Title: U.S. immigration westernizes the human gut microbiome

Authors: Pajau Vangay[1], Abigail J. Johnson[2], Tonya L. Ward[2], Gabe A. Al-Ghalith[1], Robin R. Shields-Cutler[2], Benjamin M. Hillmann[3], Sarah K. Lucas[4], Lalit K. Beura[4], Emily A. Thompson[4], Lisa M. Till[5], Rodolfo Batres[6], Bwei Paw[6], Shannon L. Pergament[6], Pimpanitta Saenyakul[6], Mary Xiong[6], Austin Kim[7], Grant Kim[8], David Masopust[4], Eric C. Martens[9], Chaisiri Angkurawaranon[10], Rose McGready[11,12], Purna C. Kashyap[5], Kathleen A. Culhane-Pera[6], Dan Knights[1,2,3]*

Affiliations

[1]Bioinformatics and Computational Biology Program, University of Minnesota, Minneapolis, Minnesota, 55455, United States

[2]Biotechnology Institute, University of Minnesota, Minneapolis, Minnesota, 55455, United States

[3]Department of Computer Science and Engineering, University of Minnesota, Minneapolis, Minnesota, 55455, United States

[4]Center for Immunology, Department of Microbiology and Immunology, University of Minnesota, Minneapolis, Minnesota, 55455, United States

[5]Division of Gastroenterology and Hepatology, Department of Internal Medicine, Mayo Clinic, Rochester, Minnesota, 55902, United States

[6]Somali, Latino, and Hmong Partnership for Health and Wellness, West Side Community Health Services, St. Paul, Minnesota, 55106, United States

[7]Department of Mathematics, Statistics, and Computer Science, Macalester College, St. Paul, Minnesota, 55105, United States

[8]College of Biological Sciences, University of Minnesota, Minneapolis, Minnesota, 55455, United States

[9]Department of Microbiology & Immunology, University of Michigan, Ann Arbor, Michigan, 48109, United States

[10]Department of Family Medicine, Faculty of Medicine, Chiang Mai University, Chiang Mai, 50200, Thailand

[11]Shoklo Malaria Research Unit, Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, 63110 Mae Sot, Thailand

[12]Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Old Road Campus, Oxford, OX3 7BN, United Kingdom

*Correspondence and Lead Contact: dknights@umn.edu

Summary

Many United States immigrant populations develop metabolic diseases post-immigration, but the causes are not well understood. Although the microbiome plays a role in metabolic disease, there have been no studies measuring the effects of U.S. immigration on the gut microbiome. We collected stool, dietary recalls, and anthropometrics from 514 Hmong and Karen individuals living in Thailand and the U.S., including first- and second-generation immigrants and 19 Karen individuals sampled before and after immigration, as well as from 36 U.S.-born Caucasian individuals. Using 16S and deep shotgun metagenomic DNA sequencing, we found that migration from a non-Western country to the U.S. is associated with immediate loss of gut microbiome diversity and function, with U.S.-associated strains and functions displacing native strains and functions. These effects increase with duration of U.S. residence, and are compounded by obesity and across generations.

Keywords

Microbiome, microbiota, immigration, immigrant health, refugee health, obesity, Bacteroides, Prevotella, acculturation, metagenomics

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) after relocation to the U.S. (Careyva et al., 2015). 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 and have similar disease risks after migration to the U.S. Although, to our knowledge, disaggregated data on long-term health changes in ethnic Karen from Burma do not yet exist, refugee children from Burma exhibit the steepest BMI increase after relocation, compared with other refugee and non-refugee children (Dawson-Hahn et al., 2016). 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) although they do not fully explain risk.

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) in 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 group (foreign-born U.S. residents) 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 were compounded across generations by birth in the U.S. Finally, we followed a unique longitudinal cohort of 19 Karen refugees for up to 9 months beginning immediately before or after arrival in the U.S to measure the short-term effects of U.S. immigration.

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 1A). We also recruited healthy Caucasian American female individuals to serve as U.S. controls (Controls; n = 36) (Figure 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 S1). 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 S1).

During face-to-face enrollment, bilingual-bicultural research team members collected migration and medical histories (Table S2), anthropometrics (weight, height, waist circumference), 24-hour dietary recalls, and a single stool sample 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 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 from whom we collected initial samples in a refugee camp in Thailand prior to their relocation to the U.S. As a result of our recruitment efforts, we collected a total of 673 stool samples comprising 531 single- and 142 multiple-time-point collections. Because we stratified recruitment by only a single BMI threshold of 25, examining the ratio of obese (BMI ≥ 30) to overweight (BMI between 25 and 29.9) individuals provided an estimate of the prevalence of obesity across groups. Consistent with the previously observed high rate of obesity in U.S. immigrants (see Introduction), we saw that obesity prevalence increased after a decade in the U.S. (Figure 1B).

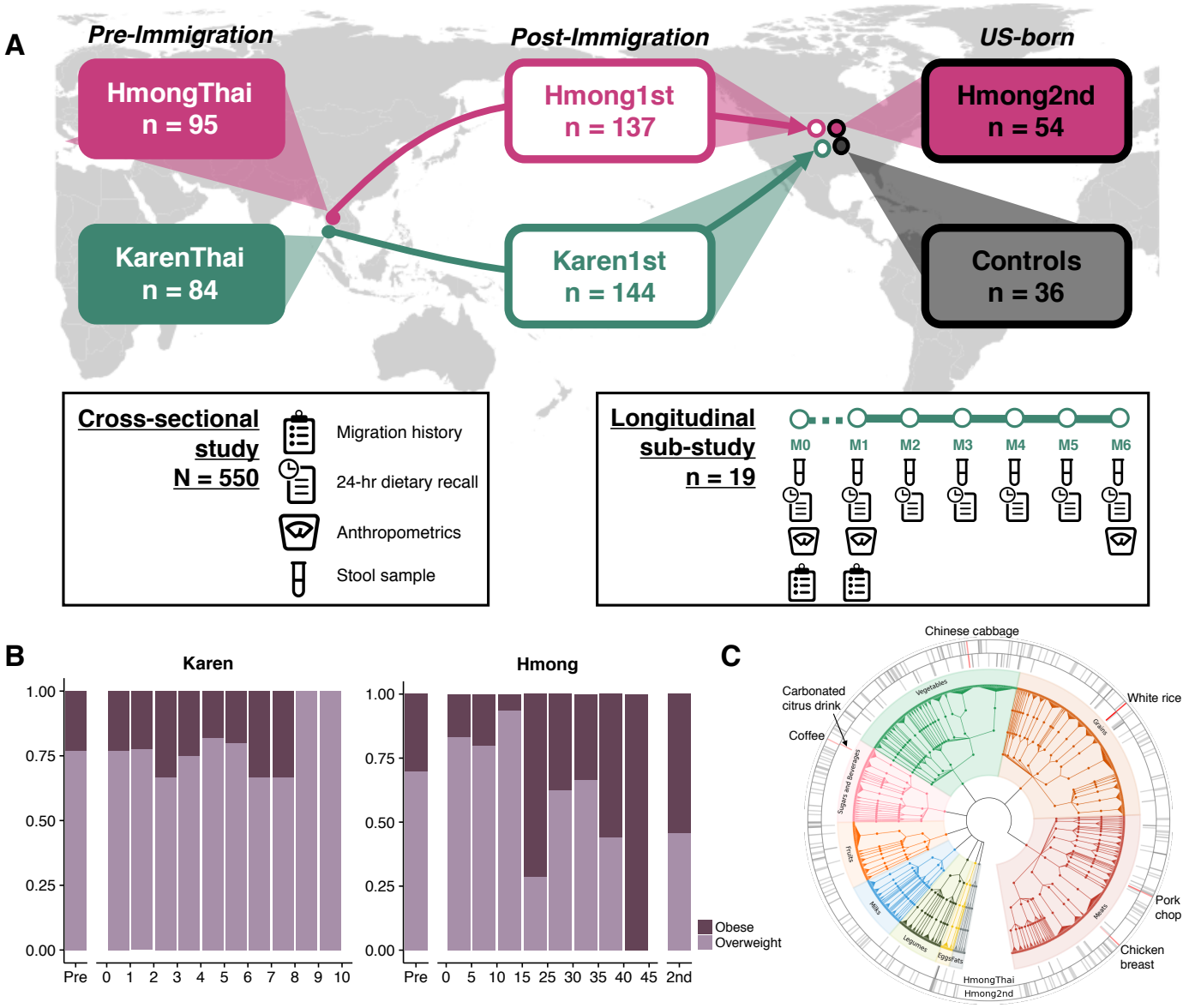

Figure 1. Assembly of a multi-generational Asian American cohort, while accounting for BMI and diet
(A) Experimental design for cross-sectional and longitudinal cohorts.
(B) Ratios of overweight-to-obese individuals across sample groups and over time in the U.S., separated by ethnicity due to differences 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 foods highlighted in red: 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.

To be able to measure the association of observed changes in the gut microbiome with changes in 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, and 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 also considered the relatedness of individual foods when assessing the similarity of dietary profiles across individuals. This approach relied on the 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 tree format where more closely related foods were grouped together (Figure 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 S3) were researched individually for macronutrient content before entry into SuperTracker, manually assigned new food codes, and inserted into the hierarchical food taxonomy, allowing us to account for all foods reported by all participants. This hierarchical food tree also allowed us to compare dietary diversity between sample groups, showing 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 1C).

U.S. immigration is associated with loss of native gut microbiome species

We performed amplicon-based sequencing of the 16S rRNA gene V4 region on 550 stool samples (one sample per participant). Principal coordinates analysis (PCoA) of unweighted UniFrac distances (Lozupone et al., 2011) revealed that Hmong and Karen ethnic groups harbor distinct gut microbial compositions regardless of country of residence, yet their microbiomes converge toward Caucasian American microbiomes after relocating to the U.S. (ANOSIM $R=0.25$, $P=0.001$). The first two principal coordinate axes show that second-generation Hmong and Caucasian American microbiomes share nearly identical cluster centroids (Figure 2A), although Caucasian American microbiomes have lower inter-individual variation. We also found that both diversity and richness are highest in the Thailand groups and decrease with each generation of residence in the U.S. (Tukey's HSD, $p < 0.01$, Figure 2B). As with other studies (Sze and Schloss, 2016; Turnbaugh and Gordon, 2009), we found that lower phylogenetic

diversity was associated with obesity across all major study groups (unbalanced two-way ANOVA, P = 0.0044, Figure 2B). This trend persisted after stratification by ethnicity (Tukey's HSD, $p < 0.01$, Figure S2). Interestingly, the median richness of obese individuals in Thailand was still higher than the median richness of any lean group in the U.S. (Figure 2B). These findings suggest that both obesity and residency in the U.S. are independently associated with loss of microbiome diversity, and that immigration has a stronger effect than obesity on microbiome diversity. Furthermore, we observed a consistent loss of certain native bacterial operational taxonomic units (OTUs) among first-generation Hmong (Figure 2C). Although 7 of the 10 most prevalent OTUs found in HmongThai were also found at similar levels in Hmong1st, others such as *otu1812* (*Faecalibacterium prausnitzii*) incurred a 45% loss in prevalence (Fisher's exact test, FDR-corrected q = 3.05E-14) (Table S4). Overall, we found 28 OTUs, or 10.5% of all OTUs in 75% of HmongThai, that incurred at least a 50% loss in prevalence among first-generation Hmong, with more than half of them belonging to the genus *Prevotella* (Table S4).

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 found that the Western-associated genus *Bacteroides* displaces the non-Western-associated genus *Prevotella* across generations in the U.S. (Figure 3A). The ratio of *Bacteroides* to *Prevotella* was lowest in Thailand-resident individuals, highest in U.S.-born Caucasian Controls, and increased in a stepwise fashion from first-generation Karen, to first-generation Hmong, to second-generation Hmong (unbalanced two-way ANOVA, Resident Continent $P=3.4e{-}13$, Birth Continent $P=0.00085$, Ethnicity $P=5.5e{-}12$). 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 within *Bacteroides* and *Prevotella*. We aligned shotgun metagenomic sequences against a custom database that included 256 *Bacteroides* genomes and 153 *Prevotella* genomes isolated from diverse body sites and habitats, retaining any *Bacteroides* and *Prevotella* strains with at least 50% genome coverage within at least one sample. We found that U.S. Controls have varied *Bacteroides* strain profiles, while those with *Prevotella* tend have only a single strain of *P. copri* (Figure 3B).

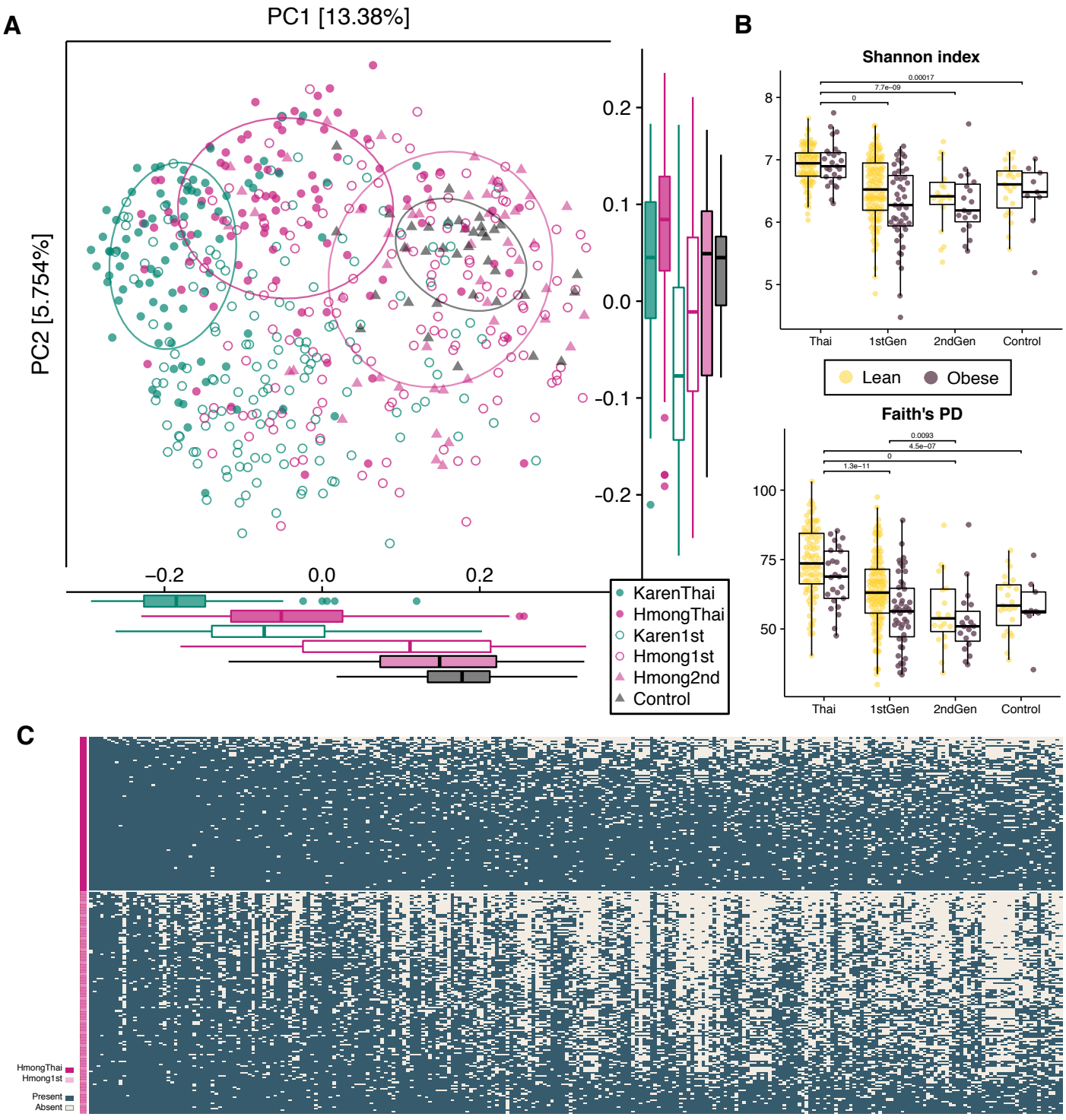

Figure 2. Loss of diversity and native bacterial taxa with time in the U.S.
(A) Principal coordinate analysis (PCoA) of unweighted UniFrac distances between bacterial communities of cross-sectional participants reveals that phylogenetic variation is strongly explained by sample group (ANOSIM R=0.25, P=0.001). 95% standard error ellipses are shown around Hmong and Karen in Thailand, second-generation Hmong, 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, $p < 0.01$). Using an unbalanced 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 at least 75% of HmongThai samples (See Table S4 for taxonomic assignments, mean group prevalences, and statistics).

Conversely, Thailand-based individuals carry up to 4 strains of *Prevotella*, with low abundance and generally low genomic coverage of *Bacteroides* strains, although we may have observed lower coverage of *Bacteroides* strains in those subjects due to more limited characterization of strains specific to Thailand residents in the current reference genome databases. 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 some *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 (Abubucker et al., 2012) in our shotgun metagenomics samples (ANOVA, FDR-corrected $q < 0.10$, Figure S3A). In long-term-resident first-generation Hmong, we observed increases in relative abundances of 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 β-(1,4)-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 and annotated carbohydrate-degrading enzymes (CAZymes) (Lombard et al., 2014; Yin et al., 2012). We found that the observed shifts in strain-level composition and functional pathways were accompanied by significant shifts in several types of CAZymes, including differential abundance of 58 CAZymes across the HmongThai, Hmong1st, and Control groups (Mann Whitney U test, FDR-corrected $q < 0.05$, Figure 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 may be associated with loss of dietary fiber sources that promote persistence of the organisms that harbor these enzymes, followed by loss or reduction of the ability of the microbiota to degrade these dietary fibers. In order to determine the organisms most likely contributing these CAZymes, we identified all shotgun metagenomics sequences that matched both a de-novo assembled CAZyme-containing scaffold and one or more known reference strains in our genome database. This analysis showed that the three CAZymes were predominantly originating from *Prevotella copri* genomes (42 ± 11.1%, Figure S3B), with smaller fractions coming from *Eubacterium*

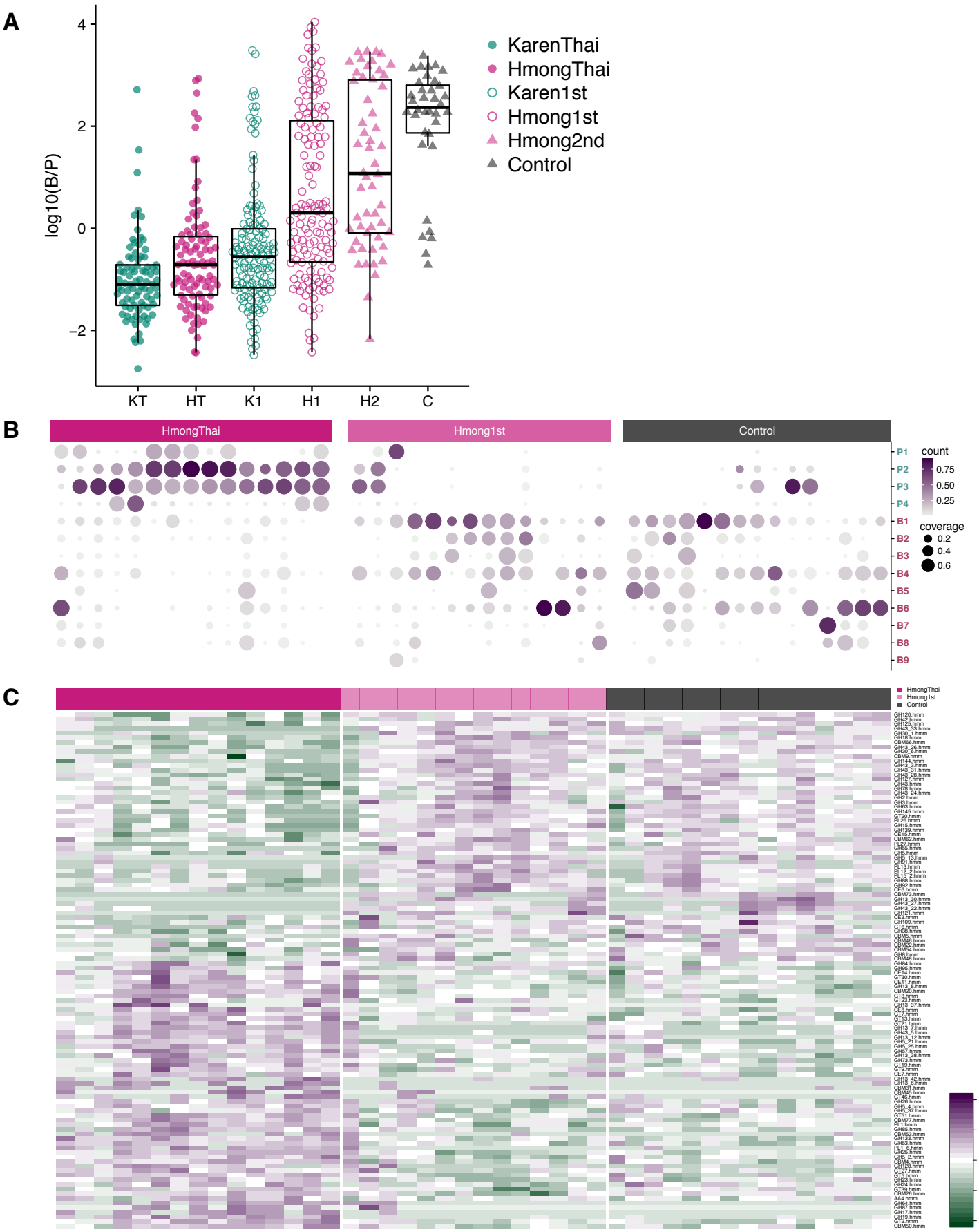

Figure 3. *Bacteroides* and *Prevotella* strain diversity and abundances

(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 S5 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 (Kruskal-Wallis test, FDR-corrected q < 0.05).

ventriosum, *Roseburia faecis*, *Blautia obeum*, *Prevotella oulurum*, and other species. This supports the hypothesis that loss of *Prevotella* strains in U.S.-resident individuals is driving loss of plant fiber degradation capability. We also observed a loss of GH5 and GH26 glycoside hydrolases from HmongThai to Hmong1st and U.S. controls, which indicates a loss of cellulose, beta-mannan and possible xyloglucan degradative potential. Beta-mannans are present in seeds, kernels, and corms, such as palm (Subrahmanyan et al., 1956), coconut (Kooiman, 1971), and konjac (Pangsri et al., 2015), and xyloglucan is found most abundantly in tamarind (Mishra and Malhotra, 2009), which interestingly are food ingredients prevalent in Southeast Asia. The loss of glycoside hydrolases for degrading cellulose, a plant cell-wall component, is another indication that the microbiota of post-immigration individuals have lost some of their ability to degrade plant-derived fibers (El Kaoutari et al., 2013).

Dietary acculturation partly explains microbiome acculturation

In our analysis of diet across sample groups, we observed significant differences in the consumption of macronutrients commonly associated with a Western diet: sugars, fats, and protein (unbalanced two-way ANOVA, $p < 0.01$, Figure 4A). Consumption of sugars and fats were associated most significantly with residency in the U.S., and protein consumption was highest among first- and second-generation Hmong when compare to the more recently arrived first-generation Karen (Tukey's HSD, $p < 0.01$, Figure 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 S4).

Our use of a hierarchical food tree enabled approximate comparisons of common American foods to non-American foods, and as a result, enabled us 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 revealed distinct separation by sample group and a gradient of increasing dietary acculturation along the first principal coordinate (ANOSIM $R=0.29$, $P=0.001$, Figure 4B). Shifts toward positive values of the first principal coordinate were 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 (Spearman's correlation, FDR-corrected $q < 0.01$, Table S6). First- and second-generation Hmong had similar food choice profiles (Figure

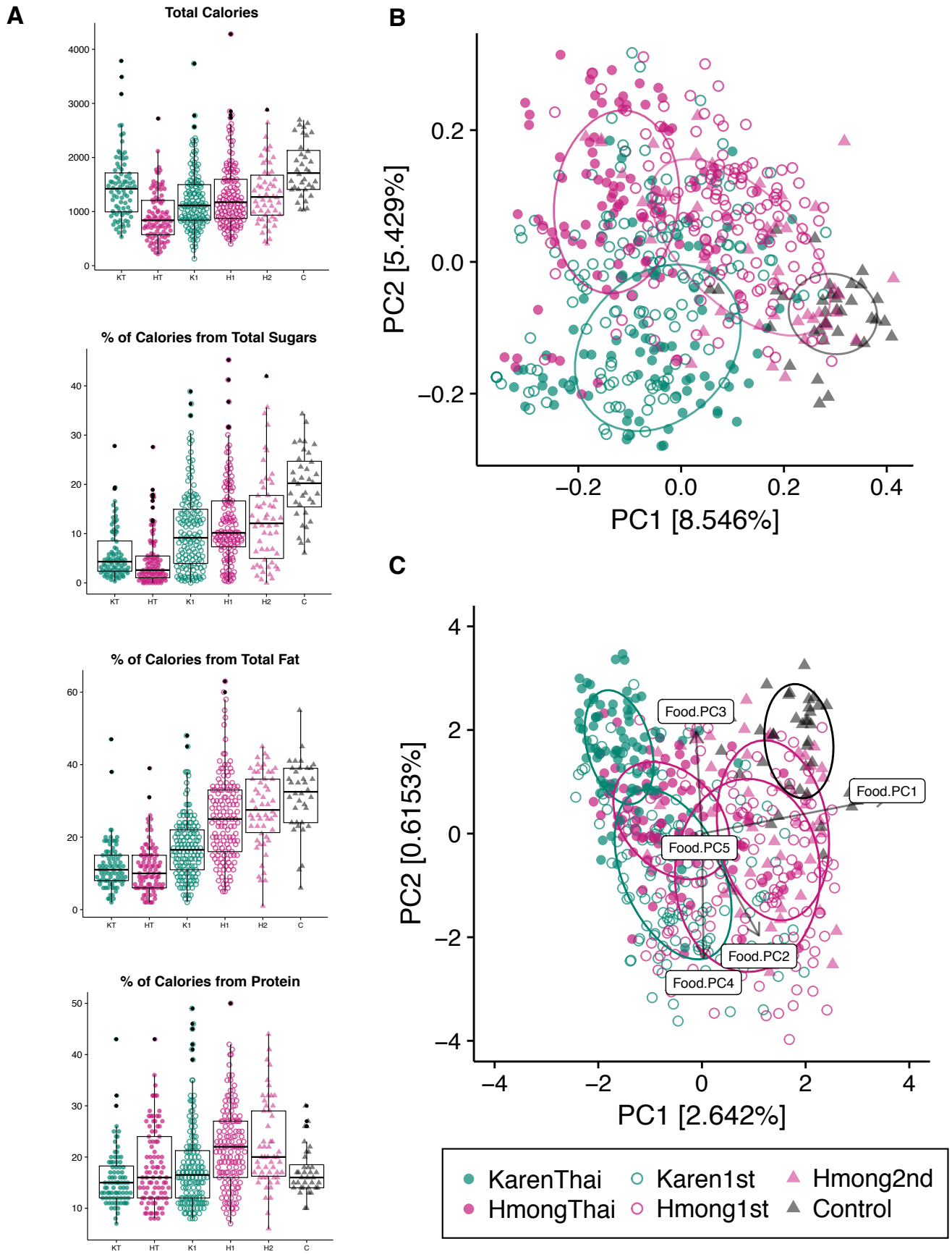

Figure 4. Dietary acculturation is detectable using novel food tree and partially explains microbiome variation
(A) Comparison of macronutrients consumption levels across sample groups. Ethnicity is significantly associated with calories (P=3.4e−05), sugars (P=0.00023), fat (P=1.3e−07), protein (P=3.2e−07), 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 2A).

4B). We confirmed that there was a high degree of shared foods between all Hmong and Karen groups, but not between these groups and the U.S. controls, using a bipartite network pairing participants with their food choices (Figure S5A). 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 S5B). The separation of U.S. Control diets from the Hmong and Thai group diets was notably different from the groupings seen in the microbiome data, where second-generation Hmong instead clustered closely with Controls (Figure 2A). The fact that the microbiome in second-generation individuals becomes more acculturated and westernized than their diet suggests that non-dietary influences, such as U.S.-based birth and early childhood development in the context of a Western diet and lifestyle, are partly responsible for the observed shifts in the microbiome.

We next wanted to understand the extent to which overall dietary variation across individuals explained overall microbiome variation across individuals. To accomplish this, we first measured the correspondence between dietary UniFrac distances and microbiome UniFrac distances and found strong similarity between the two distance matrices (Procrustes test P=0.001, n=999 permutations) (Figure S6). However, constrained ordination of the microbiome by the first 5 principal coordinates of diet variation revealed that dietary variation alone explained a relatively small fraction (16.8%) of the total variation explained in the microbiome PCoA (Figure 4C). This confirmed that although both microbiome acculturation and dietary acculturation increased with time in the U.S., diet was 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.

After finding that U.S. residence was associated with a major shift in dominant taxa in the microbiome (Figure 3A), we decided to test whether U.S. residents experienced more profound changes in microbiome composition the longer they lived in the U.S. In a PCoA of unweighted Unifrac microbiome-based distances, we found that time spent in the U.S. was strongly correlated with the first principal coordinate axis (ρ = 0.62, $p < 2.2e-16$, Figure 5A). Conversely, gut biodiversity, as measured by Faith's phylogenetic diversity, was negatively correlated with PC1 (ρ = -0.34, $p < 3.19e-09$, Figure 5B). To account for the distinct time frames of Hmong and Karen immigrant residence in the U.S., (up to 40 years versus up to 10 years, respectively), we stratified our analysis by ethnic group. We found that gut biodiversity in first-generation Hmong

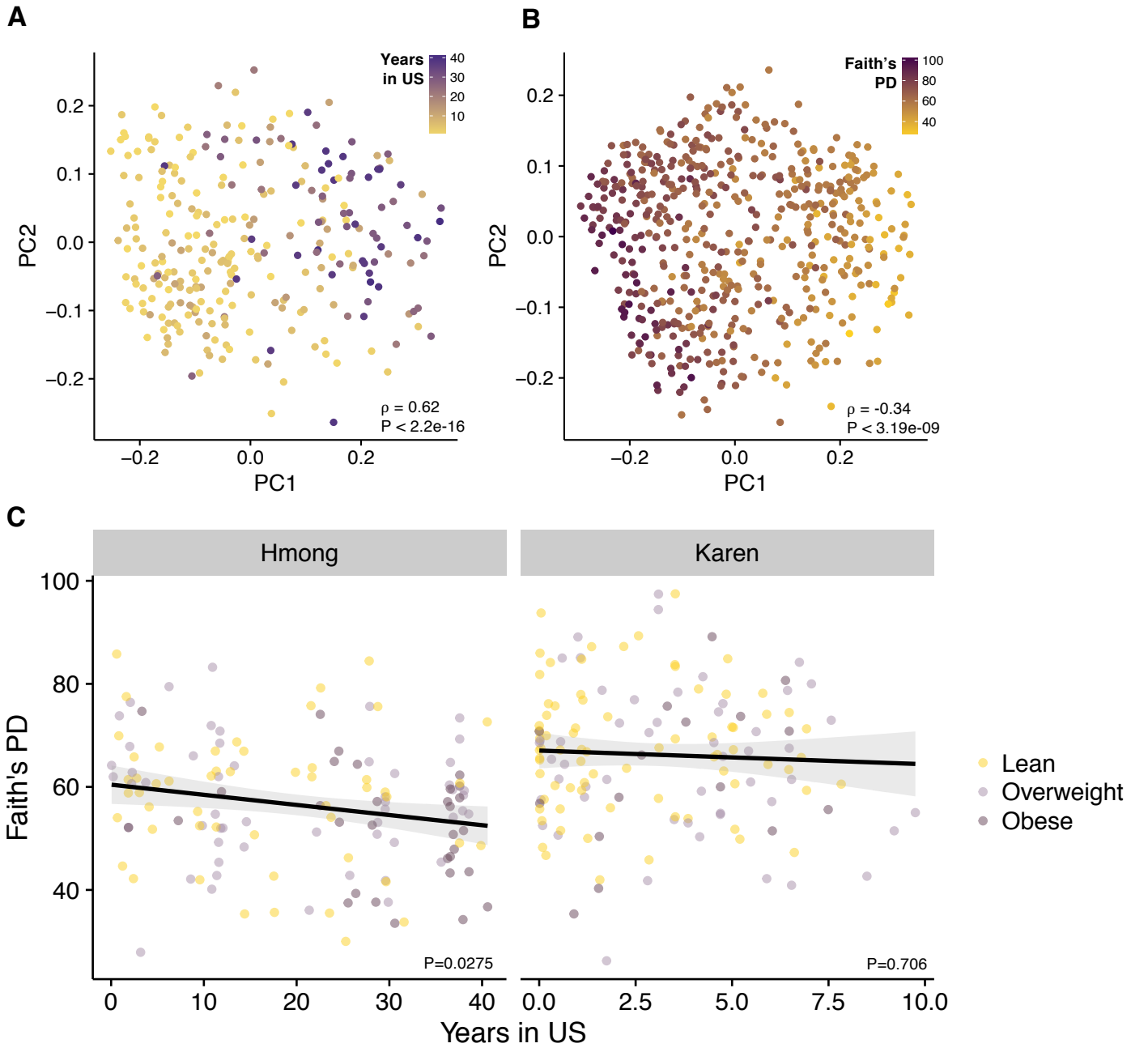

Figure 5. Gut biodiversity decreases with time spent in the U.S.
(A) Unweighted Unifrac PCoA of gut microbiomes of first-generation Hmong and Karen participants (N = 281), colored by years spent in the U.S. which ranges from 1 day to 40.6 years. PC1 is strongly correlated with the amount of time spent 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.

decreased significantly with increased time in the U.S., even while controlling for BMI (multiple linear regression, Years in US β = -0.18, P = 0.0275, Figure 5C). However, we did not find an association between gut biodiversity and time spent in the U.S. in first-generation Karen (multiple linear regression, Years in US β = -0.17, P = 0.71; BMI β = -0.27, P = 0.28, Figure 5C), which suggests that detectable changes in overall microbiome diversity may take place after 10 years of U.S. residence.

Prevotella displacement continues for more than one decade

We found that the longer immigrants spend living in the U.S., the more their microbiomes compositions diverge from their Thai counterparts (Spearman's correlation, ρ = -0.41, P = 1.3e−12) and converge toward Caucasian Controls (Spearman's correlation, ρ = 0.35, P = 1.2e−09) (Figure 6A). We find that the continuing shift in bacterial composition after decades of U.S. residence was largely driven by continuing displacement of *Prevotella* with *Bacteroides* (Spearman's correlation, ρ = 0.44, P = 8.76e-15, Figure 6B). We confirmed that this significant association persisted after stratifying the first-generation immigrants by ethnicity, despite the shorter time frame of U.S. residence in first-generation Karen (Spearman's correlation, Hmong ρ = 0.47, P = 8.16e-19; Karen ρ = 0.19, P = 0.023, Figure 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 multiple decades.

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 between first- and last-month stool samples show that within 6 to 9 months, there was a significant shift in microbial composition along the first principal coordinate axis (one sample t-test, P=0.023, Figure S7). We also found that within this short time frame, all but one participant gained weight (paired t-test, P=8.3e-05, Figure 7A), protein consumption increased (paired t-test, FDR-adjusted q=0.048, Figure 7B), while the total variety of foods consumed decreased (paired t-test, P=0.017, Figure 7C), suggesting a period of acclimation to newly available foods. Within this timeframe, we again observed the displacement of *Prevotella* by *Bacteroides* (paired t-test, P=0.0013, Figure 7D), in many cases involving a ten-fold increase in the *Bacteroides-Prevotella* ratio, indicating that microbiome westernization begins immediately after arrival to the U.S. Using deep shotgun

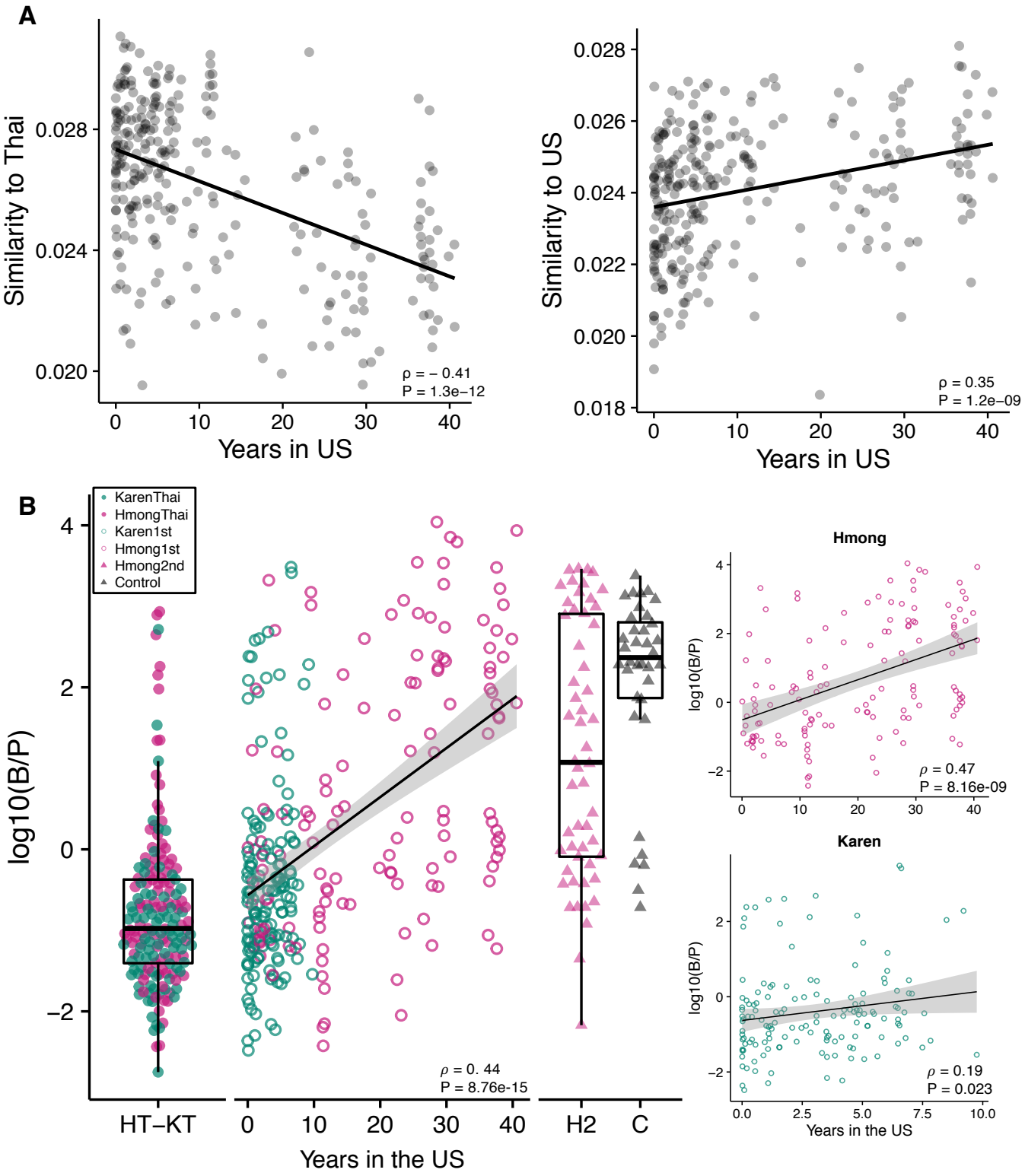

Figure 6. Prevotella displacement is observable within one decade in the U.S.
(A) Similarity (1 / Aitchison's distance) of microbiomes relative to Thai-based groups (Spearman's correlation, ρ = -0.41, P = 1.3e−12) and to Controls (Spearman's correlation, ρ = 0.35, P = 1.2e−09).
(B) Log ratio of *Bacteroides* to *Prevotella* of first-generation groups are significantly correlated to years spent in the U.S. (Spearman's correlation, ρ = 0.44, P = 8.76e-15). Significantly correlated trends persist after stratification by ethnicity (Spearman's correlation, 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).

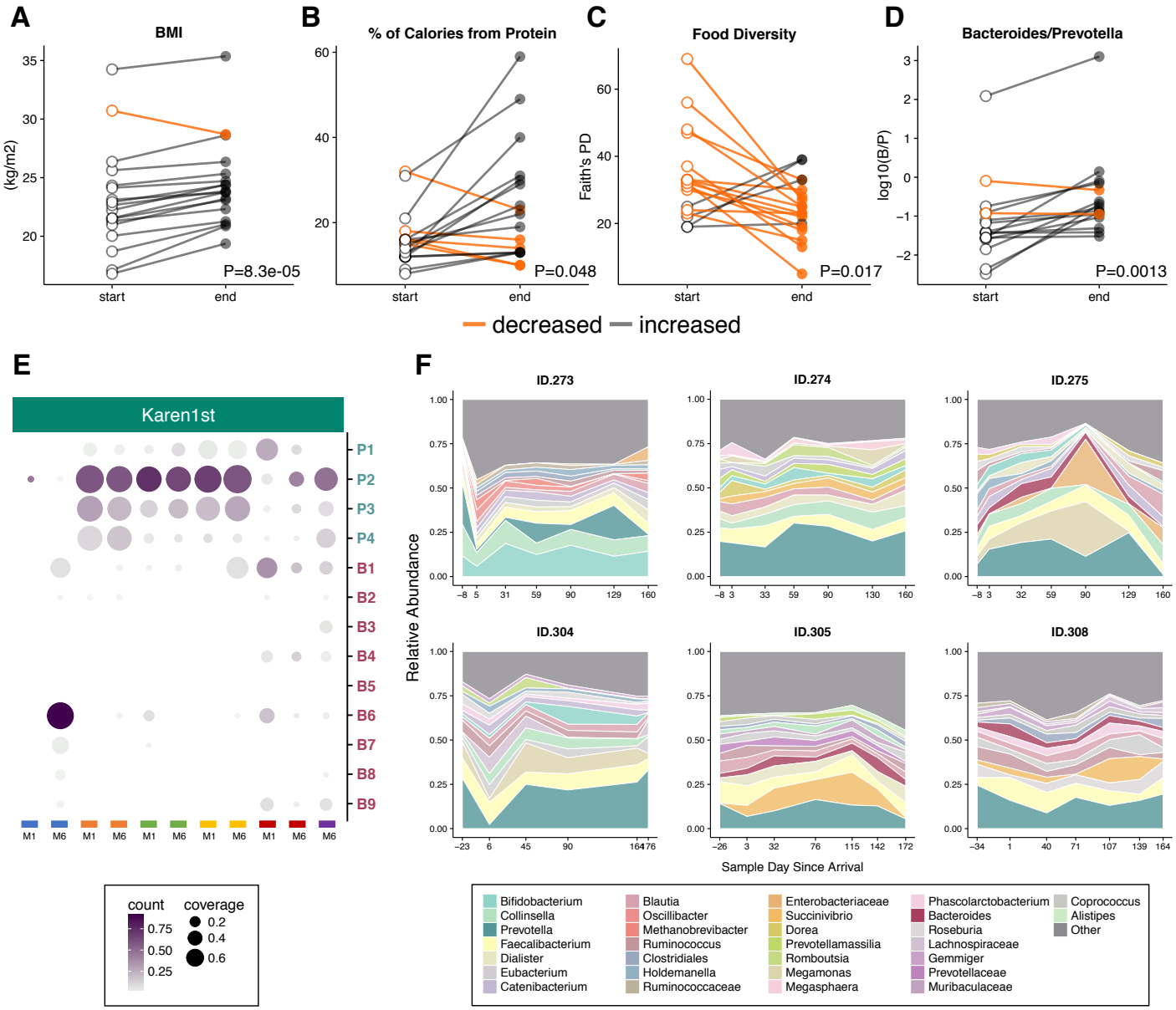

Figure 7. Longitudinal changes immediately pre- and post-arrival to the U.S.
(A) Comparison of per-participant changes between first and last months of the study in BMI (P=8.3e-05),
(B) protein consumption (P=0.048),
(C) dietary diversity (Faith's PD) (P=0.017), and
(D) *Bacteroides* to *Prevotella* ratios (P=0.0013), (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 3B (at least 50% coverage per sample across N=55 samples, see Table S5).
(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.

metagenomics sequencing on 13 samples from 6 participants as described above, we found that *Prevotella* and *Bacteroides* strain profiles remain largely stable over 6-9 months but can sometimes undergo substantial changes (subject highlighted in blue, Figure 7E). This longitudinal cohort also included six participants from whom we collected samples in Thailand, prior to their relocation to the U.S. We were able to reestablish contact with these individuals after their arrival in the U.S. in order to continue collecting longitudinal samples on a monthly basis. We analyzed their microbiome changes over the initial period of U.S. residence, and while we found examples of disruption to the gut microbiome immediately after arrival in two of these subjects (ID.273 and ID.304), we observed in general that physically relocating to the U.S. induced a variety of short-term gut microbiome responses. These responses included expansion of opportunistic pathogens (ID.305), gut disruption several months after arrival (ID.275), and stability (ID.274, ID.308) (Figure 7F). Thus, we found that short-term responses to immigration of overall microbiome composition were variable across individuals, but the displacement of dominant native taxa with dominant U.S. taxa does begin within 6 to 9 months of U.S. residence.

Discussion

This study represents the first large cohort study of the effects of migrating from a non-Western country to a Western country on the human gut microbiome. Leveraging both multi-ethnic cross-sectional and longitudinal cohorts of immigrants and refugees, including pre-immigration, first-generation immigrant, and second-generation immigrant individuals, stratified by high or low BMI, allowed an unprecedented examination of microbiome resilience and response to migration to the U.S. independent of the effects of obesity and ethnicity. In these cohorts, we observed that gut microbiome diversity, function, and strain composition are strongly impacted by U.S. immigration and that both short-term and long-term U.S. residence as well as being born in the U.S. shift 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 more than 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. Even a short period of residence in the U.S. is sufficient to induce pronounced increases, in some cases over

ten-fold, in the ratio of *Bacteroides* to *Prevotella*. We did not find sufficiently dramatic changes in dietary choices to explain this dramatic change in microbiome-dominant strains over the first 9 months of U.S. residence. This implies that certain non-dietary exposures are involved in the immediate perturbation of the microbiota. Metagenome assembly and functional annotation showed that the observed changes in bacterial strains were associated with post-immigration shifts in the profile of carbohydrate-active enzymes dominant in the gut microbiota, including a near-complete loss of certain beta-glucanases and other glycoside hydrolases that may indicate loss of ability to break down specific dietary fibers. In addition, analysis of second-generation immigrants 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.

In addition to studying the microbiome in two immigrant groups, 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). Although we observed clear patterns of dietary acculturation associated with U.S. residence, dietary variation only partly explained microbiome variation across individuals. Interestingly, the diets of second-generation immigrants remain quite distinct from the Controls, while their microbiomes do not. This is further evidence that non-dietary environmental exposures, in this case, those associated with being born and raised in the U.S., contribute to acculturation and Westernization of the microbiome.

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 factors associated with adjustment to life in the U.S. Most of these factors were not examined in the context of this study. These include changes in exposure to stress, exercise, chlorinated municipal drinking water, antibiotics, and treatment with antiparasitics. 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, our study design did not allow us to test directly whether immigration causes the observed changes in the microbiome, nor whether the changes in microbiome are directly contributing to the high incidence of obesity in U.S. immigrants.

Our findings demonstrate that U.S. immigration is associated with profound perturbations to the gut microbiome, including loss of diversity, loss of native strains, changes in fiber degradation capability, and shifts from *Prevotella* dominance to *Bacteroides* dominance. These changes begin immediately upon arrival and continue over decades of U.S. residence and are compounded in obese individuals and in second-generation individuals. These results improve our fundamental understanding of how human migration affects the microbiome, and underscore the importance of considering the impact of the gut microbiome in future research into immigrant and refugee health.

Acknowledgements

We thank all of the participants of this study. We also thank the members of our community advisory boards, who provided critical feedback throughout the 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. Our work in Thailand would not have been possible without Ntxawm Lis, Yi Lis, Blooming Zion, Htoo Lay Paw, Moo Kho Paw, See Thoj, and Wirachon Yangyuenkun. We also thank Nurul Quratulaini Abd Salim Nast, Dominique Sabas, and Max Abramson for their assistance in the lab. We thank Ryan Hunter for his advice and assistance with planning. This work was supported by the Clinical and Translational Science Institute, the Healthy Foods, Healthy Lives Institute, the Office of Diversity, and the Graduate School at the University of Minnesota.

Author Contributions

Conceptualization, P.V., K.C.P, and D.K. Methodology, P.V., K.C.P, R.B., S.P., C.A., T.W., L.T., L.B., S.L., R.H., D.M, R.M., P.K., and D.K. Software, G.A., B.H., and A.K. Formal Analysis, P.V., R.S.-C., and A.J.J. Investigation, C.A., R.M., P.V., B.P., P.S., and M.X. Data Curation, G.K. Writing - Original Draft, P.V. and D.K. Writing - Review and Editing, K.C.P, S.P., C.A., L.B., S.L., R.H., D.M, P.K., R.S.-C., P.V. and D.K. Visualization, P.V., R.S.-C., A.J.J. Supervision, D.K, K.C.P, and S.P. Project Administration, P.V. Funding Acquisition: P.V. and D.K.

Declaration of Interests

D.K. serves as CEO and holds equity in CoreBiome, a company involved in the commercialization of microbiome analysis. The University of Minnesota also has financial interests in CoreBiome under the terms of a license agreement with CoreBiome. These interests have been reviewed and managed by the University of Minnesota in accordance with its Conflict-of-Interest policies.

References

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.

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.

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.

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.

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.

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. (2016). Antibiotics, birth mode, and diet shape microbiome maturation during early life. Sci. Transl. Med. *8*, 343ra82.

Britten, P. (2013). SuperTracker Incorporates Food Composition Data into Innovative Online

Consumer Tool. Procedia Food Science *2*, 172–179.

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*.

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

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.

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.

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.

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*.

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.

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

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.

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

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

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

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

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

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

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

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

563 Gomez, A., Petrzelkova, K.J., Burns, M.B., Yeoman, C.J., Amato, K.R., Vlckova, K., Modry, D.,
564 Todd, A., Jost Robinson, C.A., Remis, M.J., et al. (2016). Gut Microbiome of Coexisting BaAka
565 Pygmies and Bantu Reflects Gradients of Traditional Subsistence Patterns. Cell Rep. *14*, 2142–
566 2153.

567 Heney, J.H., Dimock, C.C., Friedman, J.F., and Lewis, C. (2014). Pediatric refugees in Rhode
568 Island: increases in BMI percentile, overweight, and obesity following resettlement. R. I. Med. J.
569 *98*, 43–47.

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

573 Hildebrandt, M.A., Hoffmann, C., Sherrill-Mix, S.A., Keilbaugh, S.A., Hamady, M., Chen, Y.-Y.,
574 Knight, R., Ahima, R.S., Bushman, F., and Wu, G.D. (2009). High-fat diet determines the
575 composition of the murine gut microbiome independently of obesity. Gastroenterology *137*,
576 1716–1724.e1–e2.

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

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

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

584 Kooiman, P. (1971). Structures of the galactomannans from seeds of Annona muricata, Arenga

585 saccharifera, Cocos nucifera, Convolvulus tricolor, and Sophora japonica. Carbohydr. Res. *20*,
586 329–337.

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

588 Kwok, L.-Y., Zhang, J., Guo, Z., Gesudu, Q., Zheng, Y., Qiao, J., Huo, D., and Zhang, H.
589 (2014). Characterization of fecal microbiota across seven Chinese ethnic groups by quantitative
590 polymerase chain reaction. PLoS One *9*, e93631.

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

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

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

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

601 Minnesota Department of Health MDH Primary Refugee Arrival Data.

602 Mishra, A., and Malhotra, A.V. (2009). Tamarind xyloglucan : a polysaccharide with versatile
603 application potential. J. Mater. Chem. *19*, 8528–8536.

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

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

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

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

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

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

622 Pangsri, P., Piwpankaew, Y., Ingkakul, A., Nitisinprasert, S., and Keawsompong, S. (2015).

623 Characterization of mannanase from Bacillus circulans NT 6.7 and its application in
624 mannooligosaccharides preparation as prebiotic. Springerplus *4*, 771.

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

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

629 Rashidi, A., Ebadi, M., Shields-Cutler, R.R., DeFor, T.E., Al-Ghalith, G.A., Ferrieri, P., Young,
630 J.-A.H., Dunny, G.M., Knights, D., and Weisdorf, D.J. (2018). Pretransplant Gut Colonization
631 with Intrinsically Vancomycin-Resistant Enterococci (E. gallinarum and E. casseliflavus) and
632 Outcomes of Allogeneic Hematopoietic Cell Transplantation. Biol. Blood Marrow Transplant. *24*,
633 1260–1263.

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

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

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

642 Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *30*, 2068–
643 2069.

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

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

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

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

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

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

660 Subrahmanyan, V., Bains, G.S., Natarajan, C.P., and Bhatia, D.S. (1956). The carbohydrates of
661 tender kernel of palmyra palm (Borassus flabellifer, L.). Arch. Biochem. Biophys. *60*, 27–34.

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*.

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

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.

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

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.

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

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

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.

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.

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.

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.

STAR 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 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. 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. 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 studied 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, sequencing, and annotation

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). Taxonomic contributions of differentiated glycoside hydrolases were identified as follows: (1) scaffolds that contributed to GH17, GH64, GH87 were identified and respective DNA sequences were obtained and used as a reference database, (2) shotgun metagenomic reads were quality filtered as described previously, (3) quality reads were aligned against the scaffold reference database using BURST (Al-Ghalith and Knights, 2017) at 95% identity, (4) quality filtered reads from step 2 were aligned with BURST at 98% identity against the previously described custom database with taxonomy assigned from the NCBI database, (5) sequences that hit both the scaffolds reference and the custom NCBI-based reference were used to construct an OTU table.

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.

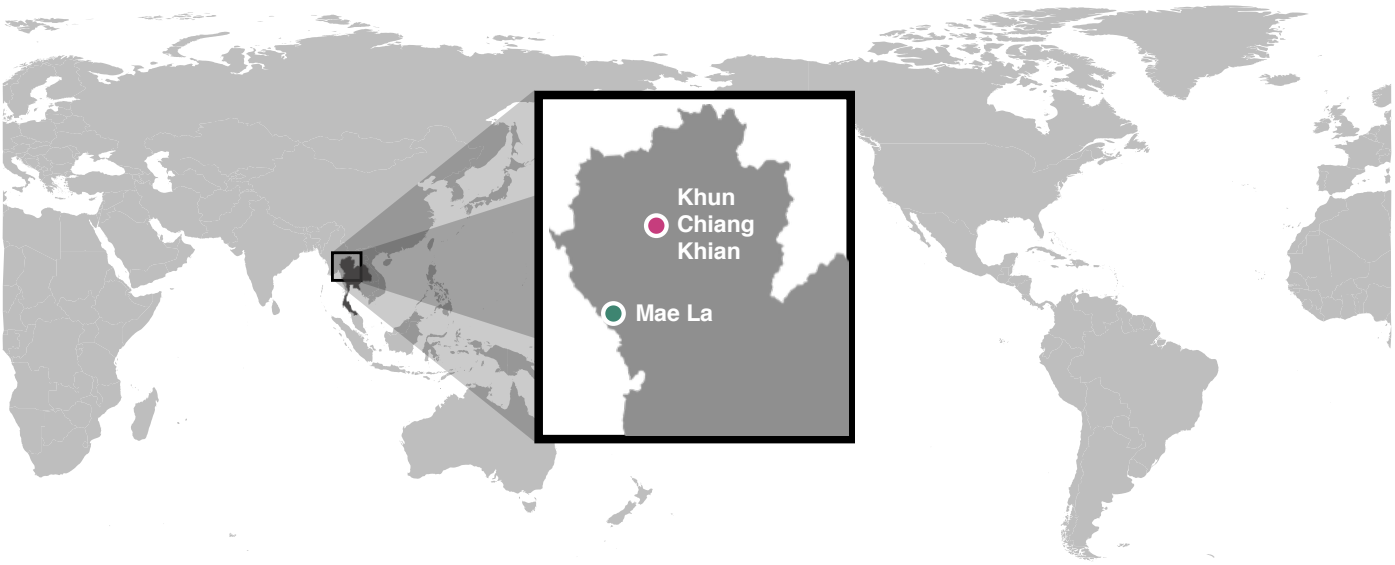

Figure S1. Geographical locations of recruitment sites in Thailand. Related to Figure 1.
Khun Chang Khian in Chiang Mai province and Mae La camp in Tak Province.

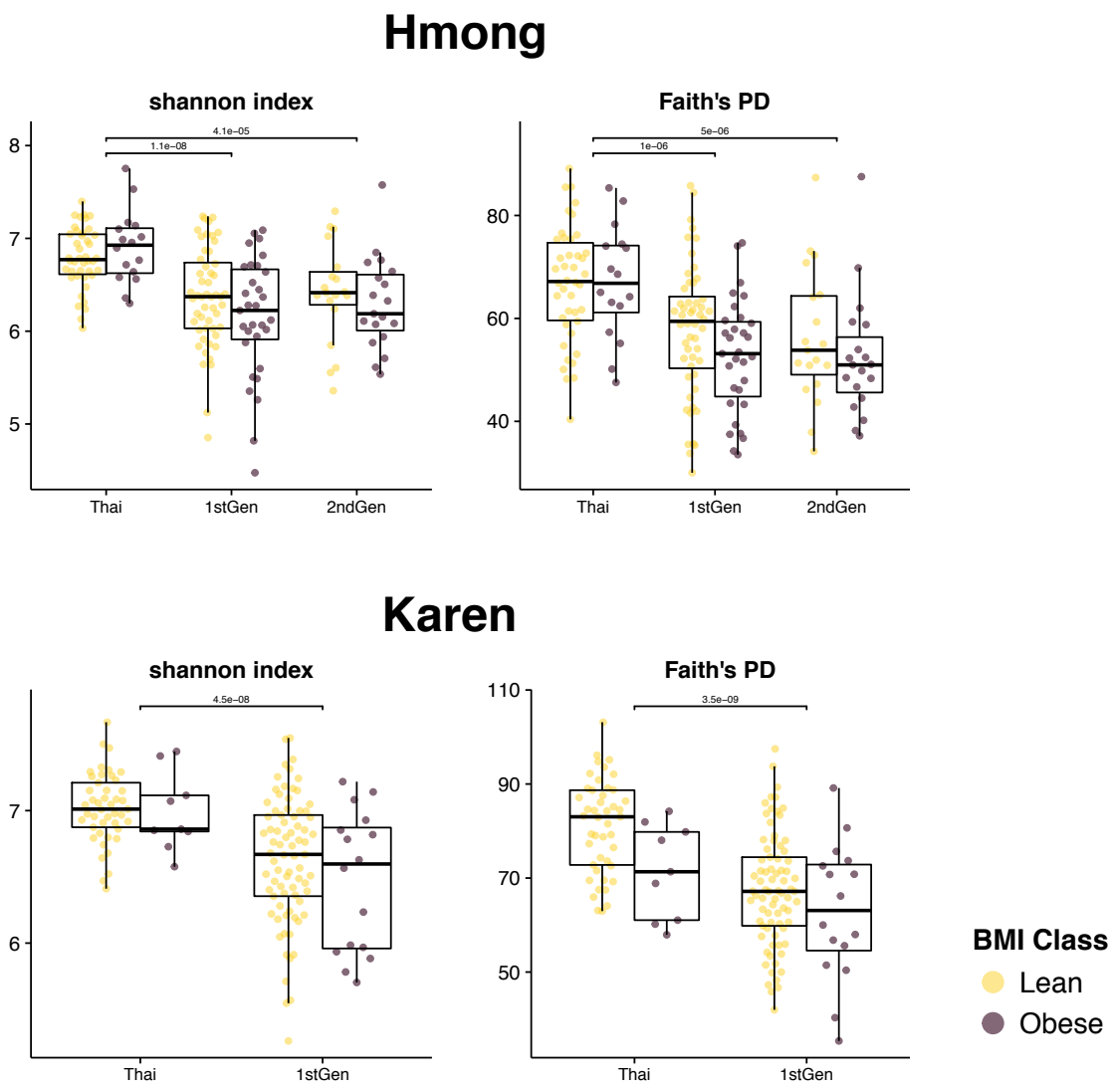

Figure S2. Alpha diversity boxplots of obese and lean individuals, separated by ethnicity. Related to Figure 2.
Post-hoc analysis with Tukey's HSD test across sample groups ($p < 0.01$).

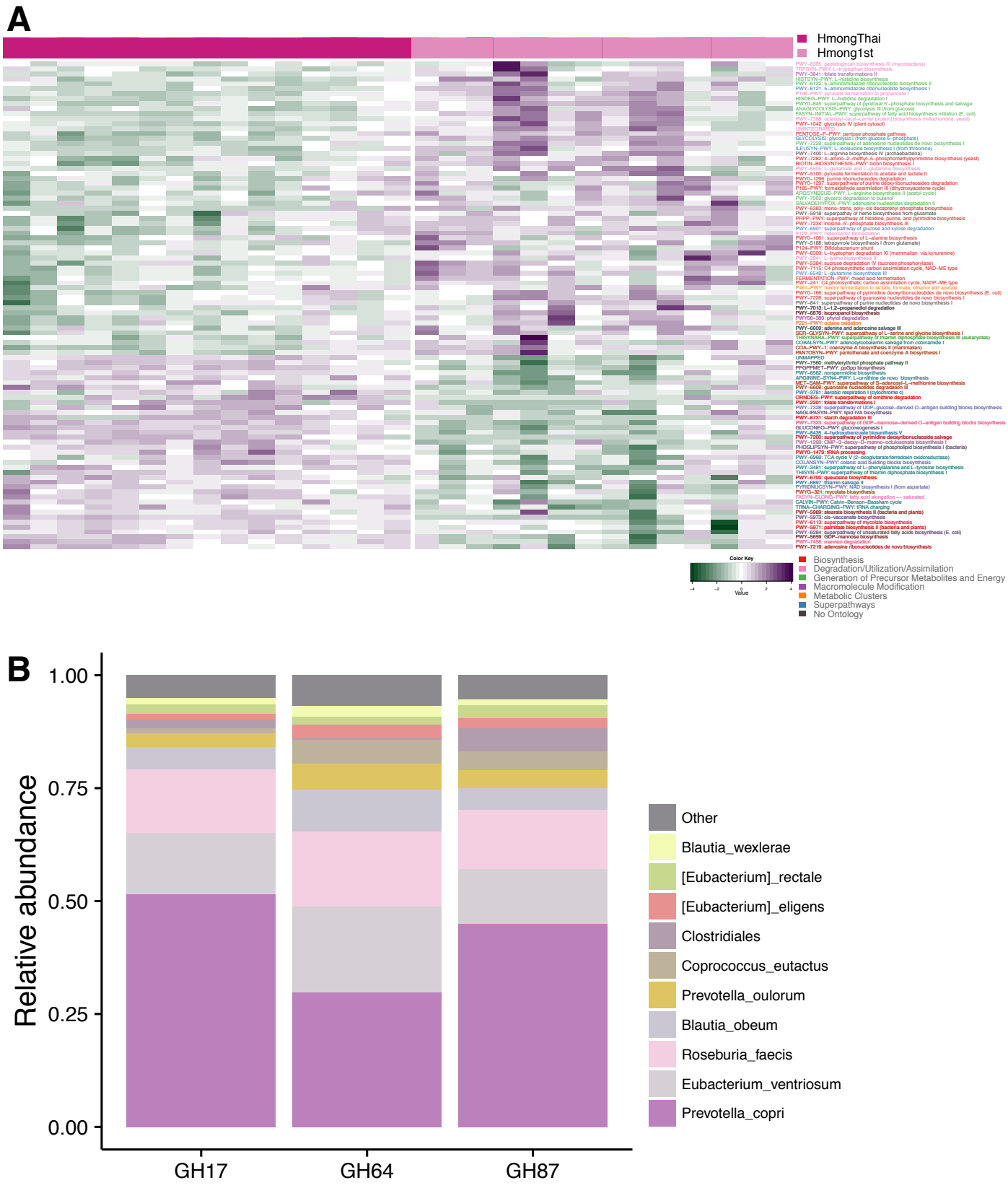

Figure S3. Functional annotations and glycoside hydrolase taxonomic contributions. Related to Figure 3.

(A) Differentiated relative abundances of functional pathways between HmongThai and Hmong1st (asin-sqrt transformed abundances, ANOVA, FDR-corrected q < 0.10).

(B) Taxonomic contributions of scaffolds contributing to beta-glucan-targeting glycoside hydrolases

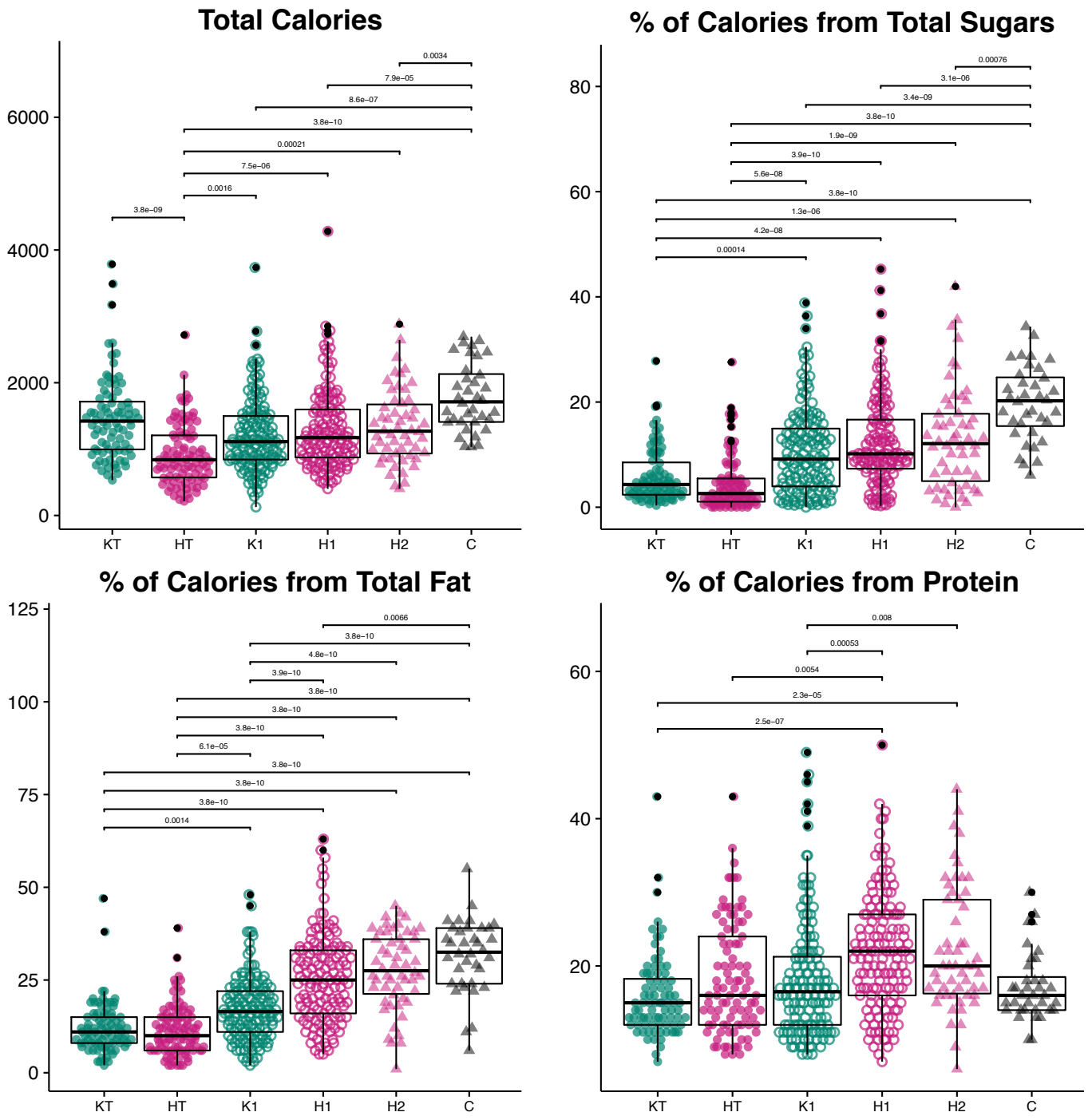

Figure S4. Macronutrient pairwise comparisons. Related to Figure 4.
Pairwise comparisons with Tukeys' HSD, significant p-values < 0.01 are shown.

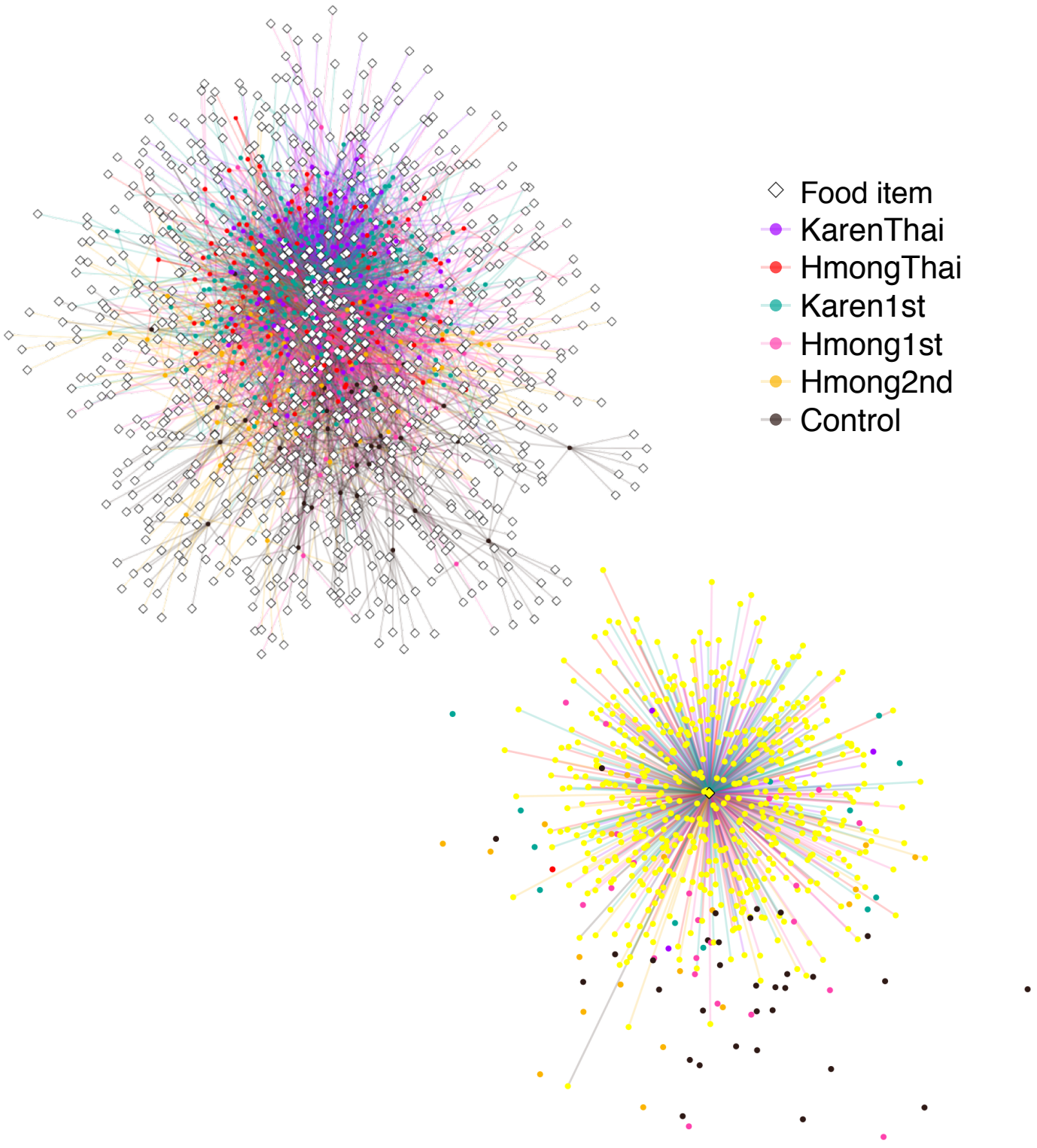

Figure S5. Bipartite network of participant dietary records and food items. Related to Figure 4.
(A) Edges and participants are colored by sample group, and food items are shown as white-filled diamonds.
(B) We highlight the high prevalence of rice consumption. Participants who consumed rice are denoted as yellow nodes and yellow edges connected to the centroid (rice), otherwise participants were colored by sample group.

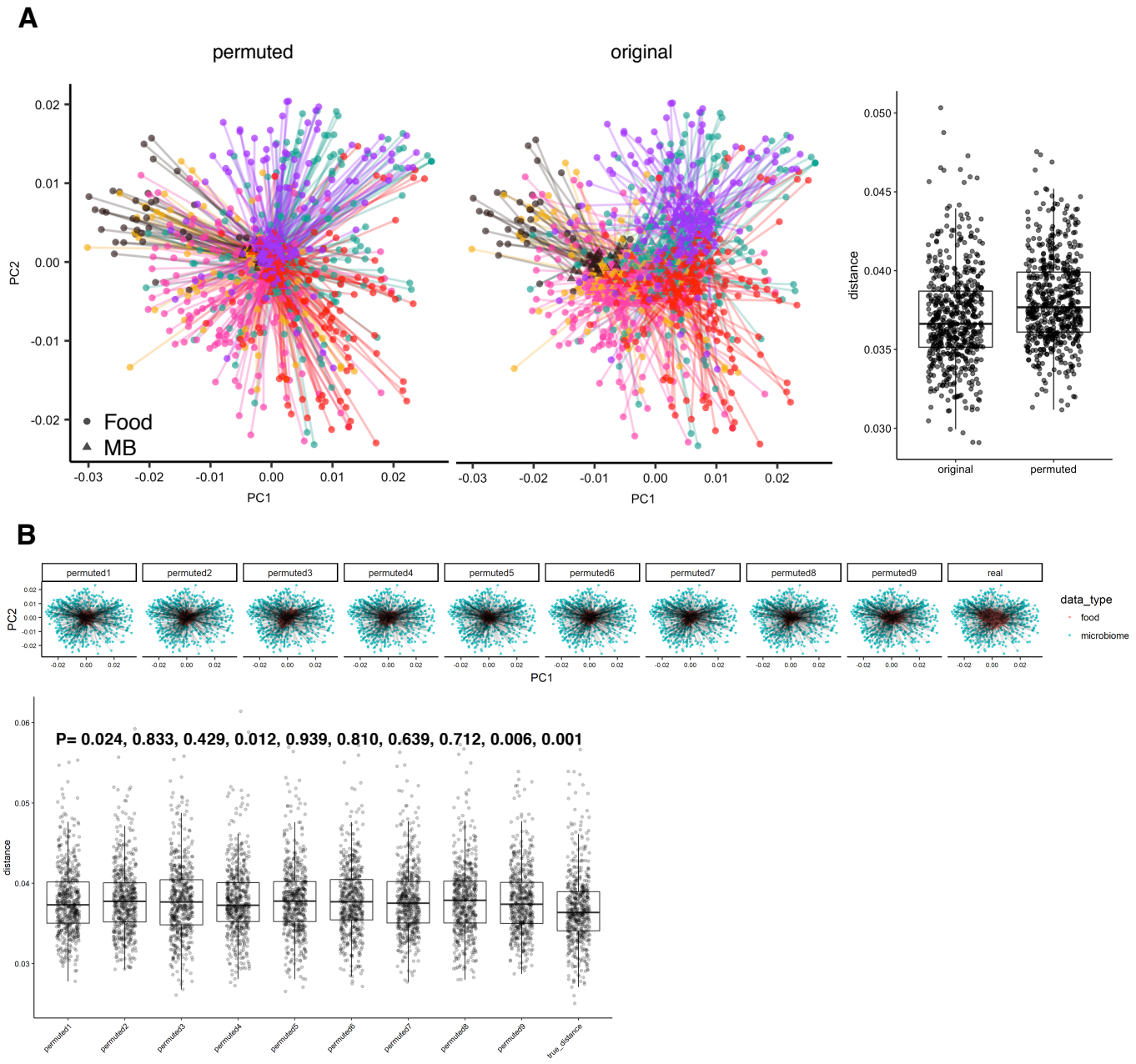

Figure S6. Procrustes of diet and microbiome distances. Related to Figure 4.
(A) Procrustes permutation shows significant relatedness between individuals' food and microbiome profiles. Shown at left is the Procrustes PCoA for a representative permutation (median Procrustes sum of squares distance from 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.

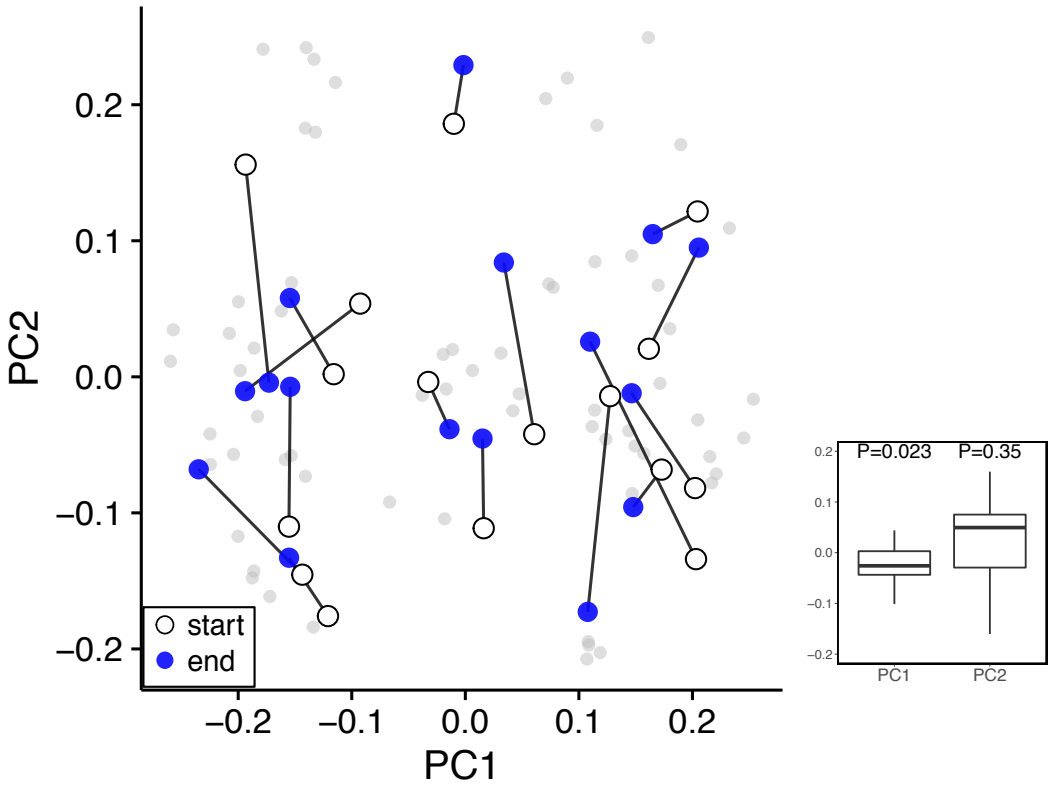

Figure S7. PCoA of unweighted Unifrac distances of longitudinal samples. Related to Figure 7.

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, PC1 P=0.023, PC2 P=0.35).

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 S1. Sample group recruitment stratified by BMI threshold of 25. Related to Figure 1.

	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 S2. Sample Group Characteristics.
Related to Figure 1. All values are represented as mean (min - max). HS = High School; ESL = English as a Second Language; < = less than.

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 S3. Manually curated foods. Related to Figure 1.

Table S4. OTU prevalences in HmongThai and Hmong1st. Related to Figure 2.

OTUID	qval	delta.prevalence	taxa
220	0.06841975	0.051094891	t__Prevotella copri DSM 18205
553	0.204775661	0.02919708	Enterobacteriaceae
899	0.11775423	0.03649635	Blautia faecis
921	0.038451473	0.058394161	Hungatella effluvii
1175	0.726474298	0.011492468	Romboutsia timonensis
58	0.50466954	-0.010638298	Faecalibacterium prausnitzii
75	0.32617301	0.033623233	Faecalibacterium prausnitzii
12	0.570002155	0.023232085	Gemmiger formicilis
1611	0.000246047	0.164900714	Clostridium
1812	3.05E-14	0.455694216	Faecalibacterium prausnitzii
267	0.275539384	0.038144573	t__Prevotella copri DSM 18205
394	0.658209736	-0.014049133	t__Bacteroides vulgatus ATCC 8482
543	0.075085901	0.067969547	Enterobacterales
818	0.404353527	0.030688329	Blautia luti
909	0.275539384	0.038144573	Dorea formicigenerans
936	0.570002155	0.023232085	Blautia
1276	0.018126819	0.10306252	Faecalibacterium prausnitzii
1667	0.812983527	0.012615043	[Eubacterium] hallii
1773	0.459930833	0.035226912	Eubacterium
1845	9.16E-06	0.223659156	Faecalibacterium prausnitzii
1905	2.89E-09	0.336718502	Faecalibacterium prausnitzii
455	0.001390193	0.148286258	t__Parabacteroides distasonis ATCC 8503
614	0.812983527	0.012615043	Butyricicoccus
63	1.43E-06	0.253808315	Faecalibacterium prausnitzii
71	0.006957309	0.118137099	Faecalibacterium prausnitzii
754	0.812983527	0.012615043	Anaerostipes hadrus
806	0.626494283	0.020152333	t__Blautia obeum ATCC 29174
822	0.028637438	0.09552523	Blautia obeum
1643	1.64E-05	0.230368172	Clostridiales
1890	9.39E-06	0.237988289	Faecalibacterium prausnitzii
20	0.000797591	0.169407235	Subdoligranulum variabile
832	0.802138015	-0.013475576	Fusicatenibacter saccharivorans
884	0.04001889	0.093206064	Lachnoclostridium
1200	0.073783973	0.090835361	Intestinibacter bartlettii
1552	0.684252293	0.02919708	Erysipelotrichaceae
1888	2.45E-08	0.337388483	Faecalibacterium prausnitzii
3761	0.002097959	0.160178427	Blautia
881	0.684252293	0.021492295	Lachnoclostridium
1458	0.247300404	0.065037316	Clostridiales
1453	0.000685619	0.189699008	Oscillospiraceae
3283	0.00109417	0.181907652	Blautia
427	0.648513941	-0.020667596	Bacteroides
576	0.335198919	0.057245961	t__Haemophilus parainfluenzae ATCC 33392
828	0.176888356	0.072828672	Fusicatenibacter
1809	5.48E-13	0.503566689	Faecalibacterium prausnitzii
1652	0.590313023	0.030773059	Lactobacillus rogosae
1728	0.48459024	-0.032266092	Roseburia faecis

1956	5.96E-08	0.353848706	Faecalibacterium
383	0.009969274	0.148971466	Alistipes shahii
43	1.19E-07	0.338088918	Faecalibacterium prausnitzii
929	0.023736724	0.125331785	Hungatella
1672	0.511123889	-0.036160752	[Eubacterium] hallii
1709	0.082337425	0.099337193	t__Roseburia intestinalis L1-82
1715	0.487467972	0.043543921	t__Roseburia hominis A2-183
1846	1.21E-06	0.31453981	Faecalibacterium prausnitzii
2541	1.39E-08	0.386274016	t__Prevotella copri DSM 18205
285	6.20E-05	0.250776072	t__Prevotella copri DSM 18205
42	1.21E-06	0.31453981	Faecalibacterium prausnitzii
4334	0.059562446	0.10730766	Blautia
738	0.208494353	-0.052101686	Actinomyces odontolyticus
856	0.000795675	0.202953268	Clostridiales
886	0.005382086	0.171071399	Ruminococcus
895	0.217852853	0.075425791	Lachnoclostridium
953	0.698346136	-0.020219817	Bacteroides xylanolyticus
1283	0.030388645	-0.080461721	Streptococcus
1288	0.37762301	-0.048209133	Streptococcus
1752	0.022889138	0.1453064	t__Ruminococcus faecis JCM 15917
930	0.127890723	0.096927517	Hungatella
1277	0.140589679	0.094461142	Faecalibacterium prausnitzii
3910	0.255039316	0.078145127	Blautia
4372	0.000797591	0.224989266	Faecalibacterium prausnitzii
743	0.05430795	0.127093173	Tyzzerella
771	0.332915648	0.069987119	Lachnoclostridium
1045	0.650958567	0.034150156	Collinsella aerofaciens
1084	0.544727028	0.042405283	Collinsella aerofaciens
1595	0.030542582	0.149721933	Phascolarctobacterium succinatutens
1463	0.153752817	0.100191171	Oscillibacter
1786	0.030757767	0.141466806	Eubacterium
2346	4.46E-08	0.397375739	t__Prevotella copri DSM 18205
3924	0.269832396	0.075425791	Blautia
435	0.204208954	-0.064911366	t__Bacteroides xylanisolvens XB1A
534	0.154181346	0.091936044	Desulfovibrio
9	1.71E-09	0.4469065	Gemmiger formicilis
928	0.082337425	0.116701425	Hespellia
1863	0.003114933	0.206314308	Faecalibacterium prausnitzii
1986	9.06E-09	0.431888136	Faecalibacterium prausnitzii
2571	5.00E-10	0.473661068	t__Prevotella copri DSM 18205
664	0.453218047	0.055931756	Acutalibacter
954	0.554066713	0.04757717	Bacteroides xylanolyticus
962	0.914702065	0.014158825	Coprococcus catus
1252	0.000137479	0.281199929	Gemmiger formicilis
1452	8.03E-05	0.2896564	Oscillibacter
1523	0.008441475	0.188178743	Holdemanella biformis
23	0.626494283	-0.040145985	Faecalibacterium
1922	2.84E-06	0.3488517	Faecalibacterium prausnitzii

1957	8.55E-07	0.374221114	Faecalibacterium
1971	0.018575557	0.1712658	Faecalibacterium
2569	6.34E-11	0.509524657	t__Prevotella copri DSM 18205
411	0.626494283	-0.040145985	Bacteroides uniformis
4326	0.008441475	0.188178743	t__[Eubacterium rectale] ATCC 33656
940	0.376747825	0.069788143	Blautia
95	8.03E-05	0.2896564	Faecalibacterium prausnitzii
1660	0.041650425	0.152473642	t__[Eubacterium] eligens ATCC 27750
1725	0.006453762	0.203838875	t__[Eubacterium rectale] ATCC 33656
1891	0.00016619	0.280886726	Faecalibacterium prausnitzii
2337	8.84E-08	0.417860683	t__Prevotella copri DSM 18205
283	1.26E-08	0.452104172	t__Prevotella copri DSM 18205
3963	0.388583748	0.066864918	Blautia
4830	3.38E-07	0.392178066	[Eubacterium] hallii
4880	3.29E-05	0.315130215	t__[Eubacterium rectale] ATCC 33656
503	0.186205838	0.101108408	t__Parabacteroides merdae ATCC 43184
798	0.057658149	0.143912769	Ruminococcus faecis
905	0.104789797	0.118230152	Clostridiales
1206	2.04E-07	0.410583942	Terrisporobacter petrolearius
1402	0.323706804	0.08120438	Veillonella
1688	0.033187762	0.159215328	t__Senegalimassilia anaerobia JC110
1939	1.19E-07	0.419251825	Faecalibacterium prausnitzii
1951	1.19E-07	0.419251825	Faecalibacterium prausnitzii
1984	7.01E-14	0.618613139	Faecalibacterium prausnitzii
2325	3.68E-11	0.54060219	t__Prevotella copri DSM 18205
255	0.082337425	0.133211679	t__Prevotella copri DSM 18205
3282	7.13E-11	0.531934307	Dorea longicatena
3760	4.38E-06	0.358576642	Dorea longicatena
4134	0.00735803	0.202554745	Collinsella aerofaciens
4786	1.29E-05	0.341240876	Lactobacillus rogosae
773	0.659451185	-0.031478102	Lachnoclostridium
918	0.684252293	0.037864964	Blautia obeum
1033	0.266437023	0.087129262	Collinsella aerofaciens
1086	0.553356	-0.044534787	Collinsella aerofaciens
1217	0.012114043	0.201238104	Sutterella
1442	0.000588102	0.27145893	Clostridium
2031	0.000138258	0.306569343	Faecalibacterium prausnitzii
208	1.64E-06	0.385567772	t__Prevotella copri DSM 18205
2106	9.05E-06	0.359234963	Faecalibacterium prausnitzii
248	1.83E-08	0.464566202	t__Prevotella copri DSM 18205
2513	1.22E-07	0.429455789	t__Prevotella copri DSM 18205
2545	1.08E-09	0.508454218	t__Prevotella copri DSM 18205
2788	3.05E-14	0.640118267	t__Prevotella copri DSM 18205
3285	0.00014631	0.29779174	Collinsella aerofaciens
3773	0.00014631	0.29779174	Faecalibacterium prausnitzii
3929	0.048754658	0.157350088	Blautia
3959	0.000907196	0.262681327	Blautia
4112	1.22E-07	0.429455789	Collinsella aerofaciens

809	0.005530239	0.218793311	[Clostridium] glycyrrhizinilyticum
917	0.117890639	0.122239675	Clostridiales
1078	0.424233161	0.066535654	Collinsella aerofaciens
1085	0.926385951	-0.013475576	Collinsella aerofaciens
1849	3.35E-12	0.591053715	Faecalibacterium prausnitzii
227	2.16E-06	0.386580573	t__Prevotella copri DSM 18205
2349	1.21E-16	0.706625491	t__Prevotella copri DSM 18205
2462	4.25E-08	0.457701666	t__Prevotella copri DSM 18205
302	4.25E-08	0.457701666	t__Prevotella copri DSM 18205
400	0.447849025	-0.057926259	Bacteroides dorei
437	0.557425463	-0.049036122	Bacteroides ovatus
4371	3.16E-10	0.528822759	Faecalibacterium prausnitzii
69	0.000455341	0.279898933	Faecalibacterium prausnitzii
938	0.926385951	0.013194834	Blautia
1332	0.099676192	0.135463077	Flintibacter
1466	0.0559278	0.162479856	Oscillibacter
1813	1.28E-06	0.405630865	Faecalibacterium prausnitzii
1909	5.12E-11	0.567731539	Faecalibacterium prausnitzii
1987	9.78E-08	0.45065883	Faecalibacterium prausnitzii
202	2.46E-08	0.477675609	t__Prevotella copri DSM 18205
2121	0.00017671	0.315574936	Faecalibacterium prausnitzii
2345	7.08E-14	0.648781875	t__Prevotella copri DSM 18205
2348	1.11E-15	0.69380984	t__Prevotella copri DSM 18205
2459	1.20E-05	0.369608494	t__Prevotella copri DSM 18205
2810	1.23E-17	0.738837805	t__Prevotella copri DSM 18205
4060	6.86E-09	0.495686795	Clostridiales
450	0.006696738	0.225519007	t__Parabacteroides distasonis ATCC 8503
4496	2.01E-12	0.60375391	Clostridiales
524	1.30E-05	0.360602901	t__Prevotella copri DSM 18205
1178	0.005006862	0.24270073	Romboutsia timonensis
1072	0.361018784	0.087591241	Collinsella aerofaciens
1461	0.103879911	0.142335766	Oscillospiraceae
1616	0.059194428	0.160583942	Clostridium
1767	0.926385951	-0.012773723	Blautia
1870	5.94E-11	0.571167883	Faecalibacterium prausnitzii
1970	0.000568481	0.288321168	Faecalibacterium
201	8.87E-09	0.498175182	t__Prevotella copri DSM 18205
221	4.72E-06	0.388686131	t__Prevotella copri DSM 18205
2354	2.81E-06	0.397810219	t__Prevotella copri DSM 18205
2465	1.88E-07	0.452554745	t__Prevotella copri DSM 18205
2653	2.81E-06	0.397810219	t__Prevotella copri DSM 18205
2705	3.19E-11	0.580291971	t__Prevotella copri DSM 18205
3840	0.078559594	0.151459854	Dorea longicatena
3876	0.043997413	0.169708029	Dorea formicigenerans
4115	0.078559594	0.151459854	Collinsella aerofaciens
666	0.926385951	0.01459854	Ruminococcus bromii
758	0.710303703	0.041970803	Ihubacter
1076	0.817920367	0.02919708	Collinsella aerofaciens

1146	0.368810626	0.084671533	Thermoactinomycetaceae
1548	0.062892914	0.15863747	Turicibacter sanguinis
1597	2.81E-06	0.408272506	Phascolarctobacterium succinatutens
1766	3.02E-11	0.593187348	Coprococcus
1819	8.53E-09	0.509975669	Gemmiger formicilis
1835	4.88E-05	0.352798054	Faecalibacterium prausnitzii
1883	4.88E-06	0.399026764	Faecalibacterium prausnitzii
1999	0.082337425	0.149391727	Faecalibacterium prausnitzii
2071	2.81E-06	0.408272506	Faecalibacterium prausnitzii
2338	1.71E-09	0.537712895	t__Prevotella copri DSM 18205
4146	0.447849025	0.075425791	Collinsella aerofaciens
432	0.812983527	-0.026277372	Bacteroides fragilis
4328	0.001438243	0.278832117	t__Ruminococcus faecis JCM 15917
4911	0.000138258	0.334306569	Faecalibacterium prausnitzii
52	1.19E-07	0.463746959	Faecalibacterium prausnitzii
5389	3.10E-08	0.491484185	Faecalibacterium prausnitzii
848	0.045994213	0.177128954	Dorea longicatena
949	0.062892914	0.15863747	Hungatella
1049	0.818574175	0.025448806	Collinsella aerofaciens
1145	0.005995551	0.240974551	Thermoactinomycetaceae
1123	0.000153482	0.334681397	Blautia faecis
1262	0.005995551	0.240974551	Parasutterella excrementihominis
1771	0.048754658	0.175379759	t__[Eubacterium rectale] ATCC 33656
1848	3.02E-05	0.372164135	Faecalibacterium prausnitzii
1916	9.99E-06	0.390905504	Faecalibacterium prausnitzii
1928	6.64E-08	0.48461235	Faecalibacterium prausnitzii
1950	0.000246047	0.325310712	Faecalibacterium prausnitzii
2009	0.018126819	0.212862498	Faecalibacterium prausnitzii
3554	3.02E-05	0.372164135	t__[Eubacterium rectale] ATCC 33656
3950	0.112858378	0.147267706	Blautia
4136	0.000246047	0.325310712	Collinsella aerofaciens
4151	0.036094188	0.184750444	Collinsella aerofaciens
4327	9.99E-06	0.390905504	Roseburia faecis
4439	2.93E-06	0.419017558	Faecalibacterium prausnitzii
6447	6.64E-08	0.48461235	Subdoligranulum variabile
14	0.000455341	0.316068393	Gemmiger formicilis
1467	0.000703832	0.306569343	Oscillibacter
1473	0.001809351	0.278072193	Sporobacter
2081	9.84E-07	0.449055094	Faecalibacterium prausnitzii
219	5.02E-07	0.458554145	t__Prevotella copri DSM 18205
2317	2.48E-09	0.544045595	Blautia
2406	1.61E-12	0.648535146	t__Prevotella copri DSM 18205
3626	3.34E-06	0.420557944	Lactobacillus rogosae
3820	0.459004763	0.078592141	Fusicatenibacter saccharivorans
3884	0.027234594	0.202079792	Hungatella effluvii
4155	0.63539325	0.050094991	Collinsella aerofaciens
4515	1.83E-08	0.515548445	Lachnospiraceae
744	0.000163252	0.344565543	Tyzzerella

783	0.089230338	0.154584542	t__[Eubacterium rectale] ATCC 33656
1486	0.014153156	0.229521492	Sporobacter
1741	0.728790445	0.036901865	t__[Eubacterium rectale] ATCC 33656
1958	3.13E-07	0.470296026	Faecalibacterium
224	7.53E-08	0.49918897	t__Prevotella copri DSM 18205
2441	4.20E-08	0.508819951	t__Prevotella copri DSM 18205
2497	3.13E-07	0.470296026	t__Prevotella copri DSM 18205
2542	1.69E-12	0.653284672	t__Prevotella copri DSM 18205
2543	3.33E-11	0.614760746	t__Prevotella copri DSM 18205
2628	1.34E-07	0.489557989	t__Prevotella copri DSM 18205
2656	4.40E-09	0.547343877	t__Prevotella copri DSM 18205
3307	2.16E-09	0.556974858	t__Prevotella copri DSM 18205
3631	2.05E-05	0.393248175	[Eubacterium] hallii
3790	0.000784924	0.306569343	t__Blautia obeum ATCC 29174
3897	6.29E-05	0.373986212	Blautia
3905	0.020090775	0.219890511	Blautia
4107	0.000784924	0.306569343	Collinsella aerofaciens
4148	0.020090775	0.219890511	Collinsella aerofaciens
423	0.817920367	-0.030515004	t__Bacteroides stercoris ATCC 43183
4324	0.000515695	0.316200324	t__Roseburia inulinivorans DSM 16841
4337	0.000474959	0.325831306	Anaerosporobacter
772	0.014153156	0.229521492	Lachnoclostridium
857	0.000290327	0.335462287	Clostridiales
984	1.98E-06	0.441403082	Prevotella copri
985	7.00E-06	0.412510138	Prevotella

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 S5. 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 S6. Foods (summarized at level 3) that are significantly correlated with PC1 of the diet-based unweighted Unifrac PCOA (Spearman's correlation, FDR-corrected q < 0.01). Related to Figure 4.