. Immunity *33*, 736–751.

Walker, P.F., Barnett, E.D., Hauck, F.R., and Pearson, R.D. (2008). Immigrant Medicine. Emerg. Infect. Dis. *14*, 1007–1008.

Wang, T., Cai, G., Qiu, Y., Fei, N., Zhang, M., Pang, X., Jia, W., Cai, S., and Zhao, L. (2012). Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. ISME J. *6*, 320–329.

Wang, Y., Hoenig, J.D., Malin, K.J., Qamar, S., Petrof, E.O., Sun, J., Antonopoulos, D.A., Chang, E.B., and Claud, E.C. (2009). 16S rRNA gene-based analysis of fecal microbiota from preterm infants with and without necrotizing enterocolitis. ISME J. *3*, 944–954.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., Lee, S.Y., Fischbach, M.A., Müller, R., Wohlleben, W., et al. (2015). antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. *43*, W237–W243.

Wen, L., Ley, R.E., Volchkov, P.Y., Stranges, P.B., Avanesyan, L., Stonebraker, A.C., Hu, C., Wong, F.S., Szot, G.L., Bluestone, J.A., et al. (2008). Innate immunity and intestinal microbiota in the development of Type 1 diabetes. Nature *455*, 1109–1113.

Wickens, K., Pearce, N., Crane, J., and Beasley, R. (1999). Antibiotic use in early childhood and the development of asthma. Clin. Exp. Allergy *29*, 766–771.

Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S.A., Bewtra, M., Knights, D., Walters, W.A., Knight, R., et al. (2011). Linking long-term dietary patterns with gut microbial enterotypes. Science *334*, 105–108.

Wu, H.-J., Ivanov, I.I., Darce, J., Hattori, K., Shima, T., Umesaki, Y., Littman, D.R., Benoist, C., and Mathis, D. (2010). Gut-residing segmented filamentous bacteria drive autoimmune arthritis via T helper 17 cells. Immunity *32*, 815–827.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al. (2012). Human gut microbiome viewed across age and geography. Nature *486*, 222–227.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. *40*, W445–W451.

Yoshioka, H., Iseki, K., and Fujita, K. (1983). Development and differences of intestinal flora in the neonatal period in breast-fed and bottle-fed infants. Pediatrics *72*, 317–321.

Zhang, X.-S., Li, J., Krautkramer, K., Badri, M., Battaglia, T., Ng, S., Sibley, R.A., Koh, H., Li, Y., Borbet, T.C., et al. (2018). Antibiotic-induced acceleration of Type 1 diabetes alters intestinal innate pathway maturation. Elife.

Appendix

Appendix A. Survey form.

1. What is your birthdate? _____
2. Do you live in public housing? _____
3. Do you receive medical assistance? _____
4. Do you have any children? _____
   If yes, do they qualify for free lunch? _____

5. How many years have you gone to school?

_____ English as a Second Language
_____ Less than High School
_____ High School
_____ College
_____ Graduate School

6. What is your ethnicity?
   ____ Hmong
   ____ Karen
   ____ Other: _____

7. What is your religion?
   ____ Buddhism
   ____ Christianity
   ____ Hmong animism
   ____ Islam
   ____ Other: _____
   ____ None

8. Where were you born? Please write name: camp/village/city, province, country

_____

9. What type of place was that?

_____ a. Refugee camp.
_____ b. Rural village.
_____ c. Urban dwelling.
_____ d. Other. _____

10. When did you arrive in the US? _____

11. Where did you live just before you arrived in the US?
Please write name: camp/village/city, province, country

_____

12. What type of place was this place?

_____ a. Refugee camp.
_____ b. Rural village.
_____ c. Urban dwelling.
_____ d. Other. _____

13. How long did you live there? _____

14. In this location, what did you usually eat per day?

	Food and Drink	**Serving Size**
Breakfast		
Snack		
Lunch		
Snack		
Dinner		
Snack		

15. What did you eat yesterday (including overnight)?

	Food and Drink	**Serving Size**
Breakfast		
Snack		
Lunch		
Snack		
Dinner		
Snack		

16. Do you use tobacco - whether smoked cigarettes/cigars/pipes, or chewed tobacco?
    a. Never
    b. Currently- ever day use
    c. Currently- occasional use
    d. Used to use, now quit.

17. Do you consume alcohol - whether beer, wine, whiskey or other liquors?
    a. Never
    b. Currently – 1-3 drinks every day
    c. Currently - 1-3 drinks every week
    d. Currently - 1-3 drinks every month
    e. Currently - 1-3 drinks occasionally (less than once a month)
    f. Used to drink, now quit.

Health Metrics

18. Height
_____feet _____ inches

19. Weight
_____ lb

20. Waist Circumference
_____ inches

21. Are you taking any medications? _____(Yes/No)
If yes, please list: _____

22. Do you know if you were breastfed as a child?
If yes, do you know for how long?

Appendix B. Fecal sampling instructions

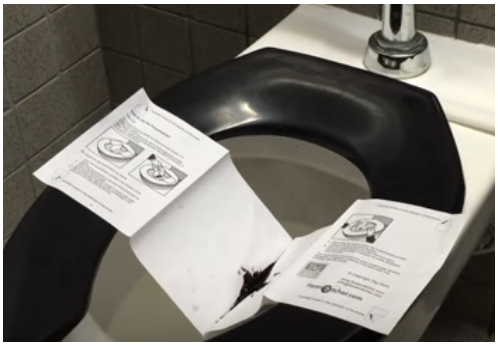
How to Get Stool Sample

Do this 1-4 days before you are going to bring us the sample.
Urinate (pee) in the toilet before starting.

For a video demonstration, please watch:
http://z.umn.edu/impenglish

The kit has these things:

1. A FecesCatcher paper
2. A tube
3. A stick
4. Latex-free gloves
5. A small envelope
6. A bag
7. A paper towel



This is what you have to do:

1. Gently open the FecesCatcher paper. Put the sticky ends on the left and right sides of the toilet. Make sure the center of the paper does not touch the water. Press hard on the sticky ends so they stick to the toilet.
2. Have your normal bowel movement onto the paper. Make sure that nothing touches your stool, like toilet paper, water or urine.
3. Wash and dry your hands. Put on the gloves.
4. Unscrew the cap on the tube. Do not spill the liquid. Put the lid on the counter, with the inside facing up, towards the ceiling.
5. Open the package with the stick.
6. Get a small amount of stool on the stick, about the size of 3-4 grains of rice.
7. Scrape the stool into the tube.
8. Repeat this again to put another small amount of stool into the tube.
9. Use the stick to evenly mix the stool and liquid in the tube. Throw away the stick.
10. Tightly screw the lid on the tube.
11. Place the tube in the small envelope and seal it. Place the envelope in the bag with the paper towel.
12. Remove the FecesCatcher paper from the sides of the toilet. Shake the FecesCatcher paper so the stool falls into the toilet.
13. Either put the paper in a waste basket **(recommended for all toilets, especially at home)** Or: Put the paper into the toilet water, let it soak for 1-2 minutes and flush it down the toilet.
14. Remove the gloves and put into the trash. Wash your hands.
15. **Keep the fecal sample at room temperature and away from children until you can return it to us.**