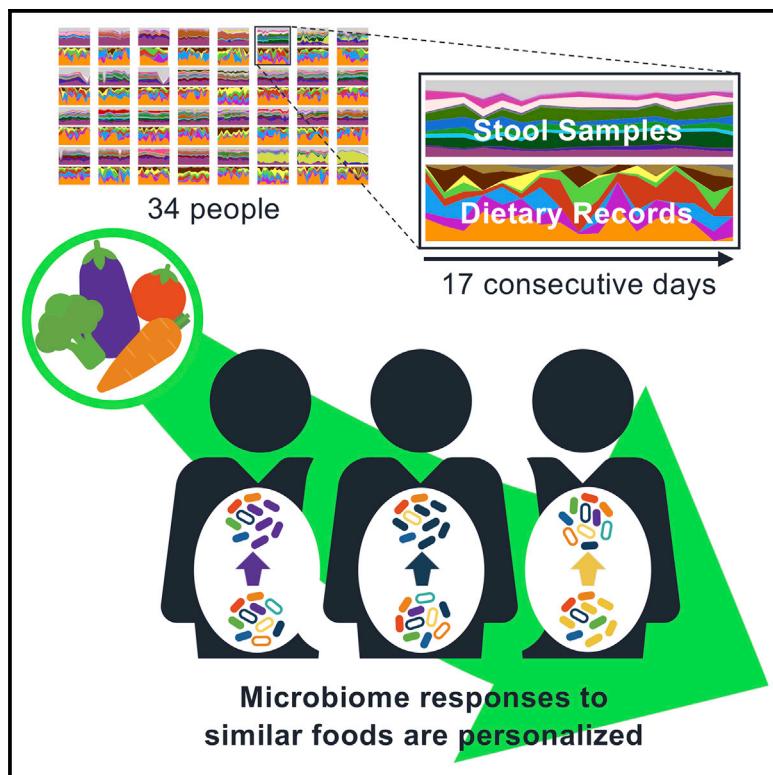


Cell Host & Microbe

Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans

Graphical Abstract



Authors

Abigail J. Johnson, Pajau Vangay,
Gabriel A. Al-Ghalith, ..., Ravi Menon,
Katie Koecher, Dan Knights

Correspondence

dknights@umn.edu

In Brief

Dietary intake is often considered to be a driver of microbiome variation. Johnson et al. use longitudinal sampling and daily dietary records to model microbiome changes in response to diet and find that microbiome responses to diet are personalized.

Highlights

- Daily microbiome variation is related to food choices, but not to conventional nutrients
- Daily microbiome variation depends on at least two days of dietary history
- Similar foods have different effects on different people's microbiomes



Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans

Abigail J. Johnson,¹ Pajau Vangay,² Gabriel A. Al-Ghalith,² Benjamin M. Hillmann,³ Tonya L. Ward,⁴ Robin R. Shields-Cutler,¹ Austin D. Kim,⁵ Anna Konstantinovna Shmagel,⁶ Arzang N. Syed,⁷

Personalized Microbiome Class Students,⁸ Jens Walter,⁹ Ravi Menon,¹⁰ Katie Koecher,¹⁰ and Dan Knights^{1,3,11,*}

¹BioTechnology Institute, University of Minnesota, Saint Paul, MN 55108, USA

²Bioinformatics and Computational Biology, University of Minnesota, Minneapolis, MN 55455, USA

³Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA

⁴CoreBiome Inc, St. Paul, MN 55114, USA

⁵Department of Mathematics, Statistics, and Computer Science, Macalester College, St. Paul, MN 55105, USA

⁶Division of Rheumatic and Autoimmune Diseases, Department of Medicine, University of Minnesota, Minneapolis, MN 55455, USA

⁷Department of Genetics, Cell Biology, and Development, University of Minnesota, Minneapolis, MN 55455, USA

⁸Microbial Engineering Program, Biotechnology Institute, University of Minnesota, Minneapolis, MN 55455, USA

⁹Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta, Canada

¹⁰Bell Institute of Health & Nutrition, General Mills Inc, Minneapolis, MN 55427, USA

¹¹Lead Contact

*Correspondence: dknight@umn.edu

<https://doi.org/10.1016/j.chom.2019.05.005>

SUMMARY

Diet is a key determinant of human gut microbiome variation. However, the fine-scale relationships between daily food choices and human gut microbiome composition remain unexplored. Here, we used multivariate methods to integrate 24-h food records and fecal shotgun metagenomes from 34 healthy human subjects collected daily over 17 days. Microbiome composition depended on multiple days of dietary history and was more strongly associated with food choices than with conventional nutrient profiles, and daily microbial responses to diet were highly personalized. Data from two subjects consuming only meal replacement beverages suggest that a monotonous diet does not induce microbiome stability in humans, and instead, overall dietary diversity associates with microbiome stability. Our work provides key methodological insights for future diet-microbiome studies and suggests that food-based interventions seeking to modulate the gut microbiota may need to be tailored to the individual microbiome. Trial Registration: ClinicalTrials.gov: NCT03610477.

INTRODUCTION

The microbial ecosystem within the human gastrointestinal tract is dynamic and complex, and its composition is known to vary widely across healthy individuals (Human Microbiome Project Consortium, 2012). When measured within the same individual over a longitudinal period, large shifts in microbial composition can take place in response to disease or environmental changes

(David et al., 2014a; Flores et al., 2014; Fukuyama et al., 2017).

Substantial changes in microbiome composition have been measured in response to changes in dietary intake, such as those seen with a shift from plant-based to animal-based diets (David et al., 2014b) or those seen after the addition of individual nutrients (Maier et al., 2017). However, controlled feeding trials have revealed that inter-subject microbiome variation remains high even after periods of identical dietary intake (Wu et al., 2011). In population-level studies, diet consistently accounts for only a small proportion of microbiome variation (Falony et al., 2016; Rothschild et al., 2018; Vangay et al., 2018), and only modest differences have been found between groups of people consuming vastly different dietary patterns, such as omnivores and vegetarians (Wu et al., 2016). Recently, microbiome composition has been found to predict biomarkers of blood glucose control (Korem et al., 2017; Zeevi et al., 2015) and weight loss (Thaiss et al., 2016) in a personalized way. While it is understood that diet broadly affects microbiome composition within an individual (Gentile and Weir, 2018), an understanding of the exact importance of specific foods and nutrients in shaping microbiome composition across populations remains elusive.

The lack of robust methods to assess and pair complex food intake from free-living populations with existing microbiome analysis pipelines is a limitation of current efforts to understand how dietary intake affects microbiome composition. Studies assessing direct diet-microbiome relationships have largely relied on food frequency questionnaires (FFQs) and conventional nutrient profiles from macro- and micronutrients (Wu et al., 2011; Zhernakova et al., 2016). Studies of broad diet-microbiome relationships have used principal-component analysis (Strate et al., 2017) and correspondence analysis (Claesson et al., 2012) to define dietary patterns (i.e., “prudent” versus “western”) or dietary indices based on *a priori* knowledge (i.e., the healthy eating index [HEI]) (Bowyer et al., 2018). While dietary patterns can be linked to differences in microbiome composition across individuals (Claesson et al., 2012; Wu et al., 2011),



Table 1. Subject Characteristics by Gender

Characteristic	Overall	Female (n = 20)	Male (n = 14)	p value (female versus male)
Age (years)	31 ± 10	29 ± 10	34 ± 11	0.2
Weight (kg)	69 ± 15	60 ± 8	81 ± 13	<0.001
Height (cm)	172 ± 9	165 ± 5	180 ± 8	<0.001
Waist circumference (cm)	87 ± 8	83 ± 6	91 ± 8	0.006
Cholesterol (mg/dL)	164 ± 32	169 ± 34	158 ± 30	0.6
Trigs (mg/dL)	66 ± 20	67 ± 19	65 ± 22	0.1
HDL (mg/dL)	57 ± 11	59 ± 12	55 ± 9	0.02
LDL (mg/dL)	94 ± 29	97 ± 31	90 ± 28	0.9
Glucose (mg/dL)	86 ± 7	85 ± 8	87 ± 6	0.4
Insulin (mU/L)	8 ± 4	8 ± 2	9 ± 4	0.5

Values shown as mean ± standard deviation. p values calculated using Student's t test. Abbreviations: Trigs, triglycerides; HDL, high-density lipoprotein; LDL, low-density lipoprotein.

longitudinal pairing of diet and microbiome data is needed to assess how temporal variation in diet alters short-term microbiome stability, composition, and function. Similarly, diet assessment methods that rely solely on conventional nutrient profiles (i.e., macro- and micronutrients from nutrient databases) from 24-h dietary recalls or self-reported food records overlook information about foods that could have important microbial-dependent influences on health (Barratt et al., 2017), such as the presence of phenolic compounds (Thaiss et al., 2016), unique conformations of polysaccharides or microbiota-accessible carbohydrates (MACs) (Smits et al., 2016; Sonnenburg et al., 2016), and plant-derived exosomal microRNAs that persist through the gut and impact microbial metabolism and composition (Teng et al., 2018).

To investigate the fine-scale relationships between daily food choices and human gut microbiome composition, we conducted an ultra-dense longitudinal study of the impact of habitual diet on the microbiome, including daily shotgun fecal metagenomic microbiome sequencing for 17 days from 34 subjects, paired with complete daily 24-h dietary records. We developed and applied multivariate methods for modeling dietary intake that move beyond conventional nutrient-based analysis. Combining daily shotgun metagenomics with daily diet provided a uniquely rich dataset for measuring the effects of diet on the personalization of microbial dynamics. Our dense longitudinal dataset also allowed us to investigate relationships between dietary intake and temporal microbial stability (Zaneveld et al., 2017).

RESULTS

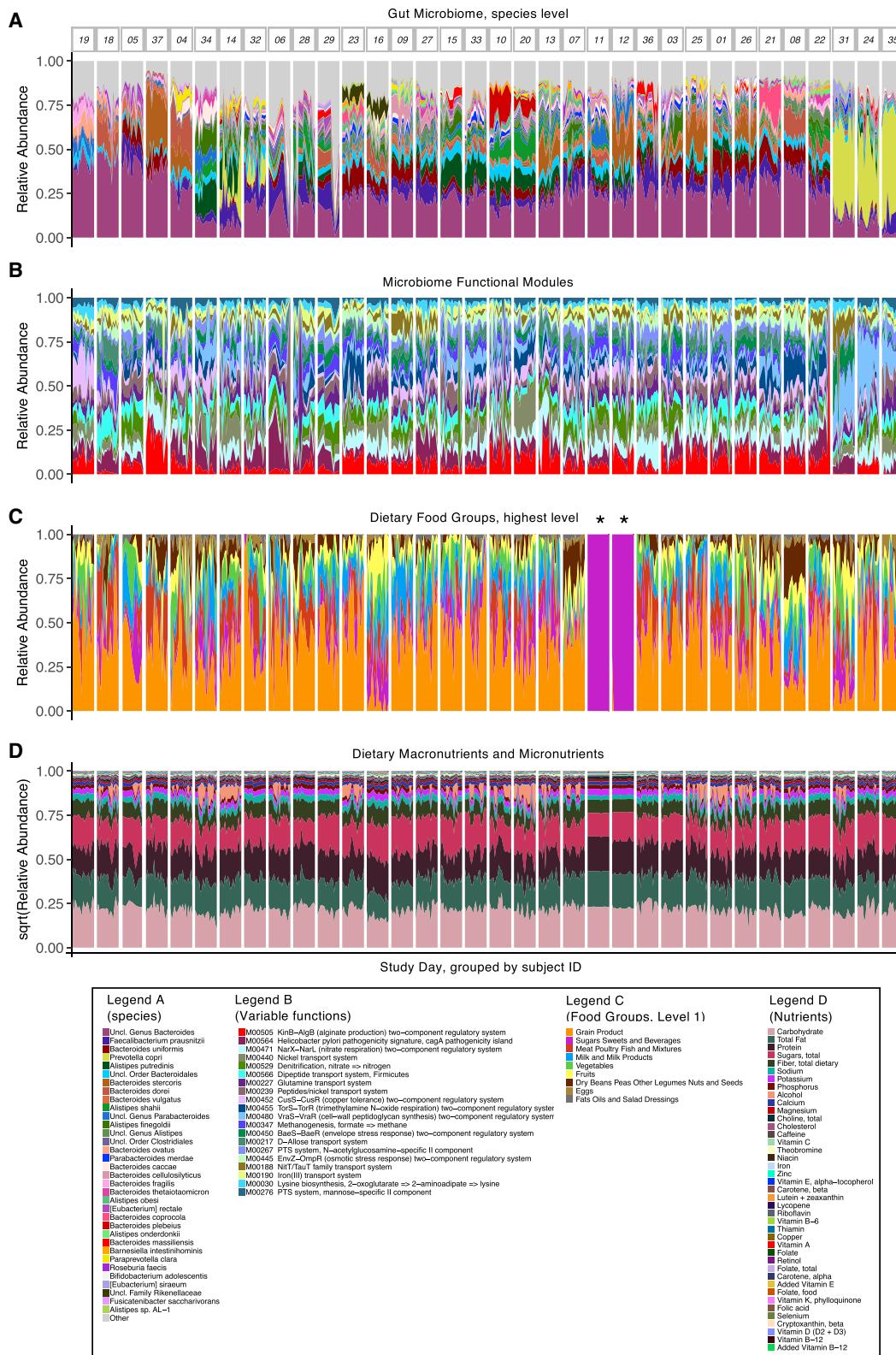
Daily Dietary Intake and Microbiome Composition Are Highly Variable and Personalized

Our study was designed to identify the relationship between habitual dietary intake and daily microbiome variation using dense, longitudinal diet-microbiome data. To characterize the longitudinal relationship between diet and microbiome composition, we collected dietary intake data and fecal samples from 34

subjects for 17 consecutive days (see Table 1 for cohort characteristics by gender and STAR Methods for detailed subject information). Female subjects had significantly lower weight, height, waist circumference, and high-density lipoprotein than did male subjects (Student's t test, p value < 0.05). Daily food records were collected using the automated self-administered 24-h (ASA24) dietary assessment tool (2016, National Cancer Institute, Bethesda, MD, USA) (Park et al., 2018). Interestingly, the reported diet of two subjects (11 and 12) consisted primarily of a nutritional meal replacement beverage in 4 different flavors (Soylent, Rosa Foods). We refer to these subjects as "shake drinkers." The study also included a 10-day, parallel, double-blind intervention trial to test the impact of medium chain triglycerides (MCTs) compared to long chain dietary triglycerides from extra virgin olive oil (EVOO) on microbiome composition (STAR Methods; see Table S1 for characteristics of study subjects by dietary supplementation arm). As there were only null findings from all tests for associations between MCT or EVOO supplementation and the microbiome (Figure S1), this manuscript focuses on the analysis of overall diet-microbiome covariation using all available samples.

We conducted shotgun metagenomic sequencing on each stool sample at a depth of $7,195,302 \pm 2,442,901$ single-end reads per subject, divided approximately evenly across the time points for each subject, with an average depth of $506,133 \pm 323,896$ reads per sample after the removal of human DNA. We have shown previously that low-depth metagenomic sequencing can recover species-level taxonomic assignments and also allows for the assessment of functional profiles (Hillmann et al., 2018). We removed sequencing adaptors, trimmed and filtered shotgun metagenomic reads according to quality using SHI7 (Al-Ghalith et al., 2018), and assigned taxonomy using BURST (Al-Ghalith and Knights, 2017) and a database consisting of all bacterial strains annotated at chromosome-level assembly or better in RefSeq version 86. We retained 483 microbiome samples for analysis after removal of those with low depth. Dietary outliers were identified according to guidelines from ASA24 by comparing macronutrient composition and total energy intake to reference levels to identify low quality reporting. We retained 566 24-h food records after removal of dietary outliers (see Figure S1 for data availability by study day). While the shake drinkers were outliers in terms of the number and types of foods consumed, they were not outliers in terms of nutritional composition (Figure 1D). These subjects were retained in the dataset for a unique subset analysis of dietary stability but were excluded from analyses of habitual dietary intake and microbiome composition.

Microbiome composition was more variable in some subjects than others across the study period (Figure 1A). As has previously been shown (Human Microbiome Project Consortium, 2012), the most prevalent microbial functional modules, as annotated by the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000), were highly consistent within and across subjects (Figure S2); however, we found a subset of functions to be highly variable (high coefficient of variation) across all samples (Figure 1B). The three most variable functions were the BaeS-BaeR (envelope stress response) two-component regulatory system module (M00450), the denitrification module (M00529) responsible

**Figure 1. Daily Microbiome Variation and Dietary Intake Are Highly Individualized**

(A) Area charts showing the relative abundance of species-level taxa in the dataset for each subject, by day, over the study time course. Low-abundance taxa are grouped together as “other.” Subjects are ordered to show microbiome similarity (see Figure S1 for clustering dendrogram). Species from related taxonomy

(legend continued on next page)

for the conversion of nitrate to nitrogen, and the methanogenesis module (M00347) responsible for the conversion of formate to methane. When comparing average functional modules with average nutrient and food intake, we found no significant correlations between the most highly variable functional modules and dietary features in our samples. We hypothesize that these metagenomic functions may be highly variable because of personalized induction of different bacteria in response to dietary compounds that are not included in our nutritional database. Alternatively, functional variation may be due to stochastic fluctuations in bacterial abundance, syntrophic, or predator-prey relationships within the microbiota that are difficult to predict.

Recovery of Dietary Diversity from Individual Food Choices

While dietary intake (in terms of food choices) was highly personalized and variable between and within subjects ([Figure 1C](#)), macro- and micronutrient profiles were relatively stable across the study period ([Figure 1D](#)). Therefore, to incorporate the maximum amount of observed dietary complexity into our analysis, we chose to study diet in terms of reported food choices. While a small number of food choices were shared by more than 20 subjects (i.e., banana, coffee, cheddar cheese, lettuce, carrots, chicken breast), many foods were consumed by only one subject over the 17-day study period ([Figure 2A](#)). This is to be expected with dietary data because, like microbiome data, food intake data are often zero-inflated with highly correlated features ([Zhang et al., 2011](#)), meaning that people episodically eat unique foods and they often consume meals within conventional food pairings. To account for the highly individualized nature of food choices, we borrowed from concepts previously developed for ecological analysis of the microbiome ([Lozupone and Knight, 2005](#)). We created a phenetic, hierarchical tree of foods from the Food and Nutrient Database for Dietary Studies (FNDDS) ([U.S. Department of Agriculture and Agricultural Research Service, 2014](#)) ([Figure 2B; STAR Methods](#)), which allowed us to apply the tree-based beta-diversity metric, UniFrac ([Lozupone and Knight, 2005](#)), and the tree-based alpha-diversity metric, Faith's phylogenetic diversity ([Faith, 1992](#)), to dietary data. This approach shares statistical information across related foods, enabling us to measure the tree-based similarity between each pair of diet profiles, the overall tree-based diversity of total foods consumed by a person, and the tree-based diversity of food sources for a particular nutrient such as fiber.

Food Choices Associate with Overall Microbiome Composition between Subjects

Microbiome beta-diversity analysis using Aitchison's distances showed strong grouping by subject ([Figure 2C; PERMANOVA; p value = 0.001; 999 permutations](#)) ([Aitchison et al., 2000](#)). Aitchison's distance was selected for the microbiome beta-diversity metric to account for compositionality of relative abundance profiles. Average microbiome beta-diversity (Aitchison's distances) did not show grouping by gender (PERMANOVA; p value = 0.2; 999 permutations). Using QIIME 1.9.1 ([Caporaso et al., 2010](#)) and our hierarchical tree of foods, we calculated unweighted UniFrac beta-diversity of food profiles. Food beta-diversity also showed significant within-subject grouping ([Figure 2D; PERMANOVA; p value = 0.001; 999 permutations](#)). This within-subject grouping is most obvious for subject 8 (lower right quadrant) and subject 28 (lower left quadrant) relative to the rest of the study cohort. Average food beta-diversity (unweighted UniFrac) did show grouping by gender (PERMANOVA; p value = 0.002; 999 permutations).

We applied Procrustes analysis to test for microbiome variation and dietary variation across subjects. Our analysis showed that a subject's average food intake corresponds with that subject's average microbiome composition when analyzed using the unweighted UniFrac-based food distances ([Figure 2E; Procrustes; Monte Carlo p value = 0.008](#)) and Aitchison's microbiome distances. This finding persisted when we used an exploratory approach to apply UniFrac to low-depth shotgun metagenomics ([Silverman et al., 2017](#)) ([STAR Methods; Figures S3C and S3D](#)). Interestingly, we did not find a similar correspondence between average microbiome and average diet when using non-tree-based food distances calculated directly from the food choice profiles ([Figure S3E; Procrustes of microbiome Aitchison and standardized food profile Euclidean distances; Monte Carlo p value = 0.6](#)), nor from average nutrient intake using 65 macro- and micronutrients ([Figure 2F; Procrustes of microbiome Aitchison and standardized nutrient profile Euclidean distances; Monte Carlo p value = 0.4](#)), further reinforcing the utility of applying tree-based metrics to food-intake profiles instead of relying on conventional nutrient profiles to understand dietary relationships with the microbiome. Average nutrient intake did correspond with tree-based food distances ([Figure S3F; Procrustes; Monte Carlo p value = 0.002](#)), demonstrating that the tree-based food distances do capture information about nutrient composition. Using constrained ordination with the first 5 principal coordinates in unweighted Unifrac tree-based food distance space, we found that diet accounted for 44% of the total variation in average microbiome composition, although food

groupings with more than one member included in the visualization are assigned similar colors: members of genus *Bacteroides* are red or pink; members of family *Rikenellaceae* are green; family *Prevotellaceae* are yellow; genus *Parabacteroides* are blue; family *Lachnospiraceae* are light purple; and order *Clostridiales* are dark purple.

(B) Relative abundance of the 20 most variable (highest coefficient of variation) functional modules, by day, over the study period ordered by microbiome similarity. See [Figure S2](#) for stable functions.

(C) Relative abundance of dietary intake shown as food groups by day over the study period. All dietary data are collapsed into 9 food groups using the Food and Nutrient Database for Dietary Studies food coding scheme ([STAR Methods](#)).

(D) Square root (sqrt) relative abundance of dietary intake shown as grams of macro- and micronutrients for each subject. *Subjects 11 and 12 primarily consumed meal replacement beverages throughout the duration of the study. Slight variation in nutrient intake within these subjects' time course is due to consumption of coffee (subject 11) or pizza and energy drinks (subject 12). While these subjects are shown next to each other, they do not cluster together based on microbiome similarity, as can be seen in the clustering dendrogram ([Figure S1](#)).

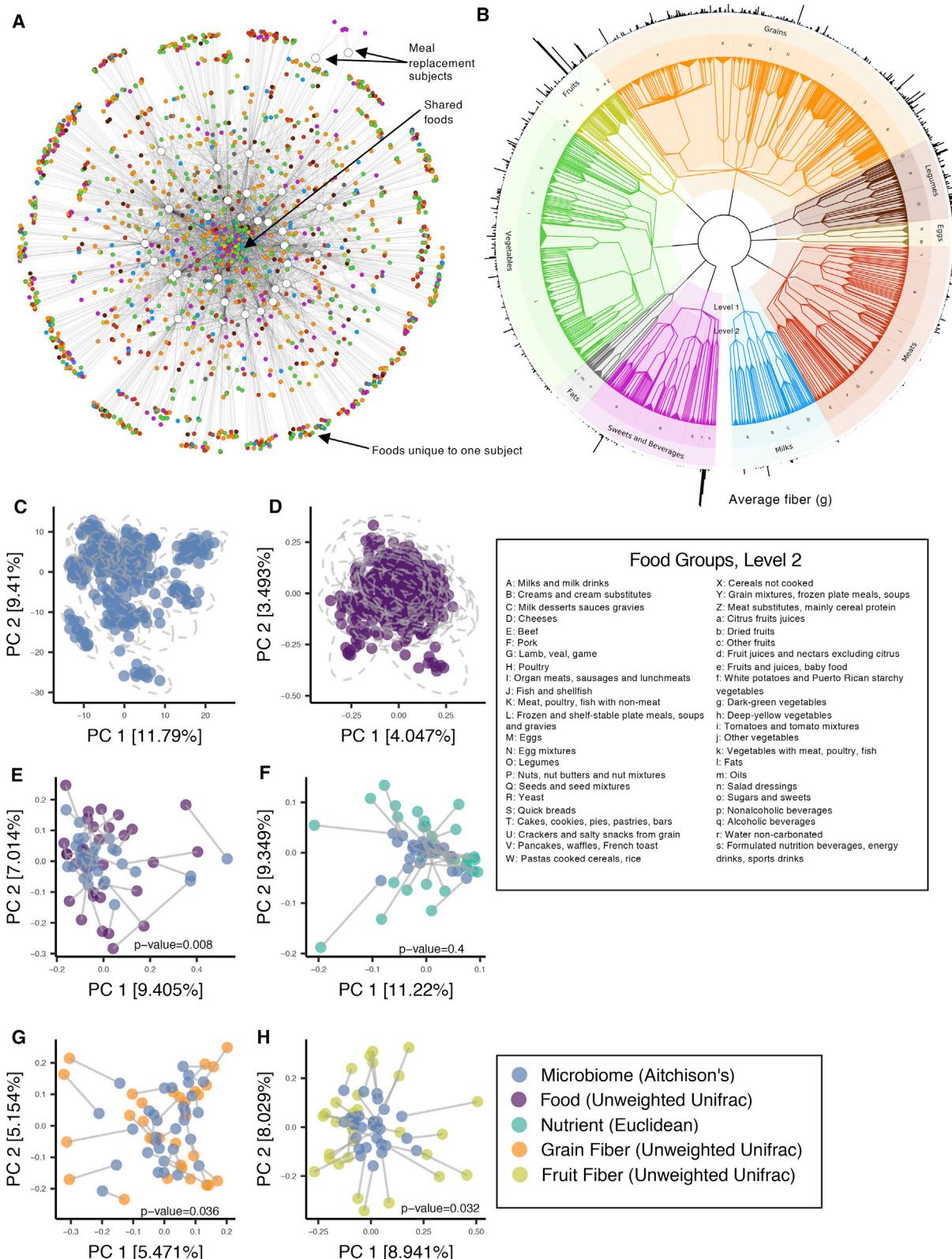


Figure 2. Dietary Diversity Associates with Microbiome Composition Using Tree-based Analysis of Food Choices

(A) Network diagram of food choices created using Cytoscape edge-weighted spring embedded layout. White circles represent subjects, and colored circles represent individual food choices. Individual food choices are colored by food group corresponding to the 9 food major food groups (see Figure 2B). Foods (legend continued on next page)

distances between people are confounded with differences in gender, BMI, and age, which independently accounted for 34% of the unconstrained explained variation in community structure.

In addition to analyzing overall between-subject dietary variation, we used the tree-based diet method to compare between-subject variation in fiber sources with the microbiome. We calculated fiber-source beta-diversity using unweighted UniFrac distances for four separate food groups that have high fiber content: grains, fruits, vegetables, and legumes. Using these distances, we compared fiber-source beta-diversity to microbiome beta-diversity and found that subjects who obtained their fruit fiber or their grain fiber from similar foods tended to have more similar microbiome profiles (Figures 2G and 2H; Procrustes; Monte Carlo p value = 0.036 and 0.032, respectively). Controlled feeding trials assessing the specific impact of fiber-source variety on microbiome composition are needed to investigate these relationships further, particularly to investigate the impact of vegetable and legume fiber (Figures S3G and S3H, respectively).

Daily Diet-Microbiome Associations Are Personalized Diet and Microbiome Covary within Subjects

We hypothesized that microbiome composition and food choices would pair longitudinally within each subject. We assessed the pairing of microbiome beta-diversity and tree-based food beta-diversity longitudinally within each subject using Procrustes analysis. We completed this analysis for 29 subjects consuming a habitual diet for whom we had at least 10 longitudinal time points available. For 78% of our subjects, we found significant longitudinal pairing of diet with the microbiome when using a decaying weighted average of dietary history weighted by 2^{-n} for the n^{th} prior day of food records to pair with a given microbiome sample (Monte Carlo p value ≤ 0.05 ; Figure 3A; STAR Methods). Interestingly, only 21% of our subjects had significant pairing using the single-day preceding food record for each microbiome sample, but pairing efficacy improved by adding additional days of food records. When considering median p values across the subjects, the decaying average described above performed optimally; using the sum of 2, 3, 4, or 5 days of food records performed almost as well (Figure 3B) and resulted in significant pairing for most of the same subjects as captured by the decaying weighted average (Figure 3C).

To test whether this improved multi-day food-record-microbiome association was simply due to the gastrointestinal transit time causing a delayed effect of single-day diet on the microbiome, versus a more complex interaction between multiple days of diet and varying growth rates among the microbial taxa, we compared single-day diet-microbiome associations when using day -1 (1 day prior) or day -2, -3, -4, or -5 as the single day. We found that longitudinal Procrustes pairing of the microbiome composition with a single day of dietary intake from between 1 and 5 days prior did not perform as well as using multiple days of dietary intake data for each microbiome sample (Figure S4). Combined, these findings suggest that daily microbiome variation depends on multiple days of recent dietary history.

Food-Species Correlations Are Personalized

Daily longitudinal sampling allowed us to test directly for associations between dietary components and microbiome composition within each subject over time. We used the decaying weighted average of dietary intake (described above) prior to a given stool sample, and we collapsed dietary food choices into FNDDS food-group categories. We calculated Spearman's correlations between the decaying historical average of level 2 food groups (see Figure 2B) and each bacterial taxon at the species-level or higher within each study subject. We excluded the shake drinkers from this analysis because of their low temporal diet variation. Visualization of the most significant correlations (false discovery rate [FDR] < 0.1) by food group and family-level taxonomy revealed several interesting patterns (Figure 4A). Some subjects have multiple species within the same family correlated with foods from the same food group. Other subjects have correlations between species in different families all responding similarly to the same food group. Two subjects (subjects 7 and 32) had an unusually high number of food-species correlations (more than 60 significant correlations; FDR < 0.1). Further interesting relationships can be visualized where more than one subject's significant values cluster together, suggesting some conservation of these relationships across people. Across the 29 individuals who had more than 10 longitudinal data points available, 83% had at least one significant correlation between the food group and species (FDR < 0.2), with a median of 9 (5th and 95th quantiles 0 and 225, respectively) significant food-species correlations per subject. Of the within-individual food-species significant correlations, 109 are found in more

closest to the center of the image are shared foods consumed by many subjects; foods on the periphery of the network are unique foods reported only by 1 or 2 subjects. Of note, the two meal replacement beverage subjects are located in the upper right quadrant of the figure and are connected only to the larger network because of their consumption of small quantities of other foods.

(B) The tree of foods representation of food groups is a hierarchical, phenetic tree that is built from the Food and Nutrient Database for Dietary Studies. Unique foods are the leaves on the tree and are grouped into one of the 9 level 1 food groups. Level 2 food groups (see legend inset) are included around the tree to highlight the specificity of the food groupings. The black bars extending from the periphery correspond to the average amount of fiber reported by subjects for each food and is included to demonstrate the flexibility of the tree for assessment of fiber-source diversity.

(C and D) (C) Microbiome composition beta-diversity (Aitchison's distance) and (D) dietary food choice beta-diversity (tree-based food diversity; unweighted UniFrac) for each submitted sample and food record, demonstrating within-subject microbiome and diet similarity. Gray ellipses show the 95% confidence intervals around each subject's points. Shake drinkers are excluded from these figures. See Figure S3 for the same plots colored by subject.

(E) Significant agreement between average diet and average microbiome distances (Aitchison's distance) by Procrustes analysis when food intake distances are determined with unweighted UniFrac using the tree of foods (Monte Carlo p value = 0.008).

(F–H) (F) No significant agreement between average diet and average microbiome distances (Aitchison's distance) by Procrustes analysis when food intake distances are determined from normalized nutritional intake variables (Monte Carlo p value = 0.4). Significant Procrustes agreement between average microbiome composition (Aitchison's distance) and fiber-source diversity from (G) grains and (H) fruits determined using the tree-based distances from fiber content of foods (Unweighted UniFrac) (Monte Carlo p values = 0.036 and 0.032, respectively).

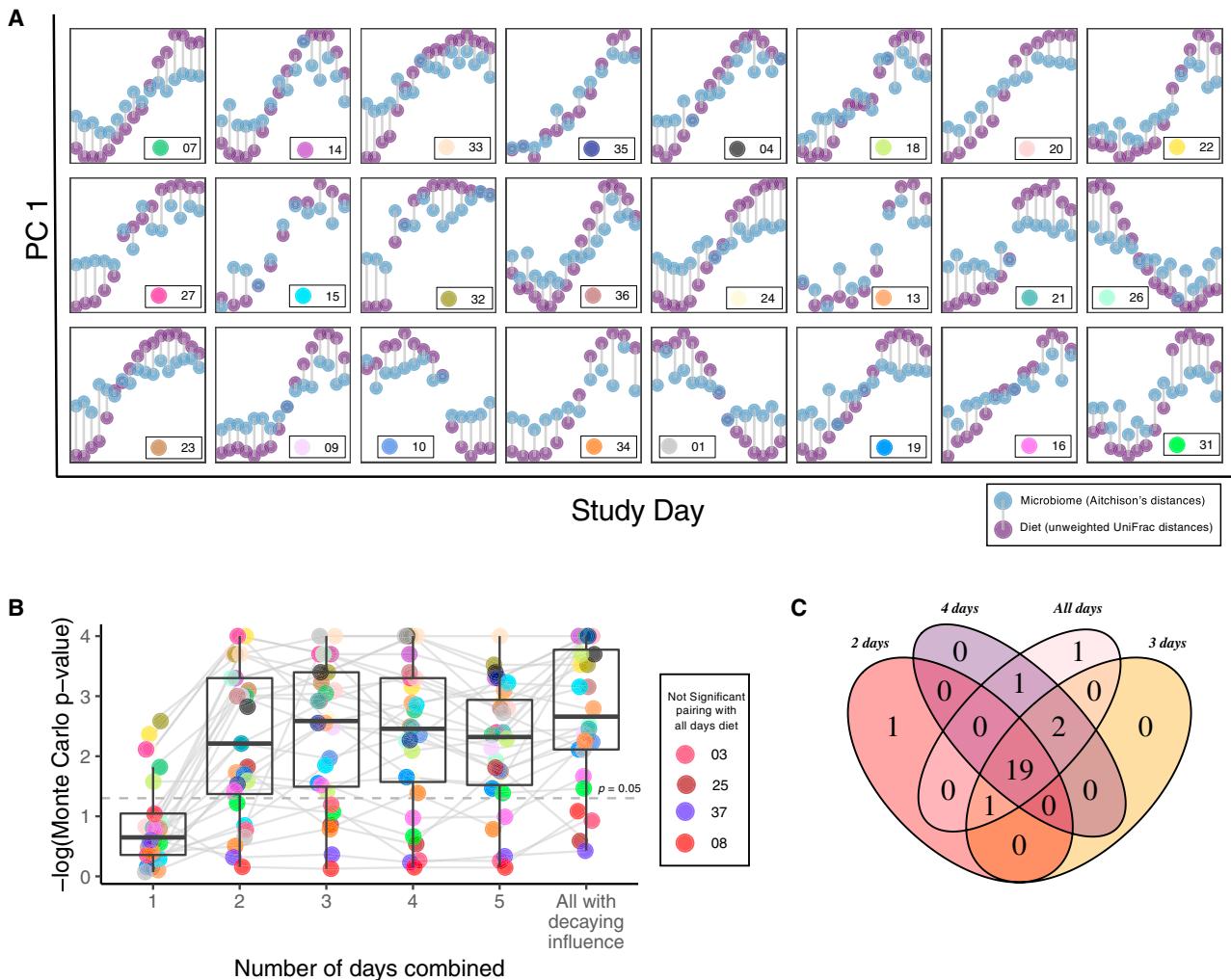


Figure 3. Individual Longitudinal Microbiome Diet-Agreement Is Improved with More than One Consecutive Day of Dietary Records

(A) Significant within-subject, longitudinal diet-microbiome pairing between the microbiome and decaying weighted average of dietary intake (Procrustes analysis; Monte Carlo p value < 0.05; plots are ordered left to right, top to bottom by lowest to highest p value).

(B) Monte Carlo p values ($-\log$ transformed) for Procrustes pairing between microbiome and diet by subject when multiple days of dietary intake are combined before pairing, demonstrating improvement in pairing when more than one day of diet is considered for within-subject longitudinal Procrustes pairing. Points are colored by subject. Dashed line indicates $-\log(0.05)$. See Figure S4 for pairing by offset days.

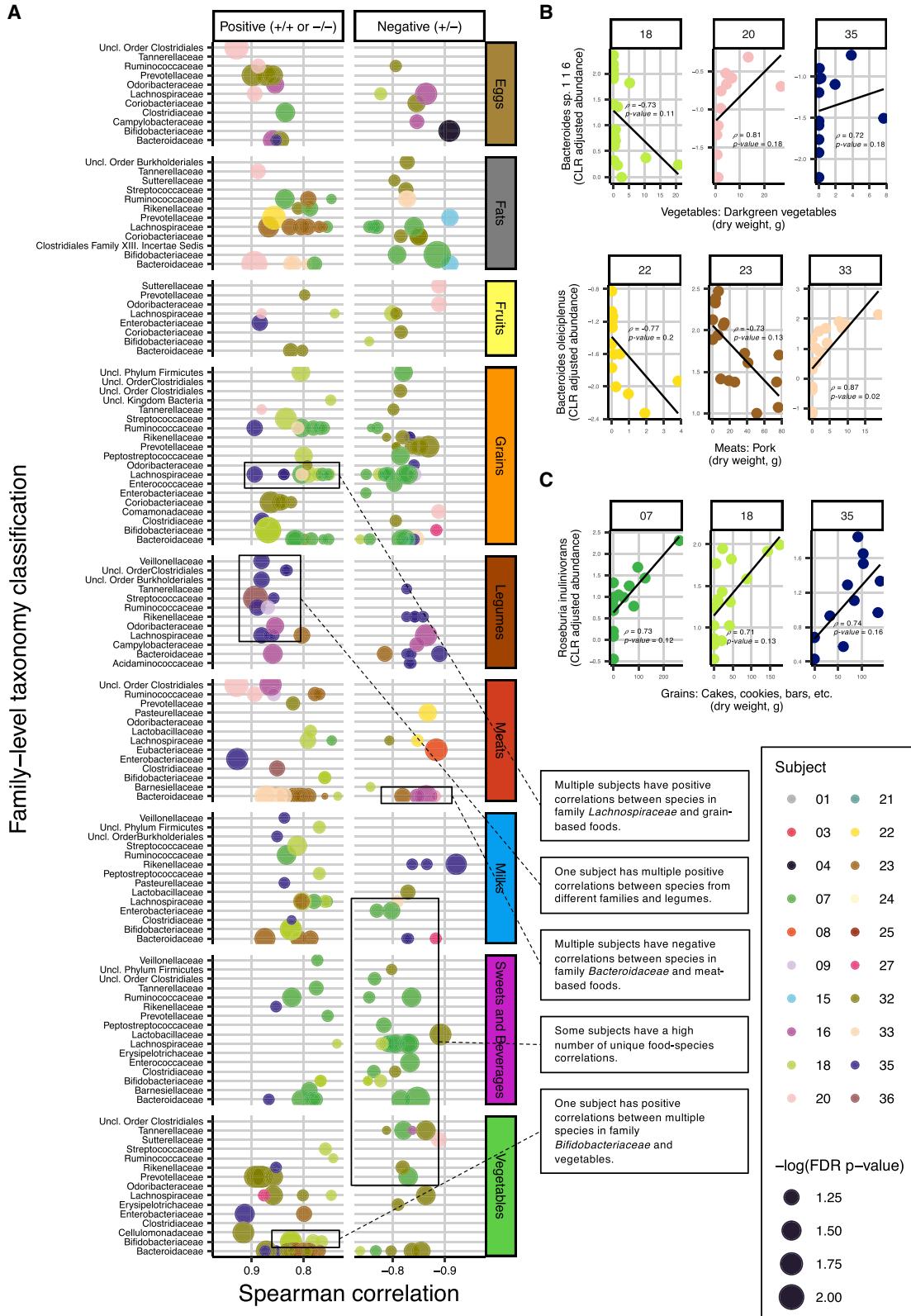
(C) Venn diagram showing the number of subjects with significant diet-microbiome Procrustes agreement for 2-, 3-, and 4-day averages and a decaying weighted average of all days (Significance defined as Monte Carlo p value < 0.05).

than 1 subject, while only 8 are found in more than 2 people (FDR < 0.2; Table S2). Interestingly, the directionality of these food-species relationships is not always conserved across people. For example, of the 8 significant correlations found in more than 2 people, 5 are in opposite directions (two examples shown in Figure 4B), and 3 have directionality conserved across people (one example shown in Figure 4C).

Dietary Patterns Improve Prediction of Next-Day Species Abundance

To confirm the personalized nature of diet-species relationships, we tested whether one subject's daily diet could predict daily changes in the microbiome. Microbiomes are already autocorrelated over time ($p < 0.001$; paired Mann-Whitney U test; Fig-

ure 5A), and indeed, we found that we could predict a subject's next-day species profiles with reasonable accuracy using the first five microbiome principal components in a linear model (mean Pearson $r = 0.7$) (Figure 5B). To determine whether recent diet improved our ability to predict the next day's microbiome, we trained a model to use dietary features to predict the residuals from the above microbiome-only prediction. The dietary predictors were the first five dietary principal components (tree-based unweighted UniFrac distances for decaying weighted average food intake). Regressing the microbiome residuals on dietary principal components isolated the impact of diet alone on predicting the next-day microbiome. Using dietary features from the same subject improved the fit of the prediction over microbiome alone by 13% (mean Pearson $r = 0.83$; Tukey

**Figure 4. Daily Diet-Microbiome Associations Are Personalized**

(A) Bubble-plot visualization of within-subject correlations between level 2 food groups and species-level taxonomy assignments (Spearman's correlation; FDR-corrected p value < 0.1). For visual simplicity, correlations between level 2 food groups and species are displayed by their membership in level 1 food groups

(legend continued on next page)

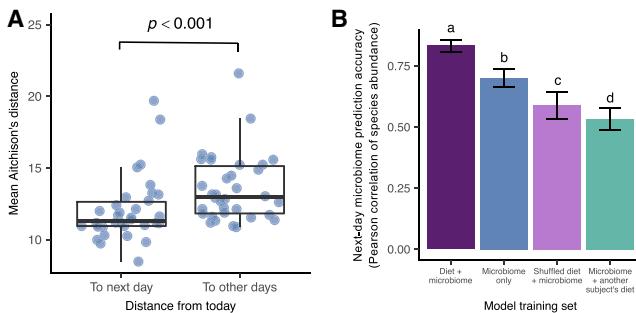


Figure 5. Dietary Patterns Improve Prediction of Next-Day Species Abundance

(A) The mean Aitchison's distance between consecutive microbiome samples is less than the mean distance between a sample and its distance to all other days (paired Wilcoxon signed rank test; $p < 0.001$).

(B) Personalized dietary features improve the prediction of next-day species abundance for the top 20 most abundant species from microbiome features alone and from microbiome features and the same subject's temporally shuffled dietary features or another subject's dietary features. Data are represented as mean Pearson correlation \pm standard error; groups that do not share a letter are significantly different; Tukey's test; $p < 0.05$.

test; $\alpha = 0.05$). In contrast, using dietary features from a randomly selected subject decreased the fit of the prediction relative to microbiome alone by 17% (mean Pearson $r = 0.53$; Tukey test; $\alpha = 0.05$), demonstrating that simply adding unrelated dietary features was not responsible for the increased prediction accuracy. Furthermore, a model using randomly shuffled dietary features from the same subject to disrupt the temporal ordering of dietary days also showed decreased predictive strength (mean Pearson $r = 0.59$). Combined, these results demonstrate that a person's recent dietary history and current microbiome state can be used to predict their microbiome state on the next day with reasonable accuracy.

A Stable Diet Is Not Associated with a Stable Microbiome

The inclusion of the shake drinkers in this dataset enabled a unique analysis of microbiome stability and dietary stability and diversity (Figure 6A). Here, we define both microbiome and dietary stability as the median inverse distance between repeated measures of day-to-day beta-diversity. We defined dietary diversity as an extension of the traditional definition, the simple number of foods consumed each day (Ruel, 2003) or between-food Jaccard distances (Otto et al., 2015), by leveraging the food tree and a tree-based alpha-diversity measure (Faith's phylogenetic diversity [Faith, 1992]).

As can be seen in the representative dietary intake of subject 10 (Figure 6A), dietary intake in most subjects varied by day. For example, subject 10 reported consumption of eggs on weekend days. We found that weekend diets are more variable (less

stable) than weekday diets. The mean diet distance between Saturday and Sunday was larger than the mean diet distance between consecutive days of the week (paired Wilcoxon signed rank test; p value = 0.008; Figure 6B). However, we found that the microbiome distances from weekend-affected microbiome days (i.e., Monday to Tuesday, 3-day offset from variable dietary intake) were not larger than the distances between weekday-diet-affected days (i.e., Wednesday to Thursday, Thursday to Friday, etc.) (Figure 6C), and median dietary food stability (i.e., to what degree a subject's food intake changes from day to day) did not correlate with median microbiome stability (Figure S5). This analysis showed that any periodicity in the dietary data was not detectable in the microbiome data.

Analysis of the shake drinkers' data further supports the hypothesis that dietary monotony alone is not sufficient to stabilize the microbiome. The two shake drinkers were outliers in dietary stability (i.e., high stability, eating the same foods consistently from day to day; Figures 6A and S5), but they had microbiome stability profiles that were not outliers when compared with other subjects. We found that the day-to-day changes in microbiome composition of these subjects appeared similar to those of other subjects, and the distance between successive microbiome days did not decrease over the study (Figure 6D). It therefore seems possible that stability is an intrinsic property of the microbiome community that is shaped by community membership, rather than the stability of diet. Indeed, we found that 5 species were positively correlated with microbiome stability (*Alistipes* sp. CHKC1003, *Alistipes onderdonkii*, unclassified genus *Alistipes*, *Bacteroides uniformis*, and *Clostridium phoceensis*) and 3 species were negatively correlated with microbiome stability (*Prevotella copri*, *Dakarella massiliensis*, and *Dorea longicatena*) (Spearman's correlation; FDR p value < 0.2 ; Figure S5).

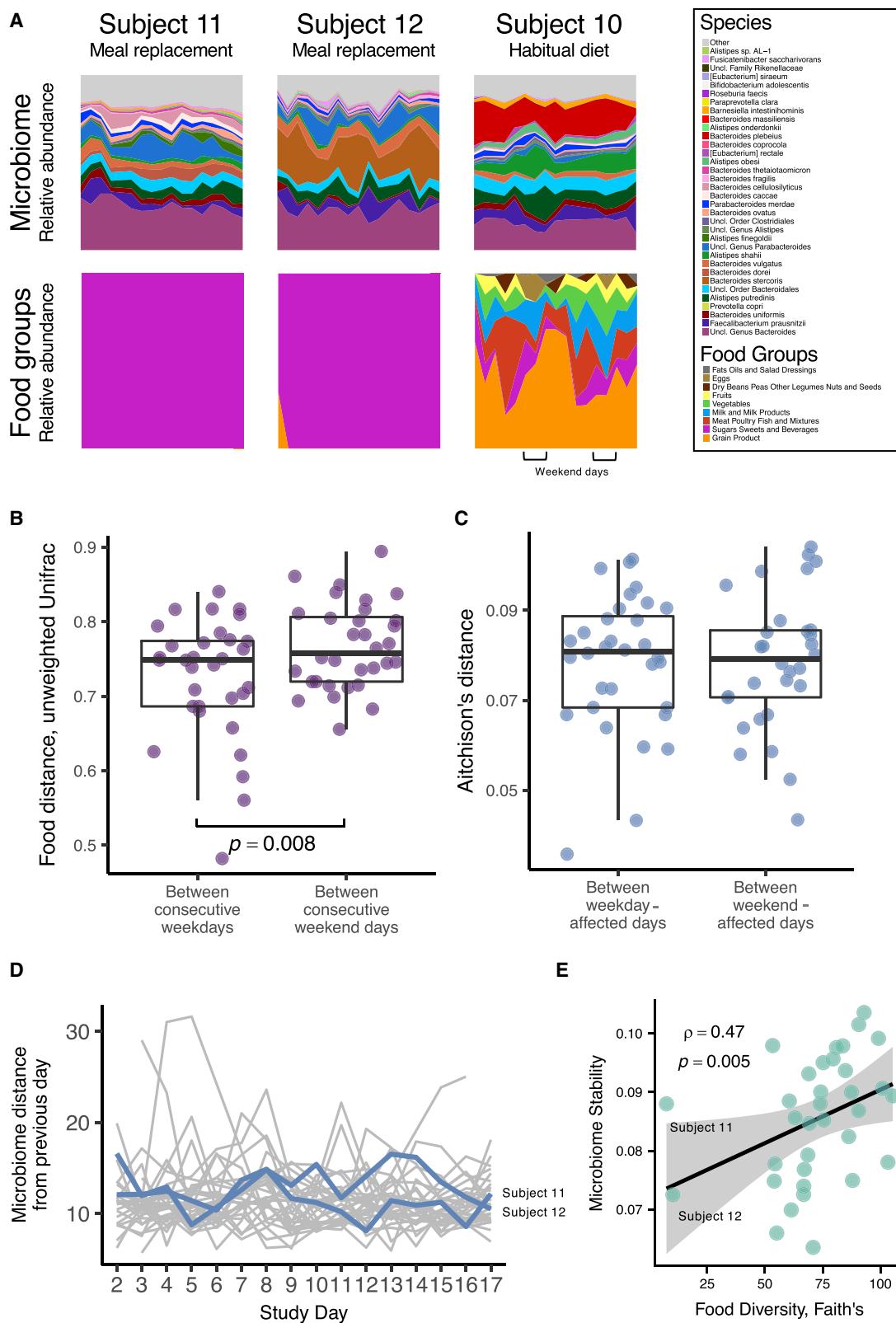
Increased dietary diversity has been hypothesized to contribute to a healthy microbiome (Heiman and Greenway, 2016). Unlike others (Claesson et al., 2012), we found no relationship between microbiome diversity and dietary diversity. However, to understand how diet affects gut ecology and resilience of gut communities, we assessed the contribution of dietary diversity to the stability of microbiome composition (Zaneveld et al., 2017). We found that microbiome stability (inverse median Aitchison's distance to self) was correlated with dietary diversity (Faith's phylogenetic diversity; Spearman's correlation $\rho = 0.47$; p value = 0.005; Figure 6E), demonstrating that subjects who habitually consumed a diverse diet (i.e., various foods from different food groups each day) had more stable microbiomes. This correlation persisted both with and without the inclusion of the shake drinkers.

DISCUSSION

We investigated daily changes in human gut microbiome composition and their relationship to daily dietary intake by collecting

and taxonomic families. Positive correlations are on the left panel, and negative correlations are on the right panel. Stronger correlations are on the outer edges of the plot. The size of each bubble corresponds to the $-\log$ (FDR-corrected p value) with larger bubbles representing lower p values. Bubbles are colored by subject.

(B and C) Two examples of significant Spearman's correlations between level 2 food groups and species with opposite directionality across people and (C) one example of conserved directionality are shown (all FDR-corrected p values < 0.2). See Table S2 for the complete list of significant p values and correlations seen in more than 2 subjects.



(legend on next page)

full daily dietary records and daily stool samples for shotgun metagenomics from 34 subjects over 17 days. Our longitudinal dataset allowed us to test whether each subject's daily diet-microbiome interactions were unique to that individual. We found that the majority of subjects had detectable interactions between daily changes in foods and daily changes in species relative abundance in their microbiome. However, very few of these food-microbe interactions were conserved across people. We built a custom predictive model for each subject that could predict the next day's microbiome profile using only the current microbiome and recent diet of that subject; substituting a model trained on another subject's diet and microbiome caused the prediction to fail. These findings showed that food-microbe interactions were highly personalized. This personalized response might be one reason for the small observed effect sizes of diet in shaping the gut microbiome in population-level studies (Falony et al., 2016; Rothschild et al., 2018; Wang et al., 2016). In other words, the effect of diet might be much greater than these studies suggest, but if the same foods impact different bacterial populations in an individualized manner, such effects may not be detectable in cross-sectional studies.

In the same way that recent findings have shown that probiotic interventions have different effects in different people (Maldonado-Gómez et al., 2016; Zmora et al., 2018; Korem et al., 2017), our findings suggest that a food-based intervention aimed at promoting or suppressing a particular bacterial species may need to be tailored to the specific composition of a subject's microbiome. Nonetheless, we find it likely that specific dietary compounds will have consistent effects on certain microbial strains or metabolic pathways. Others have recently demonstrated that blood glucose responses to foods are personalized and may depend on an individual's unique microbiome composition (Zeevi et al., 2015). However, the specifics of personalized diet-microbiome interactions still need to be investigated beyond their relationship with markers of blood glucose control. Future work with well-controlled feeding studies should be considered given that differences between individuals may also be driven by subtle anatomical variation in the gastrointestinal tract or immunological differences between people.

Here, we found that using conventional nutrient profiles alone to assess diet was insufficient to connect dietary intake with microbiome variation. We believe this is because of limited information available about nutrients and substrates that are relevant to the microbiota. The database that we used in this study, FNDDS (U.S. Department of Agriculture and Agricultural Research Service, 2014), reported over 50 macro- and micronutrients such as poly- and mono-unsaturated fats and numerous other vitamins and minerals. However, these nutrient profiles

ignore hundreds of additional chemical compounds present even in a single piece of fruit (Schwieterman et al., 2014) and many other non-nutritive components of foods such as preservatives and food additives. The fact that our nutritional databases did not contain information about dietary substrates like formates or nitrates could explain our inability to detect correlations between microbial functions and nutrient intake. Even for food components that are in our nutritional databases, such as fiber, there are many different chemical structures that are jointly referred to as "fiber" but that are used by different bacteria with differing sets of enzyme capabilities, therefore obscuring our ability to detect relationships. Our tree-based dietary analysis approach is an alternative starting point from which to consider related foods as complex admixtures of similar compounds. Future work determining the composition of foods using mass spectrometry will likely reveal meaningful associations between foods and could be used to improve our understanding about the relatedness of foods from a biochemical perspective. Projects such as the Food Biomarkers Alliance have resulted in the development of databases of known food metabolites including FooDB (Wishart, 2016), Phenol-Explorer (Rothwell et al., 2012), (Manach et al., 2016), and Phytohub (da Silva et al., 2016). These projects are beginning to address the need to better characterize food-derived nutritional molecules. Connecting these food metabolomic data and other glycobiomic data that capture the complexity of fibers to microbiome signatures has the potential to lead to significant advances in understanding both population-level and personalized diet-microbiome relationships.

An interesting discovery of this study was that multi-day food records preceding a stool sample were required to identify statistically significant interactions between diet and the microbiome longitudinally. A model using exponential decay of the effect of dietary history on daily microbiome variation performed better than simple averaging of previous days, supporting the notion that yesterday's diet has the largest effect on tomorrow's microbiome, the day before yesterday has a moderate effect, the day prior a yet smaller effect, and so on. We hypothesize that this finding is partly due to variations in transit time. Recent pilot work using ingestible sensors has clearly demonstrated that fiber intake is associated with differing small intestinal and colonic transit times (Kalanter-Zadeh et al., 2018). Using the decaying average should account for some of this variation. There may also be an ecological cause contributing to this finding: if the ecological perturbations caused by a certain change in dietary intake reverberate for several days due to cross-feeding or competitive dynamics, then the current microbiome state will be affected jointly by multiple days of recent dietary history.

Figure 6. Dietary Diversity Correlates with Microbiome Stability

- (A) Expanded view of daily species-level microbiome relative abundance and food-group relative abundance for the two shake drinkers (subject 11 and subject 12) and a representative subject (subject 10) consuming a habitual diet. Of note, weekend variation is visible in the dietary intake of subject 10, with a marked increase in egg consumption. Subject 11 and subject 12 are highlighted as dietary diversity outliers. See also Figure S5A.
- (B) Dietary intake (food beta-diversity; unweighted UniFrac) is more similar between consecutive weekdays (e.g., Monday to Tuesday, Tuesday to Wednesday, etc.) than between consecutive weekend days (Saturday and Sunday) (paired Wilcoxon signed rank test; p value = 0.008).
- (C) Microbiome composition (Aitchison's distance) is more not different between weekday affected days and weekend-affected days (3-day offset from dietary intake) (paired Wilcoxon signed rank test; p value = 0.45).
- (D) Microbiome Aitchison's distances between successive days for each subject in gray, with subjects 11 and 12 highlighted in blue.
- (E) Average daily dietary (food) diversity (Faith's phylogenetic diversity) is positively correlated with microbiome stability.

Overall, our study deepens our understanding of personalized diet-microbiome interrelationships. While conventional nutrient-based analysis of diet alone is inadequate for studying diet-microbiome interaction, we show that statistical information can be shared across related foods to overcome the lack of overlap between the foods eaten by different people. This approach, in combination with longitudinal sampling and multi-day dietary records, revealed personalized diet-microbiome interactions and may be useful in future population-wide studies that aim to determine the exact role of dietary foods in shaping the gut microbiome.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **CONTACT FOR REAGENT AND RESOURCE SHARING**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Human Subjects
 - Inclusion and Exclusion Criteria
 - Sample Size Estimation
 - Randomization
 - Health Status
 - Ethnicity
 - Other Details
- **METHOD DETAILS**
 - Experimental Design
 - Dietary Supplementation
 - Dietary Record Keeping
 - Fecal Sample Collection and Processing
 - DNA Extraction and Sequencing
 - Blood Collection and Processing
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Shotgun Metagenomic Sequencing
 - Microbiome Functional Profiling
 - Nutrient Profile Analysis
 - Food Choice Analysis
 - Cleaning Data and Removal of Outliers
 - Development and Application of Tree-based Analysis for Food
 - Decaying Weighted Average of Food Intake
 - FDR Correction
 - Procrustes Analysis
 - Species Prediction from Diet and Microbiome
- **DATA AND SOFTWARE AVAILABILITY**
 - Software
 - Data Resources

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.chom.2019.05.005>.

ACKNOWLEDGMENTS

We thank the study subjects for their dedication to this study. The citizen science course associated with this study was supported by the College of Biological Sciences and the BioTechnology Institute, both at the University

of Minnesota. We thank the Personalized Microbiome Class students who contributed to early analysis of this dataset and were the source of fruitful discussions related to these data. We specifically thank Abdirahman Sheikh-Mohamed and Patrick Johnson for their volunteer contributions to this study. We are thankful to the Knights Lab members and collaborators who provided early insight into analysis and interpretation of the data during early manuscript development. We thank the study's reviewers, who greatly improved the quality of the manuscript during review. This study was supported by General Mills.

AUTHOR CONTRIBUTIONS

Conceptualization, D.K. and A.J.J.; Methodology, D.K. and A.J.J.; Software, A.J.J., P.V., G.A.-G., B.H., T.W., and A.D.K.; Formal Analysis, A.J.J.; Investigation, A.J.J., R.S.C., and A.N.S.; Resources, D.K.; Data Curation, A.J.J.; Writing – Original Draft, A.J.J.; Writing – Review & Editing, D.K., G.A.-G., B.H., P.V., R.M., K.K., R.S.C., T.W., A.J.J., and J.W.; Visualization, A.J.J.; Supervision, D.K.; Project Administration, A.J.J. and D.K.; Funding Acquisition, D.K., R.M., and K.K. All authors read and approved the final manuscript. Personal Microbiome Class students (Jesse Abelson, Breann Elise Abernathy, Kasey Ah Pook, Adam Rashid Ahmed, Khadir Tababouanga Albert, Samantha Tappe Brook, Drew Benjamin Carter, Elizabeth S. Dean, Tong Ding, Grant Andrew Hedblom, Laura Beth Jore, Vasantha Lakshminpathy, Leeore Levinstein, Sarah Anna Marie Lucas, Morrine Auma Omolo, Kayla Donnelle Pederson, Philips, Kinsey H., Austin Michael Schachtnner, Reed Grace Owens-Kurtz, Justin Wiertzema, Sambhawa Priya, and Serina Lee Robinson) provided ideas for analysis.

DECLARATION OF INTERESTS

D.K. serves as CEO of CoreBiome, a company involved in the commercialization of microbiome analysis. CoreBiome is now a wholly owned subsidiary of OraSure. These interests have been reviewed and managed by the University of Minnesota in accordance with its conflict-of-interest policies. T.W. is an employee of CoreBiome. R.M. and K.K. are employees of General Mills.

Received: October 6, 2018

Revised: January 31, 2019

Accepted: May 9, 2019

Published: June 12, 2019

REFERENCES

- Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J.A., and Pawlowsky-Glahn, V. (2000). Logratio analysis and compositional distance. *Math. Geol.* 32, 271–275.
- Al-Ghalith, G., and Knights, D. (2017). BURST enables optimal exhaustive DNA alignment for big data.
- Al-Ghalith, G., and Knights, D. (2018). aKronyMer enables database-free metagenome comparison.
- 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.
- Asnicar, F., Weingart, G., Tickle, T.L., Huttenhower, C., and Segata, N. (2015). Compact graphical representation of phylogenetic data and metadata with GraPhAn. *PeerJ* 3, e1029.
- Barratt, M.J., Lebrilla, C., Shapiro, H.Y., and Gordon, J.I. (2017). The gut microbiota, food science, and human nutrition: a timely marriage. *Cell Host Microbe* 22, 134–141.
- Bowyer, R.C.E., Jackson, M.A., Pallister, T., Skinner, J., Spector, T.D., Welch, A.A., and Steves, C.J. (2018). Use of dietary indices to control for diet in human gut microbiota studies. *Microbiome* 6, 77.
- 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.

- Claesson, M.J., Jeffery, I.B., Conde, S., Power, S.E., O'Connor, E.M., Cusack, S., Harris, H.M.B., Coakley, M., Lakshminarayanan, B., O'Sullivan, O., et al. (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature* 488, 178–184.
- da Silva, A.B., Giacomoni, F., Pavot, B., Fillatre, Y., Rothwell, J., Sualdea, B.B., Veyrat, C., Garcia-Villalba, R., Gladine, C., Kopec, R., et al. (2016). PhytoHub V1.4: a new release for the online database dedicated to food phytochemicals and their human metabolites. In the international conference on food bioactives & health.
- David, L.A., Materna, A.C., Friedman, J., Campos-Baptista, M.I., Blackburn, M.C., Perrotta, A., Erdman, S.E., and Alm, E.J. (2014a). Host lifestyle affects human microbiota on daily timescales. *Genome Biol.* 15, R89.
- 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. (2014b). Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505, 559–563.
- Faith, D.P. (1992). Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* 61, 1–10.
- Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., Kurilshikov, A., Bonder, M.J., Valles-Colomer, M., Vandepitte, D., et al. (2016). Population-level analysis of gut microbiome variation. *Science* 352, 560–564.
- Flores, G.E., Caporaso, J.G., Henley, J.B., Rideout, J.R., Domogala, D., Chase, J., Leff, J.W., Vázquez-Baeza, Y., Gonzalez, A., Knight, R., et al. (2014). Temporal variability is a personalized feature of the human microbiome. *Genome Biol.* 15, 531.
- Fukuyama, J., Rumker, L., Sankaran, K., Jeganathan, P., Dethlefsen, L., Relman, D.A., and Holmes, S.P. (2017). Multidomain analyses of a longitudinal human microbiome intestinal cleanout perturbation experiment. *PLoS Comput. Biol.* 13, e1005706.
- Gentile, C.L., and Weir, T.L. (2018). The gut microbiota at the intersection of diet and human health. *Science* 362, 776–780.
- Gloor, G.B., Wu, J.R., Pawlowsky-Glahn, V., and Egoscue, J.J. (2016). It's all relative: analyzing microbiome data as compositions. *Ann. Epidemiol.* 26, 322–329.
- Heiman, M.L., and Greenway, F.L. (2016). A healthy gastrointestinal microbiome is dependent on dietary diversity. *Mol. Metab.* 5, 317–320.
- Hillmann, B., Al-Ghalith, G.A., Shields-Cutler, R.R., Zhu, Q., Gohl, D.M., Beckman, K.B., Knight, R., and Knights, D. (2018). Evaluating the information content of shallow shotgun metagenomics. *mSystems* 3.
- Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214.
- Kalanter-Zadeh, K., Berean, K.J., Ha, N., Chrimis, A.F., Xu, K., Grando, D., Ou, J.Z., Pillai, N., Campbell, J.L., Brkljacic, R., et al. (2018). A human pilot trial of ingestible electronic capsules capable of sensing different gases in the gut. *Nat. Electron. 1*, 79–87.
- Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30.
- Knights, D., Costello, E.K., and Knight, R. (2011). Supervised classification of human microbiota. *FEMS Microbiol. Rev.* 35, 343–359.
- Korem, T., Zeevi, D., Zmora, N., Weissbrod, O., Bar, N., Lotan-Pompan, M., Avnit-Sagi, T., Kosower, N., Malka, G., Rein, M., et al. (2017). Bread affects clinical parameters and induces gut microbiome-associated personal glycemic responses. *Cell Metab.* 25, 1243–1253.
- Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235.
- Mackay, D.S., Jew, S., and Jones, P.J.H. (2017). Best practices for design and implementation of human clinical trials studying dietary oils. *Prog. Lipid Res.* 65, 1–11.
- Maier, T.V., Lucio, M., Lee, L.H., VerBerkmoes, N.C., Brislaw, C.J., Bernhardt, J., Lamendella, R., McDermott, J.E., Bergeron, N., Heinzmann, S.S., et al. (2017). Impact of dietary resistant starch on the human gut microbiome, metaproteome, and metabolome. *MBio* 8.
- Maldonado-Gómez, M.X., Martínez, I., Bottacini, F., O'Callaghan, A., Ventura, M., van Sinderen, D., Hillmann, B., Vangay, P., Knights, D., Hutkins, R.W., et al. (2016). Stable engraftment of *Bifidobacterium longum* AH1206 in the human gut depends on individualized features of the resident microbiome. *Cell Host Microbe* 20, 515–526.
- Manach, C., Unit, H.N., and Clermont-Ferrand, U. (2016). FoodComEx, a new chemical library useful for the food bioactives community. In International conference on food bioactive and health 2016.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., and Wagner, H. (2019). Vegan: community ecology package. R Package version 2.5-4.
- O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745.
- Otto, M.C., Padhye, N.S., Bertoni, A.G., Jacobs, D.R., Jr., and Mozaffarian, D. (2015). Everything in moderation - dietary diversity and Quality, Central obesity and risk of diabetes. *PLoS One* 10, e0141341.
- Park, Y., Dodd, K.W., Kipnis, V., Thompson, F.E., Potischman, N., Schoeller, D.A., Baer, D.J., Midthune, D., Troiano, R.P., Bowles, H., et al. (2018). Comparison of self-reported dietary intakes from the Automated Self-Administered 24-h recall, 4-d food records, and food-frequency questionnaires against recovery biomarkers. *Am. J. Clin. Nutr.* 107, 80–93.
- Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P.I., Godneva, A., Kalka, I.N., Bar, N., et al. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature* 555, 210–215.
- Rothwell, J.A., Urpi-Sarda, M., Boto-Ordoñez, M., Knox, C., Llorach, R., Eisner, R., Cruz, J., Neveu, V., Wishart, D., Manach, C., et al. (2012). Phenol-Explorer 2.0: a major update of the phenol-Explorer database integrating data on polyphenol metabolism and pharmacokinetics in humans and experimental animals. *Database* 2012, bas031.
- Ruel, M.T. (2003). Operationalizing dietary diversity: a review of measurement issues and research priorities. *J. Nutr.* 133, 3911S–3926S.
- Schwieterman, M.L., Colquhoun, T.A., Jaworski, E.A., Bartoshuk, L.M., Gilbert, J.L., Tieman, D.M., Odabasi, A.Z., Moskowitz, H.R., Folta, K.M., Klee, H.J., et al. (2014). Strawberry flavor: diverse chemical compositions, a seasonal influence, and effects on sensory perception. *PLoS One* 9, e88446.
- Shah, N.D., and Limketkai, B.N. (2017). The use of medium-chain triglycerides in gastrointestinal disorders. *Pract. Gastroenterol.* 160, 20–28.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.
- Silverman, J.D., Washburne, A.D., Mukherjee, S., and David, L.A. (2017). A phylogenetic transform enhances analysis of compositional microbiota data. *Elife* 6.
- Smits, S.A., Marcobal, A., Higginbottom, S., Sonnenburg, J.L., and Kashyap, P.C. (2016). Individualized responses of gut microbiota to dietary intervention modeled in humanized mice. *mSystems* 1.
- Sonnenburg, E.D., Smits, S.A., Tikhonov, M., Higginbottom, S.K., Wingreen, N.S., and Sonnenburg, J.L. (2016). Diet-induced extinctions in the gut microbiota compound over generations. *Nature* 529, 212–215.
- Strate, L.L., Keeley, B.R., Cao, Y., Wu, K., Giovannucci, E.L., and Chan, A.T. (2017). Western dietary pattern increases, and prudent dietary pattern decreases, risk of incident diverticulitis in a prospective cohort study. *Gastroenterology* 152, 1023–1030.
- Templ, M., Hron, K., and Filzmoser, P. (2011). robCompositions: an R-package for robust statistical analysis of compositional data. In Compositional data anal. V. Pawlowsky-Glahn and A. Buccianti, eds. (John Wiley & Sons, Ltd), pp. 341–355.
- Teng, Y., Ren, Y., Sayed, M., Hu, X., Lei, C., Kumar, A., Hutchins, E., Mu, J., Deng, Z., Luo, C., et al. (2018). Plant-derived exosomal microRNAs shape the gut microbiota. *Cell Host Microbe* 24, 637–652.

- Thaiss, C.A., Itav, S., Rothschild, D., Meijer, M., Levy, M., Moresi, C., Dohnalová, L., Braverman, S., Rozin, S., Malitsky, S., et al. (2016). Persistent microbiome alterations modulate the rate of post-dieting weight regain. *Nature*.
- U.S. Department of Agriculture, Agricultural Research Service. (2014). USDA food and nutrient database for dietary studies, pp. 2011–2012.
- Vangay, P., Johnson, A.J., Ward, T.L., Al-Ghalith, G.A., Shields-Cutler, R.R., Hillmann, B.M., Lucas, S.K., Beura, L.K., Thompson, E.A., Till, L.M., et al. (2018). US immigration westernizes the human gut microbiome. *Cell* 175, 962–972.
- Wang, J., Thingholm, L.B., Skiecevičienė, J., Rausch, P., Kummen, M., Hov, J.R., Degenhardt, F., Heinsen, F.A., Rühlemann, M.C., Szemczak, S., et al. (2016). Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nat. Genet.* 48, 1396–1406.
- Wishart, D.S. (2016). FooDB. The metabolomics innovation centre 469, 470.
- Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.Y., Keilbaugh, S.A., Bewtra, M., Knights, D., Walters, W.A., Knight, R., et al. (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334, 105–108.
- Wu, G.D., Compher, C., Chen, E.Z., Smith, S.A., Shah, R.D., Bittinger, K., Chehoud, C., Albenberg, L.G., Nessel, L., Gilroy, E., et al. (2016). Comparative metabolomics in vegans and omnivores reveal constraints on diet-dependent gut microbiota metabolite production. *Gut* 65, 63–72.
- Zaneveld, J.R., McMinds, R., and Vega Thurber, R. (2017). Stress and stability: applying the Anna Karenina principle to animal microbiomes. *Nat. Microbiol.* 2, 17121.
- Zeevi, D., Korem, T., Zmora, N., Israeli, D., Rothschild, D., Weinberger, A., Ben-Yacov, O., Lador, D., Avnit-Sagi, T., Lotan-Pompan, M., et al. (2015). Personalized nutrition by prediction of glycemic responses. *Cell* 163, 1079–1094.
- Zeileis, A., and Grothendieck, G. (2005). Zoo: S3 infrastructure for regular and irregular time series. *J. Stat. Soft.* 14, 1–27.
- Zhang, S., Midthune, D., Guenther, P.M., Krebs-Smith, S.M., Kipnis, V., Dodd, K.W., Buckman, D.W., Tooze, J.A., Freedman, L., and Carroll, R.J. (2011). A new multivariate measurement error model with zero-inflated dietary data, and its application to dietary assessment. *Ann. Appl. Stat.* 5, 1456–1487.
- Zhernakova, A., Kurilshikov, A., Bonder, M.J., Tigchelaar, E.F., Schirmer, M., Vatanen, T., Mujagic, Z., Vila, A.V., Falony, G., Vieira-Silva, S., et al. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* 352, 565–569.
- Zmora, N., Zilberman-Schapira, G., SUEZ, Mor, J., Dori-Bachash, U., Bashirades, M., Kotler, S., Zur, E., Regev-Lehavi, M., Brik, D., et al. (2018). Personalized gut mucosal colonization resistance to empiric probiotics is associated with unique host and microbiome features. *Cellule* 174, 1388–1405.

STAR★METHODS**KEY RESOURCES TABLE**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Healthy, human microbiome samples	This paper	Knightslab.org
Critical Commercial Assays		
PowerSoil DNA isolation kit	Mo Bio Laboratories, USA	Catalog No. 12888
Nextera XT DNA library preparation kit	Illumina, Inc	FC-131-1096
Deposited Data		
Shotgun metagenomic sequences data	This paper; European Nucleotide Archive	ENA: PRJEB29065
RefSeq version 86	O'Leary et al. (2016)	ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/bacteria
Software and Algorithms		
BURST DNA aligner	Al-Ghalith and Knights, (2017)	doi.org/10.5281/zenodo.806850
Shi7 processing pipeline	Al-Ghalith et al., (2018)	github.com/knights-lab/shi7
FoodTree methods	This paper, (applied in Vangay et al., 2018), GitHub	github.com/knights-lab/Food_Tree
Automated Self-Administered 24-Hour Dietary Assessment Tool (ASA24-2016)	Park et al., (2018)	National Cancer Institute
Custom analysis scripts	This paper	github.com/knights-lab/dietstudy_analyses
Other		
Food and Nutrient Database for Dietary Studies 2011-2012	U.S. Department of Agriculture, Agricultural Research Service. 2014. <i>USDA Food and Nutrient Database for Dietary Studies 2011-2012</i> .	Food Surveys Research Group Home Page, http://www.ars.usda.gov/bhnrc/fsrg
Medium Chain Triglycerides	Now Foods, Bloomingdale, IL 60108, USA	N/A
Extra Virgin Olive Oil	Sea gate, Distributed by First Fishery Development Services Inc., San Diego, CA, United States of America 92123, USA	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Dan Knights (dknights@umn.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS**Human Subjects**

This study was approved by the University of Minnesota Institutional Review Board. Informed consent was obtained from all study subjects prior to any study procedures. 34 adult subjects over the age of 18, 20 female and 14 male, completed the study. Subject characteristics by gender are available in [Table 1](#). Subject characteristics by group assignment are available in [Table S1](#).

Inclusion and Exclusion Criteria

Subjects were included in the study if they self-identified as a healthy adult over 18 years old and were excluded if they had Type I diabetes mellitus or insulin dependent Type II diabetes mellitus (i.e. those at risk of developing diabetic ketoacidosis), if they were currently maintaining a ketogenic diet, if they were currently pregnant or breastfeeding, if they had used antibiotics in the last 3 months, and if they had a self-reported pre-existing history of liver disease (i.e. cirrhosis or fatty liver disease).

Sample Size Estimation

Sample size estimation was determined as follows. We performed a power analysis using MCTs as the primary exposure, and the top 10 dominant microbial taxa as the primary outcome. Using Monte Carlo simulation of microbial taxa distributed according to a lognormal distribution, and a non-parametric Mann-Whitney U test for significant differentiation, we found that a sample size of 20 subjects per group would be sufficient to detect a change of ≥ 1.24 standard deviations in any one of the dominant taxa. This analysis accounted for multiple hypothesis testing using Bonferroni correction. We aimed to enroll 40 subjects. Thirty-seven subjects began the study. Three subjects dropped out of the study during the dietary monitoring phase, before randomization to oil treatment. These subjects were not included in any analysis.

Randomization

Subjects were randomized to supplementation arms as follows: At the beginning of the supplementation phase, all currently enrolled subjects ($n=34$) were arranged into alphabetical order by first name. Treatment allocations were determined using an online random assignment calculator (graphpad.com/quickcalcs/randomize1/) to balance for equal group sizes and assigned to the subjects in alphabetical order. All subjects who were randomized to the oil intervention completed the trial.

Health Status

All subjects self-identified as healthy. Nine of the female subjects reported taking hormonal birth control, two subjects reported taking thyroid hormone, one subject with well controlled type II diabetes reported taking metformin, ranitidine, and atorvastatin, one subject reported taking lisinopril, and one subject reported taking escitalopram.

Ethnicity

Most subjects identified as non-Hispanic whites; five identified as Asian; one identified as African.

Other Details

Subjects were recruited as part of a citizen science course and from the greater University of Minnesota community. Subjects had not recently been involved in other studies. As this was a trial of dietary supplements, specifically dietary oils, subjects cannot be assumed to be test naïve as it's likely that many of these subjects consumed EVOO or MCT (from coconut oil) as part of their usual diets. Subjects were not excluded for habitual EVOO or MCT intake.

METHOD DETAILS

Experimental Design

All study data collection was completed between Jan 31, 2017 and Feb 17, 2017 in Minneapolis, MN. All participants started and ended the study on the same day. The study was designed as a double-blind, parallel-arm trial, with study subjects randomized to receive either EVOO (Sea gate, Distributed by First Fishery Development Services Inc., San Diego, CA 92123, USA) or MCT (Now Foods, Bloomingdale, IL 60108, USA) during the intervention period. The inclusion of these commercial products in this study is not meant to recommend any specific product.

The primary outcome of the supplementation phase of the study was to assess microbiome composition before and after intervention with MCT in comparison to EVOO control. Secondary outcomes included blood lipids and blood glucose measurements to characterize study subjects and assessment of the contribution of dietary intake to normal microbiome variation. Subjects were blinded to their intervention group and researchers were blinded to group assignment during data analysis. Both oils were provided in 1-gram capsules that were identical except for the color of the oil; EVOO appeared yellow in color, while MCT capsules were uncolored. Subjects were not shown both types of capsules and only ever saw the capsules they consumed.

Dietary Supplementation

On study days 10–17, subjects were randomized to receive 5% of total energy expenditure (TEE) from either olive oil or medium chain triglycerides (MCT) in 1-gram capsules. TEE was estimated using the Mifflin-St. Jeor equation from self-reported energy intake and activity level. Estimated TEE was compared to average reported caloric intake from the first 4 days of the study and adjusted up or down by increasing or decreasing the self-reported activity factor to bring the Mifflin St. Jeor estimates within 1000 kcal of reported intake. Caloric content was assumed to be 8.7 kcal/gram for MCT and 9 kcal/gram for olive oil. Oil dosage of 5% TEE was selected to balance the impact of increased energy supplied by the oils with potential gastrointestinal side effects (Mackay et al., 2017). MCT suppliers often suggest intakes ranging from 9–14 grams per day. Intakes of MCT between 50 to 100 grams have been proposed as the upper limit that should be consumed to limit gastrointestinal distress (Shah and Limketkai, 2017). Due to delivery in 1-gram capsules, intake of high doses was infeasible for this study.

Daily doses of dietary oils were provided to subjects in bags labeled with the study day. Subjects were instructed to consume the capsules throughout the day with meals and to otherwise keep their habitual diets the same, with $\frac{1}{3}$ of the supplemental pills consumed at breakfast, $\frac{1}{3}$ at lunch, and $\frac{1}{3}$ at dinner. During the intervention, subjects in the EVOO arm consumed an average of 11.9 ± 1.8 g (range 9–16 g) per day and those in the MCT arm consumed an average of 13.7 ± 3.1 g (range 10–18 g) per day.

MCT consumption was intentionally higher than EVOO due to the higher caloric content of EVOO (9.0 kcal/g v. 8.7 kcal/g). Compliance was assessed by self-report. Subjects were asked to return any unconsumed supplemental oils.

Subjects were asked to continue their habitual diet, however two subjects (subjects 11 and 12) disclosed to the researchers at the conclusion of the study that they consumed only a nutritional replacement beverage (Soylent, Rosa Foods, Inc.) throughout the study period. The dietary records from these subjects were manually checked to accurately reflect their intake of the nutritional meal replacement beverage. The nutrient content of Soylent as reported by the manufacturer was used to calculate the nutritional composition of these two subjects' dietary intake.

Dietary Record Keeping

Dietary records were entered by subjects using ASA24-2016. Subjects were instructed by a registered dietitian to record everything they had to eat and drink, including water, for the duration of the study period. Dietary record keeping began the day before the first fecal sample collection and concluded the day before the final fecal sample collection. ASA24 records were collected as dietary records, not recalls, and subjects were encouraged to record their daily intake by logging into the site as many times as necessary throughout each day. Recent comparisons of 24-hour recalls with multi-day food records and FFQs with nutritional biomarkers have concluded that a move towards multiple 24-hour recalls or multi-day food records collected consecutively may provide more accurate short-term dietary information than can currently be obtained from FFQs ([Park et al., 2018](#)).

Fecal Sample Collection and Processing

Fecal samples were collected by subjects using a stab-and-plunge method. Subjects were instructed to use FecesCatchers (Tag Hemi, Netherlands) to capture the first bowel movement from each study day. A sterile BBL CultureSwab EZ (Fisher Scientific, USA) was used to collect fecal material for two aliquots. Fecal material was placed into 2 ml cryovials pre-filled with 1 ml of 95% ethanol. Subjects were instructed to homogenize their fecal sample well within each cryovial using the CultureSwab. To account for the time it took subjects to return samples to the lab, all fecal samples were maintained in ethanol at room temperature for 10 days before freezing at -80C. Prior to freezing, samples were vortexed for approximately 15 seconds.

DNA Extraction and Sequencing

526 fecal samples were extracted and sequenced by the University of Minnesota Genomics Center. Samples were randomized to one of 6 plates prior to DNA extraction and sequencing. DNA was extracted using the PowerSoil DNA isolation kit (Mo Bio Laboratories, USA) following the manufacturer's instructions. DNA was quantified using the NanoDrop-8000 UV-Vis Spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and PicoGreen. Library prep was completed with 1/4 Nextera XT with SequalPrep prior to HiSeq 2500 Rapid Mode 100-bp single read (1x100) shallow shotgun sequencing. A subset of 72 samples, most from 1 plate, were re-sequenced because of initial low-read counts. For these samples, library prep was remade with 1/4 Nextera XT prior to MiSeq run with 150-bp single read (1x150). For the majority of these samples, this run improved read count. In all downstream analysis, sequences from second run were used instead of the first run where they were available.

Blood Collection and Processing

All subjects completing the blood draw were fasted for between 8-12 hours prior and were instructed not to smoke or exercise during the fasting period. Blood was collected into lithium heparin (gel) plasma separation tubes for measurement of glucose and lipids. Blood was collected into serum separator tubes (gel) for measurement of insulin. Serum tubes were allowed to clot for 30 minutes. All samples were spun at 2500 RPM for 10 minutes within 1 hour of the blood draw. Once separated samples were maintained at room temperature for less than 4 hours before being refrigerated. Refrigerated plasma samples were analyzed by University of Minnesota's Fairview labs within 72 hours. Serum samples were refrigerated after processing and frozen within 24 hours of collection before analysis of insulin. Insulin was assessed by chemiluminescent immunoassay, glucose by bichromatic endpoint assay, and blood lipids by bichromatic or polychromatic end points.

QUANTIFICATION AND STATISTICAL ANALYSIS

p values, sample numbers, and names of statistical tests are provided in the main text and figure legends for [Figures 2C–2H, 3A, 3B, 4A–4C, 5A–5B, and 6B–6E](#).

Shotgun Metagenomic Sequencing

All fecal samples were submitted to the University of Minnesota Genomics Center for DNA extraction, amplification and sequencing as described above. Resulting single end metagenomic shotgun reads were trimmed and processed for quality using Shi7 ([Al-Ghalith et al., 2018](#)). The quality controlled sequences were then aligned to the NCBI RefSeq representative prokaryotic genome collection (release 86) at 97% identity with BURST using default settings ([Al-Ghalith and Knights, 2017](#)).

The taxonomy table from BURST was preprocessed by dropping all samples with low read assignment counts (<23,500 reads per sample). Species level taxonomy tables were limited to those species that were present within a subject for at least 25% of the study

days, and found within 10% of the study subjects. Finally, rare species with relative abundance <0.01% were dropped, limiting the number of species to 290 annotations. The resulting taxonomy table was summarized at higher taxonomy levels (i.e. genus, family, phylum, etc.) for downstream analysis.

To correct for compositionality of microbiome sequencing data, microbiome beta-diversity was calculated using Aitchison's distance (Aitchison et al., 2000). Aitchison's distances avoids the generation of artifacts due to the fact that shotgun metagenomics provides only relative, and not absolute, abundances of species. For the majority of downstream analyses cleaned taxonomy tables were corrected for compositionality, first by imputing zeros in the taxonomy table, then using the centered log-ratio transformation (Gloor et al., 2016) using the *robCompositions* (Templ et al., 2011) package in R. Average microbiomes were determined for each subject by first normalizing taxonomy reads within each sample for each subject and averaging relative abundances within each subject. Species with mean relative abundances <0.0001 % within a subject were imputed before using the centered log-ratio transformation as described above.

To assess tree-based microbiome diversity, we applied phILR to this dataset. Because the methods to build a phylogenetic tree directly from shotgun are still unresolved, we instead aligned our shotgun sequences to the 97% GreenGenes database directly using BURST. Arguably, this is a lossy approach because we did not directly sequence the 16S region. However, if we estimate that there around 1500 bases of 16S DNA per 5-Mbp genome, then about 0.03% of the DNA should be coming from 16S in any given sequencing sample, which should give enough information to generate beta-diversity. We aligned our sequences to the GreenGenes database of 97% OTUs with BURST resulting in a median of 1230 hits per sample. Prior to applying the phILR method, OTUs that were not seen with more than 2 counts in at least 20% of the individuals in the study were filtered, as were those with a coefficient of variation less than or equal to 2. This resulted in a total of 172 OTUs remaining for diversity analysis.

Microbiome Functional Profiling

Functional annotations of KEGG modules were corrected for compositionality with the centered log-ratio adjustment. For plotting purposes (Figure 1), we selected the 20 modules with highest coefficient of variation (i.e. high SD relative to mean).

Nutrient Profile Analysis

The nutritional composition of foods was determined using ASA24-2016. ASA24 assigns nutrient information to foods using the USDA's Food and Nutrient Dietary Database (FNDDS 2011-2012). Subjects reported dietary intake as food records and entered their own dietary records directly into ASA24. For foods that were not available in the nutrient database, but consumed in high quantities, i.e. the nutritional meal replacement beverages, nutrient composition was determined from the manufacturer's website.

Food Choice Analysis

Foods were categorized according to their FNDDS food code and modification code as assigned by ASA24. The dehydrated gram weight of consumed foods and drinks was calculated using the Individual foods file from ASA24 by subtracting the moisture content from the gram weight of each food. A phenetic, hierarchical tree was created from the USDA's food code numbering system. Diversity metrics mirroring the methods used for microbiome data analysis (i.e. UniFrac for foods and Faith's diversity) were calculated from FoodTree output using QIIME 1.9.1 to calculate dietary alpha and beta-diversity.

Cleaning Data and Removal of Outliers

We ran the raw microbiome sequencing data through both aKronyMer (Al-Ghalith and Knights, 2018) and QIIME's supervised_learning.py script (Knights et al., 2011) and identified two samples (MCT.f.0086 and MCT.f.0417) with swapped labels. We confirmed that these samples were randomized to the same plate for DNA extraction and sequencing making the possibility of swapping during processing possible. We swapped the labels of these samples in the metadata for downstream analysis.

Dietary recall data were compared to the 5th and 95th percentiles of intake levels as reported by gender in NHANES for kilocalories, fat, protein, and carbohydrate as recommended by ASA24-2016. We identified 11 intake days from 8 subjects that were outside of these ranges. These reporting days were manually reviewed and two records were excluded from downstream analysis due to their incomplete or inaccurate nature. Dietary intake data and microbiome composition data from 3 subjects who dropped out of the study prior to the supplementation period were also removed from the study sample. Unless otherwise noted, all analysis of dietary intake paired with microbiome composition excluded the two subjects who primarily consumed a meal replacement beverage.

Development and Application of Tree-based Analysis for Food

First three levels of the tree structure were inferred from the FNDDS Food Code labeling scheme. Additional levels of the tree were curated manually by reviewing the Food Codes available in the ASA24 database and adding logical grouping levels to the tree. Food tree taxonomy strings are formatted with parallel structure to taxonomy strings used by QIIME for microbiome data. When a level isn't needed for food identification or grouping, it is left blank.

For example, the Levels of the Tree for a Common Food Such as Broccoli Are Denoted as L1_Vegetables;L2_Darkgreen_vegetables;L3_Darkgreen_nonleafy_vegetables;L4_Broccoli;L5_Broccoli_cooked_NS_as_to_form_NS_as_to_fat_added_in_cooking. Unlike phylogenetic trees, such as those commonly used in microbiome analysis, which use informed distances between related species, our tree of foods is phenetic and hierarchical and all distances between tree levels are the same for each food. While the uniform

branch length within the tree is a limitation of our tree, branch length distances between foods within a food group are shorter than those between foods from different food groups which allows for application of tree-based alpha and beta-diversity metrics.

Food-subject interaction network visualizations (Figure 2A) were generated using Cytoscape (Shannon et al., 2003) as described in the figure legend. Visualizations of the tree of foods (Figure 2B) were generated using Graphlan (Asnicar et al., 2015).

Decaying Weighted Average of Food Intake

We calculated the decaying weighted average for each subject's food intake as 2^{-n} for n preceding days of food records for each microbiome sample using the rollapply function from the R package zoo (Zeileis and Grothendieck, 2005). So, for example, for the microbiome sample from study day 4, the corresponding food intake profile is calculated as 100% of diet from day 3, 50% of diet from day 2, 25% of diet from day 1, and 12.5% of diet from day 0.

FDR Correction

FDR was applied at a rate of 0.2 per tested correlation (only those correlations with corrected p-values lower than 0.1 are shown in Figure 4A) and per species-level assignment (Figure S5).

Procrustes Analysis

Procrustes analysis was performed in R using the *vegan* package (Oksanen et al., 2019). Monte Carlo p-values for rotational agreement significance testing were determined from 999 permutations.

Species Prediction from Diet and Microbiome

To determine the impact of individual dietary variation on our ability to model microbiome profiles within a subject we used linear models to predict each subject's microbiome species abundance tomorrow from their microbiome principal coordinates today for the 20 most abundant species. First, we trained a linear model to predict the abundance of a species tomorrow using the first 5 principal components of a subject's microbiome today. We correlated the actual abundance tomorrow with the predicted abundance to determine the strength of the predictive fit for each species of subject A's microbiome without considering diet (Figure 5B, blue bar). Next, we calculated the microbiome residuals on a per-species basis as the difference between predicted tomorrow and actual tomorrow abundance. Then we regressed those residuals on the first 5 dietary principal components calculated from the decaying average of food intake history from unweighted UniFrac tree-based food distances for subject A. The benefit of regressing the residuals is that it decouples the microbiome coefficients from the diet coefficients so we can determine the impact of adding diet to the model when compared with the microbiome-alone prediction. Finally, we subtracted the diet predicted residuals from predicted tomorrow abundance and correlated this with actual abundance to find the species prediction model accuracy from subject A's own diet and microbiome (Figure 5B, dark purple bar). To test for a temporal association between dietary features and microbiome we repeated this process to predict the microbiome residuals but trained the model on subject A's temporally shuffled dietary features and then predicted the residual from subject A's ordered dietary features (Figure 5B, light purple bar). To test if the improvement with added diet was simply due to an increasing number of features in the model, we repeated the process described above for a new subject, subject B, selected at random from our dataset. We trained a model to predict the microbiome residuals for subject B on subject B's dietary features then predicted the residual for subject B from subject A's dietary features (Figure 5B, green bar). For clarity, the code for this predictive modeling is included as a supplementary data file, Data S1 related to Figure 5.

DATA AND SOFTWARE AVAILABILITY

Software

Processing pipelines and processed data used for the generation of figures and analysis are available at the following repository: https://github.com/knights-lab/dietstudy_analyses. Food tree methods are available here: https://github.com/knights-lab/Food_Tree

Data Resources

The shotgun metagenomic sequencing data have been deposited in the European Nucleotide Archive. The accession number for the sequencing data reported in this paper is ENA: PRJEB29065.