### Title

A posteriori dietary patterns better explain variations of the gut microbiome than individual markers in the American Gut Project




### Journal

American Journal of Clinical Nutrition, 115(2)

### ISSN

0002-9165



### Authors

Cotillard, Aurélie
Cartier-Meheust, Agnès
Litwin, Nicole S
et al.



### Publication Date

2022-02-01

### DOI

10.1093/ajcn/nqab332

Peer reviewed

See corresponding editorial on page 329.

# A posteriori dietary patterns better explain variations of the gut microbiome than individual markers in the American Gut Project

*Aurélie Cotillard,*[1] *Agnès Cartier-Meheust,*[1] *Nicole S Litwin,*[2,3] *Soline Chaumont,*[1] *Mathilde Saccareau,*[4] *Franck Lejzerowicz,*[2,3] *Julien Tap,*[1] *Hana Koutnikova,*[1] *Diana Gutierrez Lopez,*[2] *Daniel McDonald,*[3] *Se Jin Song,*[2] *Rob Knight,*[2,3,5,6] *Muriel Derrien,*[1] *and Patrick Veiga*[1]

[1]Danone Nutricia Research, Palaiseau, France; [2]Center for Microbiome Innovation, University of California San Diego, La Jolla, CA, USA; [3]Department of Pediatrics, School of Medicine, University of California San Diego, La Jolla, CA, USA; [4]Soladis, Paris, France; [5]Department of Bioengineering, University of California San Diego, La Jolla, CA, USA; and [6]Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, USA

## ABSTRACT

**Background:** Individual diet components and specific dietary regimens have been shown to impact the gut microbiome.

**Objectives:** Here, we explored the contribution of long-term diet by searching for dietary patterns that would best associate with the gut microbiome in a population-based cohort.

**Methods:** Using a priori and a posteriori approaches, we constructed dietary patterns from an FFQ completed by 1800 adults in the American Gut Project. Dietary patterns were defined as groups of participants or combinations of food variables (factors) driven by criteria ranging from individual nutrients to overall diet. We associated these patterns with 16S ribosomal RNA–based gut microbiome data for a subset of 744 participants.

**Results:** Compared to individual features (e.g., fiber and protein), or to factors representing a reduced number of dietary features, 5 a posteriori dietary patterns based on food groups were best associated with gut microbiome beta diversity ($P \leq 0.0002$). Two patterns followed Prudent-like diets—Plant-Based and Flexitarian—and exhibited the highest Healthy Eating Index 2010 (HEI-2010) scores. Two other patterns presented Western-like diets with a gradient in HEI-2010 scores. A fifth pattern consisted mostly of participants following an Exclusion diet (e.g., low carbohydrate). Notably, gut microbiome alpha diversity was significantly lower in the most Western pattern compared to the Flexitarian pattern ($P \leq 0.009$), and the Exclusion diet pattern was associated with low relative abundance of *Bifidobacterium* ($P \leq 1.2 \times 10^{-7}$), which was better explained by diet than health status.

**Conclusions:** We demonstrated that global-diet a posteriori patterns were more associated with gut microbiome variations than individual dietary features among adults in the United States. These results confirm that evaluating diet as a whole is important when studying the gut microbiome. It will also facilitate the design of more personalized dietary strategies in general populations. *Am J Clin Nutr* 2022;115:432–443.

**Keywords:** dietary patterns, gut microbiome, alpha diversity, beta diversity, American Gut Project, cohort study, food frequency questionnaire, Healthy Eating Index, 16S rRNA gene sequencing

## Introduction

Gut microbiota has emerged as a fundamental factor in human health (1). The spread of a Western lifestyle and

This work was funded by Danone Nutricia Research and supported by the Microsetta initiative.

AC-M and NSL contributed equally to this work.

Supplemental Methods, Supplemental Tables 1–8, and Supplemental Figures 1–8 are available from the "Supplementary data" link in the online posting of the article and from the same link in the online table of contents at https://academic.oup.com/ajcn/.

Address correspondence to AC (e-mail: aurelie.cotillard@danone.com) or MD (e-mail: muriel.derrien@danone.com). .

Abbreviations used: AGP, American Gut Project; ASV, amplicon sequence variant; db-RDA, distance-based redundancy analysis; DGA, Dietary Guidelines for Americans; DMM, Dirichlet Multinomial Mixture; DP5, 5 dietary patterns; ED, Exclusion diet; FL, Flexitarian diet; GI, gastrointestinal; HEI-2010, Healthy Eating Index 2010; HW, Health-Conscious Western diet; IBS, irritable bowel syndrome; MPED, MyPyramid Equivalents Database; PB, Plant-Based diet; PCoA, principal coordinate analysis; PD, phylogenetic diversity; rRNA, ribosomal RNA; SW, Standard Western diet.

Received June 29, 2021. Accepted for publication September 27, 2021.

First published online October 7, 2021; doi: https://doi.org/10.1093/ajcn/nqab332.

 

associated changes in dietary habits (e.g., increased consumption of ultra-processed foods, decreased dietary fiber, etc.), have been accompanied by an alteration of gut microbiota, thought to underlie the emergence of chronic diseases (2). As diet is a potent modulator of gut microbiota, in which cointeractions are highly personalized (3), studying the wide variation of dietary habits in the general populations is key to identifying dietary components that are best associated with the gut microbiota and human health. Notably, dietary fibers and vegetal proteins are of great relevance in regard to current recommendations and interest in plant-based diets for human health and environmental sustainability (4). However individual dietary components may not be the most suited for personalized nutrition, as the current dietary recommendations have shifted from individual foods/nutrients to eating patterns as a whole, which should be adapted to current eating habits to promote greater adherence (5, 6). As of today, there is no standard method to assess overall diet, and multiple approaches are currently used to examine dietary patterns. The 2 most common approaches include an a priori analysis, which relies on existing nutritional knowledge or evidence-based diet-health relationships (7), and an a posteriori analysis, which is a purely data-driven, exploratory approach deriving common food consumption patterns within a given population (8). Numerous cross-sectional studies have explored the relationship between diet and gut microbiota, most relying on a priori approaches using predefined diet quality scores (e.g., Healthy Eating Index, Mediterranean Diet Score) (9, 10). Recent large, population-based cohorts, exceeding thousands of subjects, have identified multiple associations between the gut microbiome and overall dietary quality (11, 12), specific dietary components (11, 13–17), and clinical outcomes (11, 18). However, most of these large studies have also typically relied on a priori–defined indices to characterize diet. Some studies have combined a posteriori dietary analysis with the study of the gut microbiome (9, 18–21), but studies that compare all of these approaches are lacking.

In this study, we explored the long-term dietary intake of adult participants from the American Gut Project using both a priori and a posteriori approaches, and identified dietary patterns that were the most associated with the gut microbiome.

## Methods

### Participant recruitment

In order to investigate the associations between long-term diet and the gut microbiome, we performed a retrospective analysis of the American Gut Project (AGP) cohort (16). This project provides access to human microbiome data and a variety of other information, including demographic, lifestyle, and dietary data (metadata) on a cohort of citizen-scientists. The entire AGP data set was subsetted using the metadata version accessed on 22 October 2019 for all samples (e.g., stool, hands, mouth) from adult participants (age $\geq$18 years) who reside in the United States ($N$ = 10,085). Given the self-reported nature of the data, we performed some basic curation steps before analysis (see **Supplemental Methods**). Participants' consent was obtained and research was conducted under the University of Colorado's Institutional Review Board and the University of California San Diego's Human Research Protection Program protocols (numbers 12-0582 and 141853, respectively).

### Dietary intake assessment

Informal dietary questions (e.g., "in an average week, how often do you consume meat/eggs?") from the AGP general questionnaire were available. In addition, long-term dietary intake was evaluated in a subset of 1948 participants who completed the VioScreen FFQ (version 4; VioCare) between July 2012 and June 2019. This web-based, graphical, and self-administered dietary assessment tool has been previously validated for use in general US adult populations (22) and utilizes the Nutrition Data System for Research (Nutrition Coordinating Center, University of Minnesota, Minneapolis, MN) to estimate nutritional compositions from VioScreen FFQ entries. Several types of dietary information were derived from raw VioScreen entries, including: custom food items and food groups (described in Supplemental Methods and **Supplemental Table 1**), total energy intake, micro- and macronutrient intakes, and MyPyramid Equivalents Database (MPED) values (23). Overall diet quality was assessed using the Healthy Eating Index 2010 (HEI-2010) (24), which was available in the database. In addition, to ensure a high level of quality for the dietary intake data, we applied several filtering criteria to the VioScreen FFQ reports. We included reports in analyses if they met the following criteria: *1*) a minimum of 25 food items consumed (22); *2*) a total energy intake between the 5th and 95th gender-specific percentiles from NHANES, as proposed previously (3, 25); and *3*) completion of the questionnaire in a window of time between 10 and 20,000 minutes (around 2 weeks).

### Dietary pattern analysis

The term dietary pattern is often used to describe the set of dietary behaviors that characterizes a group or subset of study participants. The same term can also refer to quantitative scores originating from a combination of several diet variables, called factors (26). In this study, we use the term dietary patterns to describe both.

We analyzed long-term diets based on VioScreen FFQ data using both a priori and a posteriori analyses. First, we defined a priori patterns based on 3 diet features: 2 individual dietary components (total dietary fibers and proteins) and HEI-2010 scores. We derived diet groups as quartiles based on total dietary fiber or total protein intake. For fibers, we analyzed the daily total dietary fiber intake (g/d) and their type (defined as the ratio g/d of soluble fiber:g/d of insoluble fiber), and created a combined indicator (dietary fiber quantity:dietary fiber type), which can simply be interpreted as fiber quantity normalized by fiber type. For protein, we focused on the total daily intake of protein of both animal and vegetal origins, as well as their ratio (g/d of animal protein:g/d of vegetable protein). Moreover, as an a priori factor to evaluate overall diet quality, we used the HEI-2010 total score, a composite score based on 12 components that measures adherence to the 2010 US Dietary Guidelines for Americans (DGA) (24).

In addition, we applied factor analysis to identify nonredundant reduced sets of variables (i.e., factors), that are linear combinations of the original variables and explain most of the variability in the data set. We identified factors based on MPED component values, micronutrients, and food items. In the case of MPED values and micronutrients, we first normalized variables by total energy intake and filtered those that were considered redundant (Pearson correlation above 0.95). We then

used a principal component analysis with varimax rotation [psych v2.0.9 R package (27)] and chose the number of factors by parallel analysis of the correlation matrix [nFactors v2.4.1 R package (28)]. Given the sparsity (69%) of the data, we built factors for food items using a principal coordinate analysis (PCoA) on the Jaccard distance between participants (computed from presence/absence information; Supplemental Table 1). We selected the number of factors by the Kaiser rule, which, in brief, retains factors with eigenvalues greater than 1 [nFactors v2.4.1 R package (28)].

Finally, we built a posteriori diet groups from food group data expressed in kcal/d (Supplemental Table 1). We used Dirichlet Multinomial Mixture (DMM) models [DirichletMultinomial v1.30.0 R package (29)], which apply a compositional approach for clustering (i.e., based on the relative kcal/d contribution; see Supplemental Methods). We defined our diet groups based on a robustness criteria (**Supplemental Figure 1**) and chose the largest number of groups that allowed a mean stability above 0.6 in each of them [using 50 bootstrap and Jaccard similarity with the fpc v2.2–8 R package (30)]. Moreover, to increase confidence in the group categorization, we defined core diet groups as participants having a probability above 0.8 of belonging to a particular group.

### Stool sample processing and 16S ribosomal RNA gene sequencing

In the AGP initiative, stool samples were collected at home and shipped at room temperature before microbial DNA extraction and 16S ribosomal RNA (rRNA) amplicon sequencing, which were performed as previously described (16). We used Redbiom (31) to fetch Deblur (32) feature tables from the Qiita platform (33), and extracted 20,454 stool samples on 5 December 2019 from the Deblur-Illumina-16S-V4-100nt-fbc5b2 context. In cases where a fecal sample was sequenced multiple times, the sequencing run with the most reads was kept. We used 100 nucleotides in order to be inclusive of AGP sequencing runs performed at 125 cycles.

We performed the following additional bioinformatic steps using QIIME 2 (v2019.10) (34). Bloom sequences, as identified in Amir et al. (35), were removed, and samples with fewer than 1000 reads left after filtering for blooms were excluded. We then rarefied the obtained amplicon sequence variant (ASV) table to a depth of 1000 sequences per sample in order to compute several alpha (within-subjects) and beta (between-subjects) diversity indices (Supplemental Methods). Subsequently, we performed taxonomic assignment for ASVs on nonrarefied data using the sklearn-based taxonomy classifier trained on the Greengenes reference database 13_8 (36), and we aggregated data at the genus level for further analyses. We excluded 1580 samples flagged as outliers (Supplemental Methods). In cases where participants had submitted multiple fecal samples, we selected a unique sample based on 16S data availability, metadata completion, and the shortest time between sample collection and VioScreen FFQ completion date.

### Gut microbiome statistical analysis

We looked for associations between dietary patterns and 16S rRNA gene-based gut microbiome profiles in a subset of 744 participants (**Figure 1**A) using multiple alpha- and beta-diversity assessments (Supplemental Methods), as well as taxonomic composition (genera abundances). Age, sex, and BMI were used as confounding variables. We linearly adjusted alpha-diversity indices by these confounding variables and compared them between diet groups with Kruskal-Wallis tests or associated them with factors using Spearman correlations. For diet groups, we performed post hoc comparisons using Mann-Whitney tests. We also looked for associations between beta-diversity indices and diet groups or factors using partial distance-based redundancy analysis (db-RDA), which allowed us to remove the effect of the selected confounding variables. In addition, for diet groups, we verified that the observed differences were not due to dispersion heterogeneity through beta-dispersion tests. All beta-diversity analyses were performed using the vegan v2.5–6 R package (37). Finally, we filtered out 409 genera with a low prevalence (present in less than 10% of samples) or low relative abundance (mean below 0.01%), and analyzed the 111 remaining genera for differential abundance. We used DESeq2 (v1.28.1) models with the poscounts option (38), including confounding variables and diet groups or factors effects. We estimated the global diet groups effects by likelihood ratio tests and the factor effects, as well as diet groups 2 by 2 comparisons, by Wald tests. Log2 fold change results are expressed as the estimate ± SE. In addition, we evaluated the robustness of our results using Songbird (39), an alternative statistical method often used in the microbiome field to account for data compositionality (Supplemental Methods). As a complementary analysis, we predicted *Bifidobacterium* low or high relative abundance from a subset of metadata using random forests (Supplemental Methods).

### General statistics

Data are expressed as medians (IQRs) for quantitative variables and as *n* (%) for qualitative variables. We used Kruskal-Wallis and chi-squared independence tests to compare quantitative and qualitative variables, respectively, between groups with the compareGroups v4.4.5 R package (40).

We handled multiple testing adjustments with the Benjamini-Hochberg procedure (41) (see details in table legends and figure captions) and we used a 2-sided 5% alpha error threshold. All analyses were performed in R v4.0.3 (42) except Songbird, which was run with Python 2.6.

## Results

### Description of the study cohort

After applying the FFQ quality filtering criteria, a final sample set of 1800 US adult participants was included in the dietary pattern analysis (Figure 1A). Participant characteristics are detailed in **Table 1**. Compared to the general US adult population (NHANES data), this AGP adult cohort was older, with a higher proportion of women, a higher level of education, and a lower BMI. Additionally, this cohort had lower prevalences of diabetes (3.3% compared with 9.5%, respectively) and cardiovascular diseases (4.5% compared with 11.2%, respectively) compared to what was reported among US adults in the 2018 National Health Interview Survey (43). However, the population under study may be enriched in gastrointestinal (GI) disorders, as a higher prevalence of inflammatory bowel disease was self-reported (3.6% compared with 1.3% in the general population) (44). The

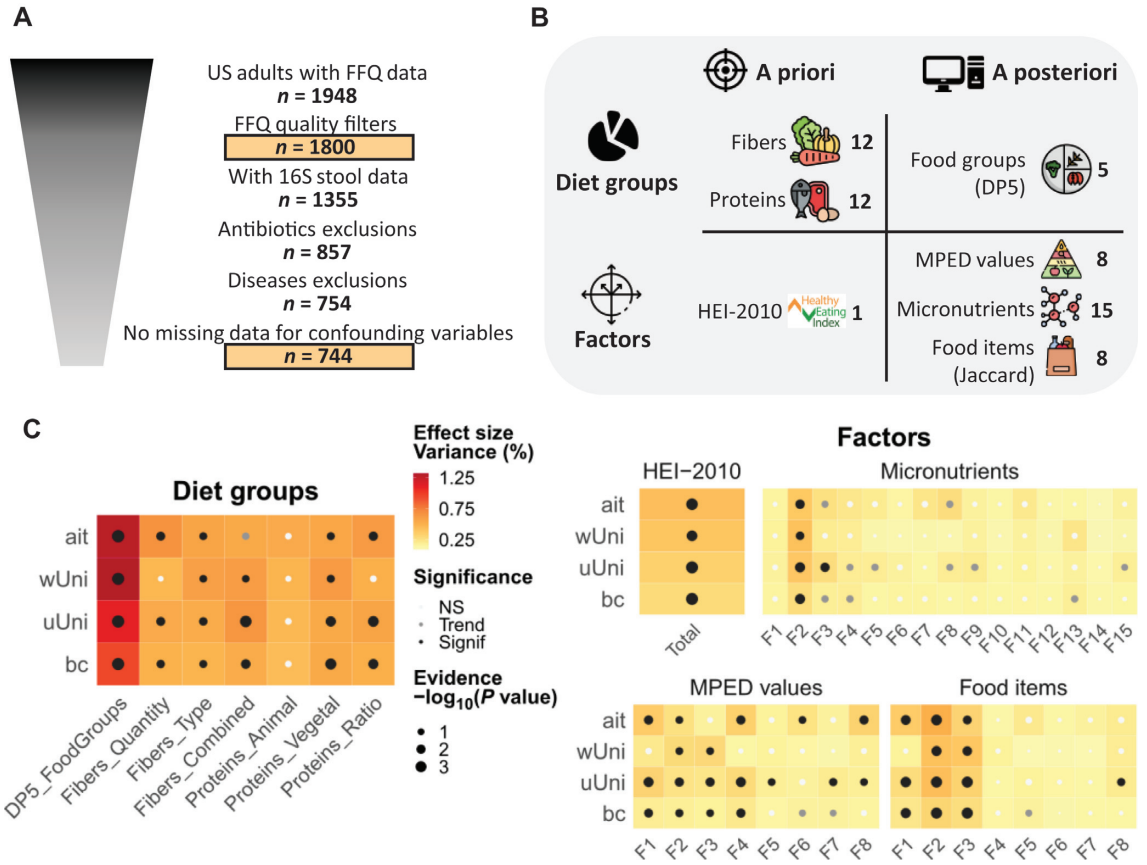


**FIGURE 1** Summary of the exploration of dietary patterns and their associations with gut microbiome beta-diversity. (A) Population selection. Populations used for main analyses (dietary patterns and associations with microbiome) are surrounded by boxes. Some outliers were removed for 16S rRNA-based microbiome data (Supplemental Methods). No antibiotics were taken in the last year. No cases of diabetes, liver disease, or IBD were diagnosed by a medical practitioner. There were no declared cases of multiple sclerosis, Hashimoto's, Graves', Behcet's, Lupus, hyperthyroidism, or chronic Lyme disease. Confounding variables were age, sex, and BMI. (B) Numbers of dietary patterns obtained with a priori or a posteriori approaches counting either diet groups or factors (61 in total). For dietary fibers, there are quartiles of quantity, type (soluble:insoluble) and combined quantity and type (quantity:type). For dietary proteins, there are quartiles of animal and vegetable proteins, as well as their ratio. The food groups analysis was based on data in kcal and focused on core diet groups: that is, participants with a probability ≥80% of belonging to his/her partition, referred to as DP5 patterns. The food items analysis was based on the presence/absence of data using the Jaccard distance. This image has been designed using resources from Flaticon.com made by Freepik, Good Ware, Eucalyp, DinosoftLabs, iconixar and surang. (C) The 16S rRNA-based gut microbiome beta-diversity analyses. Partial db-RDA models with diet groups/factors as explanatory variable and confounding variables (age, sex, and BMI) partialled out. We used permutation tests (9999 permutations). There was multiple testing adjustment by Benjamini-Hochberg on the global effects obtained with the 4 indices (diet groups) or on the global effects obtained with the 5 indices * k factors (factors). Evidence [–log10(*P* value)] cannot be higher than 4 due to the permutation scheme. The Fk factors are from the corresponding data set as described in Supplemental Tables 4–6. NS: *P* value ≥ 0.1; trend: 0.05 ≤ *P* value < 0.1; significance: *P* value < 0.05. Abbreviations: ait, Aitchison distance; bc, Bray-Curtis dissimilarity; db-RDA, distance-based redundancy analysis; DP5, 5 dietary patterns; Fk, factor number k; HEI-2010, Healthy Eating Index 2010; IBD, inflammatory bowel disease; MPED, MyPyramid Equivalents Database; NS, not significant; rRNA, ribosomal RNA; uUni, unweighted UniFrac distance; wUni, weighted UniFrac distance.

overall diet quality in this cohort, as assessed by the HEI-2010 total score, appeared to be high, with a median of 72.2 (IQR, 64.4–78.8). Of note, only 4% of the participants had an HEI-2010 score <50 (indicative of poor diet quality). In addition, while men and women showed expected differences in terms of health and diet (**Supplemental Table 2**), they had similar HEI-2010 scores and the total calorie consumptions for both sexes were low compared to the current DGA recommendations (5). This suggests that the majority of participants in this cohort reported a diet of high nutritional quality, as well as high adherence to the US DGA.

### Dietary patterns most associated with gut microbiome

To identify the dietary patterns that were most associated with the gut microbiome in this cohort, we utilized both a priori and a posteriori approaches and constructed patterns in the form of diet groups or factors (Figure 1B). We first derived a priori diet groups from individual dietary components (fibers and proteins; **Supplemental Table 3**), and then looked for a more global representation of habitual diet through combinations of food variables. To this end, we used the HEI-2010 total score, which combines both food groups and individual nutrients. A posteriori factors identified from MPED components, micronutrients, and food items led to 3 sets of 8, 15, and 8 factors explaining 54%, 75%, and 22%, respectively, of the variance in their data set (**Supplemental Tables 4–6**; **Supplemental Figure 2**). Last, our cluster-based global-diet analysis on food groups data resulted in 5 diet groups, further refined into 5 core diet groups including only participants with the highest probability of belonging to their group. These core diet groups are hereon denoted as DP5 (5 dietary patterns), and were defined for 1466

**TABLE 1** Description of the Study Cohort

| | Study cohort, $n = 1800$ | Microbiome cohort, $n = 744$ | NHANES, $n = 14,584$ |
|---|---|---|---|
| Demography & lifestyle | | | |
| Age, years | 53.0 [41.0–63.0] | 52.0 [41.0–62.0] | 45.0 [31.0–59.0] |
| Sex, female | 1166 (64.8%) | 462 (62.1%) | 7639 (52.4%) |
| Education, graduate | 1076 (60.1%) | 456 (61.5%) | 4108 (28.2%) |
| Alcohol frequency, regularly[1] | 547 (30.6%) | 235 (31.6%) | — |
| Health | | | |
| BMI | 23.7 [21.5–26.6] | 23.4 [21.2–25.8] | 27.5 [24.0–32.0] |
| Diabetes | 59 (3.31%) | 0 (0%) | — |
| CVD | 80 (4.50%) | 23 (3.10%) | — |
| Autoimmune disease | 256 (14.4%) | 53 (7.15%) | — |
| IBD | 63 (3.58%) | 0 (0%) | — |
| IBS | 250 (14.2%) | 70 (9.46%) | — |
| Gluten intolerance | 435 (24.8%) | 166 (22.6%) | — |
| Lactose intolerance | 342 (19.5%) | 137 (18.7%) | — |
| Bowel movement, normal[1] | 1288 (74.4%) | 582 (80.3%) | — |
| Diet–AGP Questionnaire | | | |
| Diet type, vegetarian[1] | 136 (7.67%) | 73 (9.92%) | — |
| Plant diversity, more than 20[1] | 455 (35.6%) | 213 (41.8%) | — |
| Vegetable frequency, regularly[1] | 1603 (89.9%) | 683 (92.2%) | — |
| Fruit frequency, regularly[1] | 1130 (63.5%) | 477 (64.5%) | — |
| Whole-grain frequency, regularly[1] | 836 (47.1%) | 368 (50.0%) | — |
| Red meat frequency, regularly[1] | 384 (21.5%) | 171 (23.0%) | — |
| Milk & cheese frequency, regularly[1] | 835 (46.7%) | 339 (45.8%) | — |
| SSB, regularly[1] | 54 (3.03%) | 220 (29.7%) | — |
| Diet–FFQ | | | |
| Total energy intake, kcal/d | 1772 [1399–2275] | 1766 [1423–2271] | — |
| Carbohydrates, % of calories | 41.2 [33.3–48.0] | 41.0 [33.2–47.8] | — |
| Fats, % of calories | 38.0 [31.8–45.4] | 38.4 [31.5–46.1] | — |
| Protein, % of calories | 15.6 [13.4–18.0] | 15.4 [13.3–17.8] | — |
| HEI-2010 | 72.2 [64.4–78.8] | 72.9 [65.7–79.3] | — |

Descriptions of the US adult participants who were analyzed for dietary patterns (study cohort) and who were analyzed for gut microbiome associations (microbiome cohort). Data are presented as medians [IQRs] for quantitative variables and as *n* (%) for qualitative variables. Missing values were excluded for the percentage calculation. Data from NHANES adults are given as a proxy for the general US population (Supplemental Methods). Participants with antibiotic intake in the last year, declared IBD, liver diseases, diabetes, or specific autoimmune diseases were excluded from the microbiome cohort to minimize confounding effects. No statistical test was performed. Abbreviations: AGP, American Gut Project; CVD, cardiovascular disease; HEI-2010, Healthy Eating Index 2010; IBD, inflammatory bowel disease; IBS, irritable bowel syndrome; SSB, sugar-sweetened beverages.

[1]Only 1 representative modality is shown for compactness reasons.

participants out of 1800. In total, we obtained 61 dietary patterns (Figure 1B), ranging from individual components to a global diet, that we associated with 16S rRNA-based gut microbiome data for a subset of 744 participants (620 for DP5; Figure 1A; Table 1).

Diet groups explained higher amounts of gut microbiome beta-diversity variation than factors using 4 different metrics, with a maximum value of 1.3% (Figure 1C), while factors were prone to more significant associations with gut microbiome beta diversity, likely due to the statistical power of quantitative analyses. Among the diet groups identified based on individual dietary components and factors representing a reduced number of nonredundant food variables, the most significant associations were obtained for *1*) the second ($P \leq 0.0008$) and third ($P \leq 0.0043$) factors of food items, both driven positively by vegan foods and negatively by meat and fats; *2*) the HEI-2010 total score ($P \leq 0.0014$); and *3*) the second factor of micronutrients ($P \leq 0.0016$), mostly driven by artificial sweeteners, fruits, vegetables, and other foods rich in fibers (Supplemental Tables 5 and 6; Supplemental Figure 2). Amongst all the tested dietary patterns, the gut microbiome composition was best associated with the DP5 diet groups obtained from the global-diet food group analysis. Indeed, they exhibited the highest significance ($P \leq 0.0002$) and the highest percentage of variance explained for beta-diversity metrics (Figure 1C), even after extending the analysis to noncore DP5 diet groups (**Supplemental Figure 3**A). In addition, the DP5 diet groups were the only patterns significantly associated with alpha diversity, and they showed the highest statistical evidence and effect size among diet groups (Supplemental Figure 3B and C). Overall, our results show that the global composition of the gut microbiome was best explained by the DP5, which were retained for deeper characterization.

## Characterization of the DP5

We identified 5 dietary patterns based on energy-adjusted food groups. Two of the patterns were most similar to a previously described Prudent diet (45), a second set of 2 patterns were most similar to Western-style diets, and the last to an Exclusion diet. An additional analysis of the data via PCoA confirmed that the Prudent patterns were similar to each other, while the 2 Western-style patterns exhibited the greatest

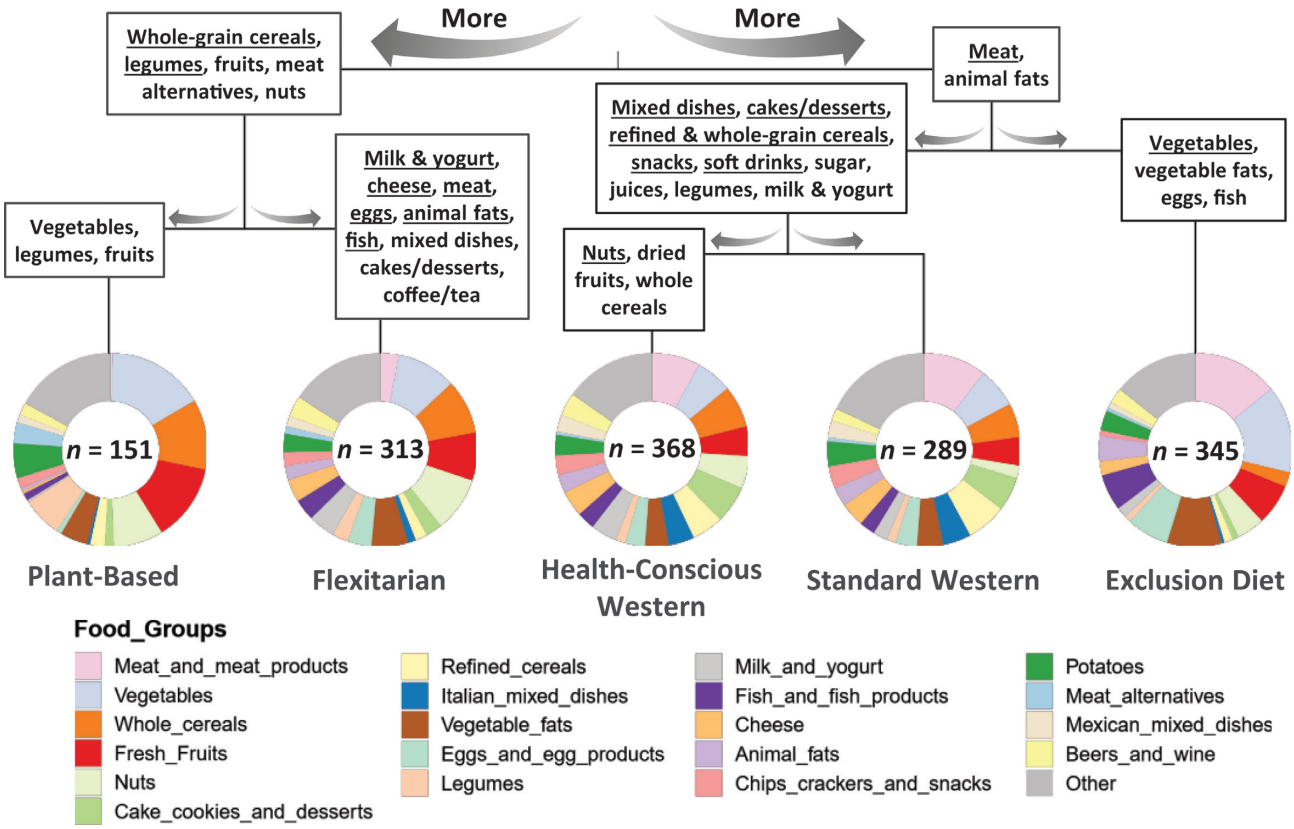


**FIGURE 2** Contribution of food groups to DP5 patterns. Pie charts represent Dirichlet scaled contributions of food groups for each dietary pattern. Food groups are ordered by their contribution to the clustering. Patterns are grouped together using a hierarchical ascending clustering on standardized data. Food groups were mentioned in each branch if their Dirichlet contribution was higher in all left/right patterns compared with right/left patterns and if their Cliff's Delta effect size was medium (≥0.33; bold) or large (≥0.47; bold underlined). Cliff's Delta effect sizes were computed based on individual relative kcal intakes. *n* is the size of the pattern. Abbreviations: DP5, 5 dietary patterns.

amount of overlap in dietary intake (**Supplemental Figure 4**). In addition, the Exclusion diet pattern appears to contain the most distinct and the most variable dietary habits of the 5 eating patterns.

Participants categorized in the Prudent-like dietary patterns presented lower consumption of animal products and higher consumption of fruits, whole-grain cereals, legumes, and nuts (**Figure 2**). One of them, referred to as the Plant-Based diet (PB), consisted of almost no meat, very few animal products, few cakes and desserts, and a high quantity of fruits, vegetables, and whole-grain cereals. Predictably, more participants in this diet group declared themselves as vegetarian or vegan compared to those in other groups. Consistently, they reported consuming a higher variety of plants and were associated with the highest dietary fiber intake (**Table 2**). Moreover, for the 12 HEI-2010 subcomponents, the PB dietary pattern presented the best fit to the DGA for total fruit, whole fruit, total vegetables, greens and beans, whole grains, fatty acids, and empty calories compared to the other groups (**Supplemental Figure 5**A). Similar to the PB group, the second pattern, denoted as a Flexitarian diet (FL), consisted of high amounts of fruits, whole-grain cereals, and nuts, but, conversely, low but not null amounts of meat and high amounts of dairy products. Interestingly, participants categorized in this FL pattern were slightly older compared to individuals with more meat-based diets and presented the best overall diet quality as measured by total HEI-2010 score (Table 2; Supplemental Figure 5B).

Participants in the 2 patterns identified with a Western-like diet reported higher consumption of mixed dishes and sugar-sweetened products, including beverages and refined cereals, and lower consumption of vegetables (Figure 2). They also declared higher ready-to-eat meal frequency (e.g., boxed macaroni and cheese, ramen noodles, frozen meals) and had higher BMIs (Table 2). One of these patterns comprised individuals with higher nut and whole-grain cereal consumption, who also appeared to have a more diverse diet compared to the other Western-like pattern [27 (IQR, 26–28) compared with 23 (IQR, 20–24) food groups and 81 (IQR, 72–93) compared with 58 (IQR, 48–68) food items]. We denoted this particular pattern as a Health-Conscious Western diet (HW), which included the highest consumption of sweets, red wine, and dairy products compared to all other groups. The second Western-like pattern, referred to as a Standard Western diet (SW), showed the poorest diet quality, having the lowest HEI-2010 score (Table 2; Supplemental Figure 5B). The SW group presented the lowest fiber intake and variety of plants consumed, as well as the highest consumption of sugar-sweetened beverages. While the HW pattern presented the highest energy intake, it was not associated with a higher BMI value compared to the SW diet group. This may be partly explained by a higher reported exercise frequency (Table 2).

**TABLE 2** DP5 Patterns Characterization

| | Plant-Based, $n = 151$ | Flexitarian, $n = 313$ | Health-Conscious Western, $n = 368$ | Standard Western, $n = 289$ | Exclusion Diet, $n = 345$ | Adjusted $P$ value |
|---|---|---|---|---|---|---|
| Demography | | | | | | |
| Age, years | 55.0 [42.5–64.0]$^{ab}$ | 57.0 [44.0–65.0]$^{a}$ | 52.0 [40.0–62.0]$^{bc}$ | 48.0 [38.0–61.0]$^{c}$ | 53.0 [43.0–63.0]$^{b}$ | <0.001 |
| Sex, female | 92 (60.9%)$^{ab}$ | 228 (72.8%)$^{c}$ | 216 (58.7%)$^{b}$ | 174 (60.2%)$^{ab}$ | 238 (69.0%)$^{ac}$ | <0.001 |
| Education, graduate | 93 (62.4%)$^{ab}$ | 202 (65.2%)$^{a}$ | 235 (64.0%)$^{a}$ | 143 (49.8%)$^{b}$ | 198 (57.7%)$^{ab}$ | <0.001 |
| Lifestyle | | | | | | |
| Exercise frequency, regularly[1] | 108 (71.5%)$^{a}$ | 241 (77.5%)$^{a}$ | 220 (59.9%)$^{b}$ | 130 (45.6%)$^{c}$ | 247 (71.8%)$^{a}$ | <0.001 |
| Ready-to-eat meals frequency, regularly[1] | 3 (2.00%)$^{ab}$ | 7 (2.24%)$^{a}$ | 18 (4.92%)$^{c}$ | 22 (7.72%)$^{c}$ | 3 (0.88%)$^{b}$ | <0.001 |
| Alcohol frequency, regularly[1] | 27 (17.9%)$^{a}$ | 116 (37.2%)$^{b}$ | 152 (41.6%)$^{c}$ | 63 (22.1%)$^{a}$ | 79 (23.0%)$^{a}$ | <0.001 |
| Red wine | 58 (38.4%)$^{a}$ | 201 (64.2%)$^{b}$ | 283 (76.9%)$^{d}$ | 109 (37.7%)$^{a}$ | 172 (49.9%)$^{c}$ | <0.001 |
| Health | | | | | | |
| BMI | 22.3 [20.7–24.4]$^{a}$ | 22.9 [21.3–25.1]$^{a}$ | 24.8 [22.1–27.7]$^{b}$ | 25.0 [22.1–29.0]$^{b}$ | 23.1 [20.9–25.5]$^{a}$ | <0.001 |
| Autoimmune disease | 21 (14.1%)$^{ab}$ | 33 (10.6%)$^{a}$ | 44 (12.1%)$^{a}$ | 43 (15.1%)$^{ab}$ | 70 (20.8%)$^{b}$ | 0.003 |
| IBS | 10 (6.80%)$^{a}$ | 40 (12.8%)$^{ab}$ | 35 (9.75%)$^{ab}$ | 44 (15.5%)$^{bc}$ | 74 (22.1%)$^{c}$ | <0.001 |
| Fungal overgrowth | 6 (4.08%)$^{ab}$ | 13 (4.29%)$^{ab}$ | 8 (2.25%)$^{b}$ | 17 (6.03%)$^{ab}$ | 31 (9.23%)$^{a}$ | 0.001 |
| SIBO | 5 (3.40%)$^{abc}$ | 5 (1.64%)$^{ab}$ | 4 (1.13%)$^{a}$ | 13 (4.63%)$^{bc}$ | 22 (6.61%)$^{c}$ | <0.001 |
| Liver disease | 5 (3.38%)$^{a}$ | 5 (1.61%)$^{a}$ | 3 (0.82%)$^{a}$ | 12 (4.18%)$^{a}$ | 4 (1.18%)$^{a}$ | 0.013 |
| Thyroid disease | 17 (11.5%)$^{ab}$ | 49 (15.9%)$^{a}$ | 32 (8.89%)$^{b}$ | 42 (14.7%)$^{ab}$ | 71 (21.1%)$^{a}$ | <0.001 |
| Gluten-intolerance | 40 (27.2%)$^{a}$ | 48 (15.7%)$^{bc}$ | 41 (11.2%)$^{b}$ | 53 (19.1%)$^{ac}$ | 181 (54.7%)$^{d}$ | <0.001 |
| Diet–AGP Questionnaire | | | | | | |
| Diet type, vegetarian[1] | 96 (64.0%)$^{a}$ | 23 (7.40%)$^{b}$ | 1 (0.27%)$^{c}$ | 7 (2.47%)$^{d}$ | 1 (0.29%)$^{c}$ | <0.001 |
| Specialized diet, exclude dairy | 45 (58.4%)$^{a}$ | 7 (4.86%)$^{b}$ | 2 (1.18%)$^{b}$ | 4 (2.65%)$^{b}$ | 42 (30.2%)$^{c}$ | <0.001 |
| Specialized diet, exclude refined sugars | 26 (33.8%)$^{a}$ | 19 (13.2%)$^{b}$ | 7 (4.12%)$^{d}$ | 8 (5.30%)$^{d}$ | 72 (51.8%)$^{c}$ | <0.001 |
| Specialized diet, modified paleo diet | 1 (1.30%)$^{a}$ | 12 (8.33%)$^{a}$ | 7 (4.12%)$^{a}$ | 8 (5.30%)$^{a}$ | 56 (40.3%)$^{b}$ | <0.001 |
| Plant diversity, more than 201 | 72 (64.9%)$^{a}$ | 107 (47.8%)$^{b}$ | 99 (37.8%)$^{b}$ | 41 (19.0%)$^{d}$ | 73 (32.2%)$^{c}$ | <0.001 |
| Whole-grain frequency, regularly[1] | 92 (62.2%)$^{ab}$ | 211 (67.8%)$^{a}$ | 225 (62.2%)$^{b}$ | 102 (35.9%)$^{d}$ | 58 (17.0%)$^{c}$ | <0.001 |
| Meat & eggs frequency, regularly[1] | 17 (11.3%)$^{a}$ | 195 (62.7%)$^{b}$ | 318 (86.9%)$^{cd}$ | 234 (81.5%)$^{d}$ | 309 (90.1%)$^{c}$ | <0.001 |
| Milk & cheese frequency, regularly[1] | 12 (8.00%)$^{a}$ | 166 (53.4%)$^{b}$ | 238 (65.4%)$^{d}$ | 141 (49.3%)$^{b}$ | 118 (34.3%)$^{c}$ | <0.001 |
| Sugary sweets frequency, regularly[1] | 26 (17.4%)$^{a}$ | 94 (30.1%)$^{b}$ | 181 (50.0%)$^{d}$ | 125 (43.9%)$^{e}$ | 32 (9.36%)$^{c}$ | <0.001 |
| SSB frequency, regularly[1] | 3 (2.01%)$^{a}$ | 2 (0.64%)$^{a}$ | 12 (3.29%)$^{c}$ | 29 (10.2%)$^{d}$ | 1 (0.29%)$^{b}$ | <0.001 |
| Diet–FFQ | | | | | | |
| Total energy intake, kcal/d | 1593 [1279–1969]$^{a}$ | 1716 [1395–2223]$^{b}$ | 2094 [1693–2607]$^{c}$ | 1869 [1399–2367]$^{b}$ | 1570 [1227–2007]$^{a}$ | <0.001 |
| Total water intake, g/d | 1066 [592–1599]$^{a}$ | 1066 [592–1599]$^{a}$ | 1066 [592–1599]$^{a}$ | 1066 [592–1599]$^{a}$ | 1422 [888–2133]$^{b}$ | <0.001 |
| Carbohydrates, % of calories | 54.8 [44.6–61.1]$^{a}$ | 43.6 [37.5–49.5]$^{b}$ | 42.8 [37.4–46.9]$^{b}$ | 42.6 [35.2–49.6]$^{b}$ | 28.4 [20.7–37.5]$^{c}$ | <0.001 |
| Fats, % of calories | 28.0 [20.6–38.7]$^{a}$ | 36.4 [31.8–42.3]$^{b}$ | 37.0 [33.0–41.7]$^{b}$ | 36.8 [30.7–42.3]$^{b}$ | 49.9 [38.5–58.0]$^{c}$ | <0.001 |
| Protein, % of calories | 12.5 [11.3–14.4]$^{a}$ | 14.7 [12.6–17.0]$^{b}$ | 15.4 [13.7–17.1]$^{d}$ | 15.8 [13.5–18.1]$^{d}$ | 17.9 [15.2–20.5]$^{c}$ | <0.001 |
| Added sugar, % of calories | 1.11 [0.75–1.49]$^{a}$ | 1.38 [1.00–1.89]$^{b}$ | 1.58 [1.19–2.03]$^{d}$ | 1.77 [0.99–2.56]$^{d}$ | 0.78 [0.46–1.22]$^{c}$ | <0.001 |
| Total fiber intake, g/d | 38.8 [28.6–50.3]$^{a}$ | 27.7 [21.3–35.5]$^{b}$ | 25.6 [19.8–30.9]$^{d}$ | 19.7 [13.7–25.7]$^{e}$ | 21.5 [15.5–28.3]$^{c}$ | <0.001 |
| Vegetable proteins, g/d | 45.2 [31.3–59.1]$^{a}$ | 32.1 [23.9–42.2]$^{b}$ | 30.7 [24.2–40.2]$^{b}$ | 22.8 [16.8–33.2]$^{d}$ | 18.9 [13.4–27.2]$^{c}$ | <0.001 |
| Animal proteins, g/d | 5.83 [2.06–14.0]$^{a}$ | 32.9 [21.2–45.8]$^{b}$ | 49.8 [38.7–63.8]$^{c}$ | 48.5 [31.5–68.9]$^{c}$ | 50.2 [34.7–69.3]$^{c}$ | <0.001 |
| HEI-2010 | 77.8 [74.2–81.0]$^{a}$ | 80.0 [74.0–83.4]$^{b}$ | 72.4 [67.2–77.4]$^{d}$ | 62.9 [55.1–70.7]$^{e}$ | 68.6 [60.4–73.0]$^{c}$ | <0.001 |

Characterization of the DP5 patterns compared with metadata of the general AGP questionnaire and compared with some general information from the FFQ. Only a selection of results is shown here. Full results are provided in Supplemental Table 8. Data are presented as median [IQR] for quantitative variables and as $n$ (%) for qualitative variables. Missing values were excluded for the percentage calculation. Kruskal-Wallis tests were used to compare quantitative variables between patterns and Chi-squared tests were used for qualitative variables. $P$ values for the global pattern effects were adjusted for multiple testing with the Benjamini-Hochberg procedure. Significant $P$ values (<0.05) in bold. If the pattern effect was significant, all 2 by 2 comparisons were reported using a Benjamini-Hochberg adjustment for each parameter separately. Groups with the same letter are not significantly different ($P$ value $\geq$ 0.05). Abbreviations: AGP, American Gut Project; DP5: 5 dietary patterns; HEI-2010, Healthy Eating Index 2010; IBS, irritable bowel syndrome; SIBO, small intestinal bacterial overgrowth; SSB, sugar-sweetened beverages.

[1]Only 1 representative modality is shown for compactness reasons.

The last pattern consisted of participants who appeared to follow low-carbohydrate diets with nearly no consumption of starchy foods or sweet products. As such, this pattern was named the Exclusion diet (ED). These participants had a diet enriched in fats and animal products, as well as nonstarchy vegetables (Figure 2). In the main AGP questionnaire, 40% (based on 139 subjects) declared following a modified paleo diet (Table 2). This pattern also presented the highest total water intake, the lowest vegetable protein intake, and, as expected, the lowest percentage of carbohydrates consumed compared to all other dietary patterns. Moreover, this pattern included a higher number of individuals with self-reported health conditions compared to some or all other patterns, including GI disorders like irritable bowel syndrome (IBS), small intestinal bacterial overgrowth, and specifically gluten intolerance, which was reported by more than half of this group. Of note, most of these observations were unrelated to age, sex, or BMI differences among the diet patterns (**Supplemental Table 7**).

### Characterization of gut microbiome based on dietary patterns

We then investigated how the gut microbiome composition differed among the DP5 patterns. Variations in alpha-diversity indices were small among diet patterns, but the SW pattern exhibited a lower alpha diversity compared to both the FL [Faith's phylogenetic diversity (PD): $P = 0.009$; observed ASVs: $P = 0.007$] and the HW (Faith's PD: $P = 0.048$) patterns (**Figure 3**A). Taking a deeper look at the composition of the microbiota, we found that 25 genera were significantly differentially abundant among the DP5 patterns using DESeq2 (Figure 3B). While differences between the Prudent diets (FL compared with PB) were modest, differences between the Western diets (HW compared with SW) were more pronounced (Figure 3B). In addition, when comparing the most typical Western-style diet (i.e., SW) to the Prudent-like diets, 16 genera were differentially abundant compared to the PB pattern and 11 genera were differentially abundant compared to the FL pattern (Figure 3B).

The following results describe the genera that were found to be significantly differentially abundant between the Prudent and Western diets with DESeq2, and are focused on those with higher relative abundances (≥0.5%). In order to increase the robustness of our findings, only those that were confirmed using the Songbird method are highlighted (Figure 3B; **Supplemental Figure 6**). *Parabacteroides'* relative abundance was significantly higher in the HW and SW patterns compared with PB (Figure 3C). In contrast, the relative abundance of *Oribacterium* was significantly higher in both Prudent-like diets compared to both Western-style diets, and was negatively associated with the animal:vegetable protein ratio (**Supplemental Figure 7**A). Lastly, an unidentified genus from the *RF39* order showed significantly lower relative abundance in the SW pattern compared to all the other diet patterns, highlighting that its depletion may be specific to the most "typical" Western-style diet (Figure 3C). To note, *Lactococcus'* relative abundance was also significantly lower in the Western patterns compared with the Prudent patterns, but both its relative abundance and prevalence were low (0.2% and 21%, respectively).

### The ED pattern is associated with a depletion in *Bifidobacterium*

The most striking difference in microbial genera abundances among the DP5 was a significantly lower relative abundance of *Bifidobacterium* in the ED pattern compared to all other patterns, with 1.7 ± 0.3 to 2.7 ± 0.4 log2 fold changes (i.e., relative abundance divided approximately by 3 to 6; Figure 3C). This observation was confirmed by Songbird analyses (Supplemental Figure 6). To identify the drivers of these differences, we used a prediction model (random forest) to test which variables (diet, demography, health status) contributed the most to *Bifidobacterium*'s relative abundance. The classification error of the obtained model was 37.4%, with an area under the receiver operating characteristic of 0.68, indicating moderate predictive power. Amongst the 20 tested variables, the DP5 was the top predictor based on variable importance, closely followed by age (Figure 3D). This result highlights the probable contribution of the diet itself, and not only preexisting health conditions, to the observed low *Bifidobacterium* relative abundance in the ED pattern. Of note, *Bifidobacterium*'s relative abundance was also positively associated with HEI-2010 scores (log2 fold change = 0.4 ± 0.1, $P = 0.001$; Supplemental Figure 7B). As a follow-up, we tried to identify dietary components that may contribute the most to the observed differential abundance of *Bifidobacterium*. We found that *Bifidobacterium* was positively associated with vegetable proteins and negatively associated with the ratio between animal and vegetable proteins (Supplemental Figure 7A). These observations are in line with the lower consumption of vegetable protein and the higher protein ratio in the ED pattern (Table 2; **Supplemental Table 8**). In addition, we added other confounding factors to the DESeq2 model to identify additional contributors to this association. The significant negative association of *Bifidobacterium* with the ED pattern decreased the most after adjusting for percentages of carbohydrates (fibers and both complex and simple polysaccharides) and fats from total calories consumed (**Supplemental Figure 8**). Nevertheless, this association was maintained, suggesting that other dietary components may also influence the abundance of *Bifidobacterium*, which further highlights the importance of considering diet as a whole when trying to understand diet-microbiome associations.

## Discussion

In this study, we explored the association between habitual diets and gut microbiomes in a large, population-based adult cohort using multiple approaches to examine dietary patterns. Our results showed that the overall diet exhibited more significant associations with the gut microbiome than individual dietary components.

The association between a habitual diet and the gut microbiota is gaining major interest, and an increasing number of cross-sectional studies have recently investigated this topic within human health (10, 11, 13, 18). When evaluating the impact of diet on gut microbiota, dietary patterns can be studied using either a priori (i.e., knowledge driven) or a posteriori (i.e., data-driven) analyses (7, 8). Yet, to the best of our knowledge, no study has compared both approaches to identify those providing the best associations with the gut microbiome. In this study, we

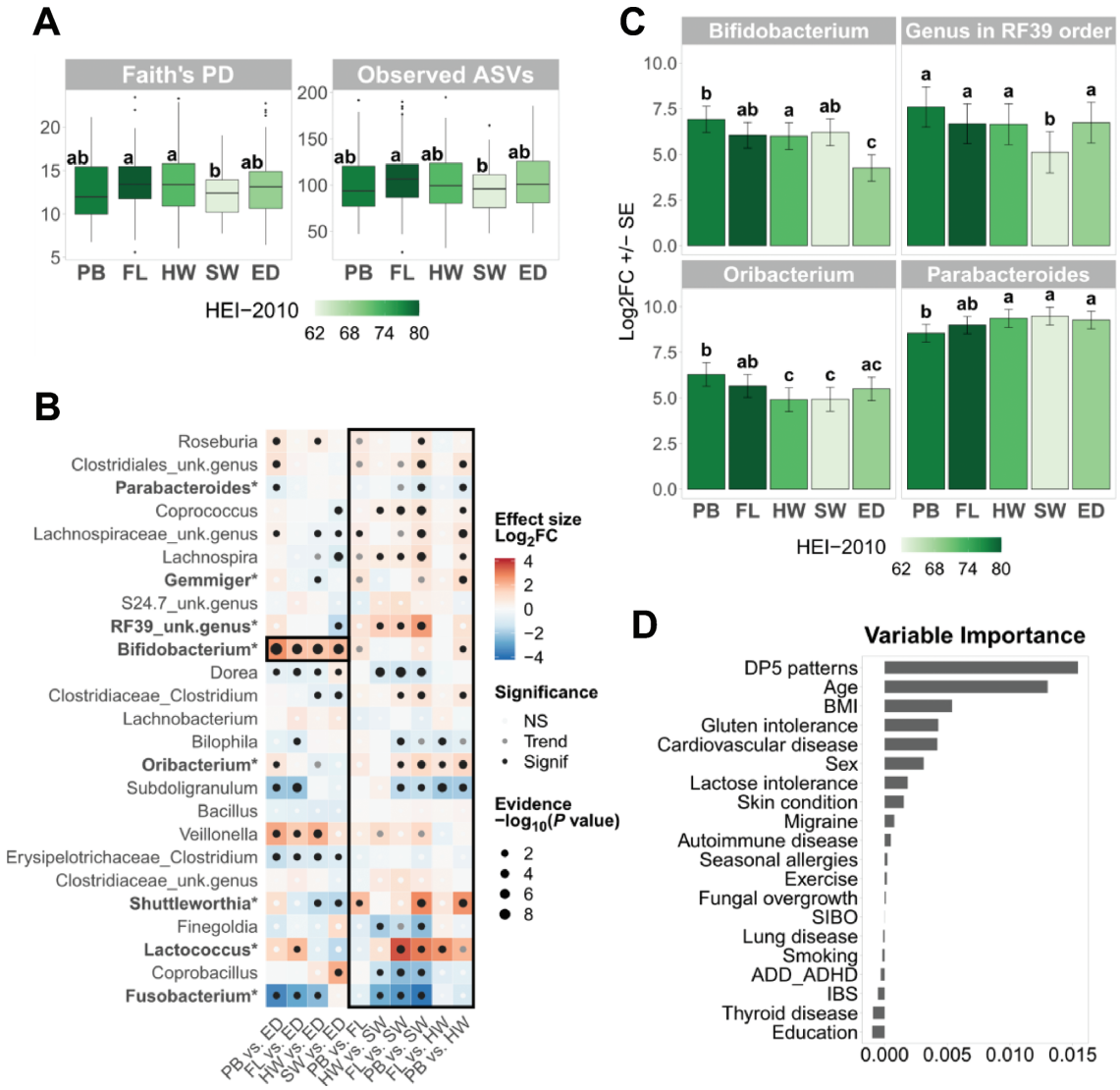


**FIGURE 3** Associations of DP5 patterns with gut microbiome. (A) Box plots for significant alpha-diversity indices. Plotted values are adjusted for age, sex, and BMI with a linear model. We performed Kruskal-Wallis tests with multiple testing adjustment by Benjamini-Hochberg on the global effects obtained with the 5 alpha diversity indices. If the global effect was significant ($P$ value $< 0.05$), post hoc comparisons were performed using Mann-Whitney tests with a Benjamini-Hochberg adjustment for each alpha diversity index separately. Groups with the same letter are not significantly different ($P$ value $\geq 0.05$). Boxes are colored by median HEI-2010 total score. (B) Heat map of 2 x 2 comparison results for significant genera. Genera are ordered from bottom to top by increasing abundance. Genera below the dotted grey line have mean relative abundances below 0.5% in the analyzed data set. Results for the *Bifidobacterium* genus and the Prudent/Western framework are highlighted in black boxes. Unknown genera are annotated at the lower available taxonomic level. DESeq2 (v1.28.1) models include pattern, age, sex, and BMI effects. The global pattern effect (likelihood ratio tests) was adjusted with the Benjamini-Hochberg procedure for multiple testing. If the global effect was significant ($P$ value $< 0.05$), all 2 x 2 comparisons (Wald tests) were reported using a Benjamini-Hochberg adjustment for each genus separately. For example, for PB vs. ED, if the Log2FC is positive, then the genus' relative abundance is higher in the PB pattern. Genera in bold with a star were found in Songbird Top10 differentials for at least 50% of DESeq2 significant results (Supplemental Figure 6). NS: $P$ value $\geq 0.1$; trend: $0.05 \leq P$ value $< 0.1$; and significance: $P$ value $< 0.05$. (C) Bar plots for selected DESeq2 genera results. Log2FC $\pm$ SE values were estimated by a DESeq2 model with no intercept. Groups with the same letter are not significantly different ($P$ value $\geq 0.05$). Bars are colored by the median HEI-2010 total score. (D) Variable importance in a random forest model for prediction of "low" or "high" *Bifidobacterium* status (see Supplemental Methods). Abbreviations: ADD, attention deficit disorder; ADHD, attention deficit hyperactivity disorder; ASV, amplicon sequence variant; DP5, 5 dietary patterns; ED, Exclusion diet; FL, Flexitarian diet; HEI-2010, Healthy Eating Index 2010; HW, Health-Conscious Western diet; IBS, irritable bowel syndrome; Log2FC, log2 fold change; NS, not significant; PB, Plant-Based diet; PD, phylogenetic diversity; SIBO, small intestinal bacterial overgrowth; SW, Standard Western diet; unk., unknown.

defined 61 dietary patterns as being further associated with the gut microbiome. These patterns included conventional nutrient analysis (e.g., fiber and protein quartiles); an a priori–defined diet quality index (i.e., HEI-2010), previously reported to better associate with gut microbiota than other diet indices (12); and a posteriori data-driven approaches. In contrast to recent findings (12, 46, 47), we did not find an association between HEI-2010 scores or dietary fiber and microbiome alpha diversity. This could be due to the high HEI-2010 scores and fiber intake values in this study cohort, a lower variation in gut microbiota/HEI-2010 scores/fiber intake, or a confounding factor that we did not control for. The data-driven approach used in

this study was based on several types of diet components using both factor and clustering analyses. We performed clustering of dietary data on the percentage of total energy intake from foods consumed, which can account for differences in energy needs (e.g., age, sex, physical activity level) (48) and provide more interpretable results (49). In addition, while compositional analysis has recently arisen in the nutrition field (50), our work employed a similar approach (DMM models), which is used in the microbiome field (51). This resulted in 5 a posteriori dietary patterns (the DP5) which take into consideration global diet and were best associated with the gut microbiome. The DP5 patterns correspond to the typical Western and Prudent diets (45), as well as a specific Exclusion (low-carbohydrate) diet. The Prudent and Western diets have more often been studied in the form of 2 independent factors (45, 52, 53) than in the form of clusters (54). Here, we identified a gradient in dietary intake with 2 different Western (HW and SW) and 2 different Prudent (PB and FL) dietary patterns. To our knowledge, a similar range of Western to Prudent diets has only been addressed in 1 previous study, which identified 3 diet clusters in women only (55). This dietary gradient was made evident through the evaluation of dietary quality. The SW pattern, exhibiting the poorest overall diet quality, was characterized by high consumption of sugar-sweetened beverages, animal products, and processed foods, and the lowest intakes of vegetables and dietary fiber. Larger amounts of plant foods and fewer amounts of animal foods, particularly meat, were associated with better overall diet quality and thus healthier dietary patterns (e.g., HW, FL, and PB). Such a gradient may be used as a basis for personalized nutrition recommendations, as shifting to a similar but healthier dietary pattern (e.g., from the HW pattern to the SW pattern) may be more sustainable and easily achieved than shifting to a less similar dietary pattern (e.g., from the HW pattern to the FL pattern).

While many studies have reported Western and/or Prudent dietary patterns in relation to metabolic health, only recently have food group–based a posteriori analyses been associated with gut microbiota (19, 56). We found that the FL pattern, which is enriched in plant-based foods but not depleted in animal products, was associated with the best HEI-2010 score and exhibited higher gut microbiota alpha diversity compared to the most typical Western-like diet (SW). This is partially in line with other cross-sectional studies in which the intake and diversity of fruits and vegetables have been reported as main factors associated with variation of the gut microbiota (14–16, 46). This particular association is primarily attributed to the role of dietary fiber and resistant starch in promoting microbial diversity (57). Surprisingly, we observed that gut microbiota alpha diversity in the PB pattern was not significantly different from that in the SW pattern. This may be due to the depletion of some animal products, such as meat and dairy products, as animal protein has been shown to increase microbial diversity [reviewed by Singh et al. (58)]. At the microbial genus level, adherence to the 2 Prudent-like diets, especially the PB pattern, was associated with a higher abundance of *Oribacterium* and lower abundance of *Parabacteroides*. *Parabacteroides* has been shown to be low or absent in vegetarian diets compared to nonvegetarian diets (59) and increased after resistant starch intake (60) and with high adherence to a Mediterranean diet (61), but also decreased after a short-term (8-week) Mediterranean diet intervention in overweight and obese subjects (62). These conflicting results highlight the variability between studies, and possibly the contribution of different *Parabacteroides* species. In addition, the most typical Western diet (SW) was associated with a lower abundance of an unassigned genus in the *RF39* order (affiliated to *Mollicutes*), even when compared to the other Western pattern (HW). *RF39* has been found to be higher in a Prudent diet (19), as well as following almond consumption (63), and is associated with a lean BMI (64). Since the SW pattern exhibited a lower consumption of nuts, almonds may be a contributor to this association. Together with the higher alpha diversity (Faith's PD) observed in the HW compared to the SW pattern, these results highlight the importance of identifying variations in dietary behavior within Western diets. The most pronounced difference among the 5 dietary patterns was the depletion of *Bifidobacterium* in subjects adopting the ED, who also self-reported more GI disorders (e.g., IBS, gluten intolerance) and appeared to follow a lower-carbohydrate diet with less grains and, as such, gluten. Exclusion of gluten and carbohydrates, such as fermentable oligo-, di-, and mono-saccharides and polyols, in order to alleviate GI symptoms, has been previously associated with a lower abundance of *Bifidobacterium* (65). Notably, these 5 dietary patterns (i.e., the DP5) had higher contributions than age and GI disorders to explain *Bifidobacterium* abundances in this study cohort. This reinforces that overall diet is a main factor affecting the *Bifidobacterium* relative abundance in US adults. This finding is further supported by our confounder analysis, which indicated that the lower percentage of carbohydrates and the higher percentage of fats in the ED pattern were both contributors to this association, but not exclusive ones. In addition, we observed that a higher intake of plant proteins compared with animal proteins was positively associated with *Bifidobacterium*, which is consistent with previous reports (18, 66, 67). This supports the hypothesis that plant protein might be a potential but currently underexplored contributor of the beneficial effects of plant-based products on gut microbiota and specifically on *Bifidobacterium*.

The present study has several limitations. First, a causal relationship between diet and the gut microbiota cannot be determined due to the cross-sectional nature of this study. Second, dietary intake was self-reported and solely assessed by an FFQ. Even though FFQs are traditionally used in observational studies, they have inherent measurement error and bias (68, 69). Third, the study cohort is not representative of the general US adult population and appears to adopt good overall dietary habits. Due to country-specific dietary guidelines (e.g., MPED components) and the specific food groupings of the FFQ, our dietary pattern findings are not generalizable across countries/populations. In addition, this work was based on 16S rRNA gene sequencing, which allows for exploration of the overall microbiota composition, but limits taxonomy resolution and functional characterization.

In conclusion, we showed that a global approach better explains gut microbiota variations than single nutrient/food data in a large, population-based cohort. This suggests that future studies should consider diet patterns, such as the DP5, as a covariate in analyses on gut microbiota to disentangle the effects of single nutrients/foods from the overall dietary pattern. In addition, further longitudinal and intervention studies including better characterizations of diet and gut microbiota (e.g., shotgun or metabolomic analyses) in larger and diverse populations

are still needed to better understand diet and microbiome interactions.

We thank all present and former members of the THDMI collaboration for their valuable input. In particular, we thank Laetitia Demaretz and Matthieu Pichaud for insightful discussions, Lauriane Raidot and Shahrokh Farokhinia for data management, Bénédicte Monnerie, Daniel Freed and Marie Poupin for project management, Renata Korczak for dietary assessment guidance.

The authors' responsibilities were as follows—PV, RK, AC, and AC-M: designed the research; RK, DM, and SJS: conducted the research; AC, NSL, SC, AC-M, MS, JT, and FL: analyzed the data; AC, NSL, DGL, SJS, AC-M, MD, HK, and PV: interpreted the results; AC, NSL, AC-M, and MD: wrote the manuscript; and all authors: read and approved the final manuscript.

Author disclosures: AC, AC-M, SC, JT, HK, MD, and PV are employees of Danone Research. MS was a consultant for Danone Research. DGL, FL, NSL, DM, and SJS are supported through a collaborative research agreement with Danone Research. The other author reports no conflicts of interest.

## Data Availability

The data used in this study are available publicly in Qiita (https://qiita.ucsd.edu/) under the study ID 10317, and the associated sequences can be found under EBI accession ERP012803. Raw VioScreen data can be downloaded at ftp://ftp.microbio.me/AmericanGut/raw-vioscreen/. The source code for defining the a posteriori diet groups with Dirichlet Multinomial Mixture models and for associating them with the 16S microbiome can be obtained at: https://github.com/danone/dp5.analysis.

## References

1. Sommer F, Bäckhed F. The gut microbiota—masters of host development and physiology. Nat Rev Microbiol 2013;11(4):227–38.
2. Zinöcker MK, Lindseth IA. The Western diet–microbiome-host interaction and its role in metabolic disease. Nutrients 2018;10:365.
3. Johnson AJ, Vangay P, Al-Ghalith GA, Hillmann BM, Ward TL, Shields-Cutler RR, Kim AD, Shmagel AK, Syed ANPersonalized Microbiome Class Students,, et al., Personalized Microbiome Class Students Daily sampling reveals personalized diet-microbiome associations in humans. Cell Host Microbe 2019;25(6):789–802.e5.e5.
4. Hemler EC, Hu FB. Plant-based diets for personal, population, and planetary health. Adv Nutr 2019;10(Suppl 4):S275–83.
5. US Department of Agriculture and US Department of Health and Human Service Dietary Guidelines for Americans, 2020–2025. 9th Edition [Internet]. 2020. Available from: https://www.dietaryguidelines.gov/resources/2020-2025-dietary-guidelines-online-materials.
6. Ordovas JM, Ferguson LR, Tai ES, Mathers JC. Personalised nutrition and health BMJ. 2018;361:bmj.k2173.
7. Burggraf C, Teuber R, Brosig S, Meier T. Review of a priori dietary quality indices in relation to their construction criteria. Nutr Rev 2018;76(10):747–64.
8. Newby PK, Tucker KL. Empirically derived eating patterns using factor or cluster analysis: a review. Nutr Rev 2004;62(5):177–203.
9. Claesson MJ, Jeffery IB, Conde S, Power SE, O'Connor EM, Cusack S, Harris HMB, Coakley M, Lakshminarayanan B, O'Sullivan O, et al. Gut microbiota composition correlates with diet and health in the elderly. Nature 2012;488(7410):178–84.
10. Filippis FD, Pellegrini N, Vannini L, Jeffery IB, Storia AL, Laghi L, Serrazanetti DI, Cagno RD, Ferrocino I, Lazzi C, et al. High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. Gut 2016;65(11):1812–21.
11. Asnicar F, Berry SE, Valdes AM, Nguyen LH, Piccinno G, Drew DA, Leeming E, Gibson R, Le Roy C, Khatib HA, et al. Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. Nat Med 2021;27(2):321–32.
12. Bowyer RCE, Jackson MA, Pallister T, Skinner J, Spector TD, Welch AA, Steves CJ. Use of dietary indices to control for diet in human gut microbiota studies. Microbiome 2018;6(1):77.
13. Partula V, Mondot S, Torres MJ, Kesse-Guyot E, Deschasaux M, Assmann K, Latino-Martel P, Buscail C, Julia C, Galan P, et al. Associations between usual diet and gut microbiota composition: results from the Milieu Intérieur cross-sectional study. Am J Clin Nutr 2019;109(5):1472–83.
14. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, Kurilshikov A, Bonder MJ, Valles-Colomer M, Vandeputte D, et al. Population-level analysis of gut microbiome variation. Science 2016;352(6285):560–4.
15. Manor O, Dai CL, Kornilov SA, Smith B, Price ND, Lovejoy JC, Gibbons SM, Magis AT. Health and disease markers correlate with gut microbiome composition across thousands of people. Nat Commun 2020;11(1):5206.
16. McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, Aksenov AA, Behsaz B, Brennan C, Chen Y, et al. American gut: an open platform for citizen science microbiome research. mSystems 2018;3(3):e00031–18.
17. Taylor BC, Lejzerowicz F, Poirel M, Shaffer JP, Jiang L, Aksenov A, Litwin N, Humphrey G, Martino C, Miller-Montgomery S, et al. Consumption of fermented foods is associated with systematic differences in the gut microbiome and metabolome. mSystems 2020;5(2):e00901–19.
18. Bolte LA, Vich Vila A, Imhann F, Collij V, Gacesa R, Peters V, Wijmenga C, Kurilshikov A, Campmans-Kuijpers MJE, Fu J, et al. Long-term dietary patterns are associated with pro-inflammatory and anti-inflammatory features of the gut microbiome. Gut 2021;70(7):1287–98.
19. Ericson U, Brunkwall L, Hellstrand S, Nilsson PM, Orho-Melander M. A health-conscious food pattern is associated with prediabetes and gut microbiota in the Malmö Offspring study. J Nutr 2020;150(4):861–72.
20. Shikany JM, Demmer RT, Johnson AJ, Fino NF, Meyer K, Ensrud KE, Lane NE, Orwoll ES, Kado DM, Zmuda JM, et al. Association of dietary patterns with the gut microbiota in older, community-dwelling men. Am J Clin Nutr 2019;110(4):1003–14.
21. Vangay P, Johnson AJ, Ward TL, Al-Ghalith GA, Shields-Cutler RR, Hillmann BM, Lucas SK, Beura LK, Thompson EA, Till LM, et al. US immigration Westernizes the human gut microbiome. Cell 2018;175(4):962–72.e10.
22. Kristal AR, Kolar AS, Fisher JL, Plascak JJ, Stumbo PJ, Weiss R, Paskett ED. Evaluation of web-based, self-administered, graphical food frequency questionnaire. J Acad Nutr Diet 2014;114(4):613–21.
23. Bowman SA, Friday JE, Moshfegh AJ. MyPyramid Equivalents Database, 2.0 for USDA survey foods, 2003–2004. [Internet]. Beltsville (MD): Food Surveys Research Group, Beltsville Human Nutrition Research Center, Agricultural Research Service, USDA; 2008. Available from: http://www.ars.usda.gov/ba/bhnrc/fsrg.
24. Guenther PM, Kirkpatrick SI, Reedy J, Krebs-Smith SM, Buckman DW, Dodd KW, Casavale KO, Carroll RJ. The Healthy Eating Index-2010 is a valid and reliable measure of diet quality according to the 2010 Dietary Guidelines for Americans. J Nutr 2014;144(3):399–407.
25. National Cancer Institute. Reviewing and cleaning ASA24® data[Internet]. 2020. Available from: https://epi.grants.cancer.gov/asa24/resources/asa24-data-cleaning-2020.pdf.
26. Ryman TK, Austin MA, Hopkins S, Philip J, O'Brien D, Thummel K, Boyer BB. Using exploratory factor analysis of FFQ data to identify dietary patterns among Yup'ik people Public Health Nutr 2014;17:510–8.
27. Revelle W. Psych: procedures for psychological, psychometric, and personality research. [Internet]. Evanston (Illinois): Northwestern University; 2020. Available from: https://CRAN.R-project.org/package=psych.
28. Raiche G, Magis D. nFactors: parallel analysis and other non graphical solutions to the Cattell Scree test[Internet]. 2020. Available from: https://mran.microsoft.com/snapshot/2020-02-28/web/packages/nFactors/citation.html.
29. Morgan M. Dirichlet-Multinomial: dirichlet-multinomial mixture model machine learning for microbiome data. 2020.
30. Hennig C. Fpc: flexible procedures for clustering. [Internet]. 2020. Available from: https://CRAN.R-project.org/package=fpc.
31. McDonald D, Kaehler B, Gonzalez A, DeReus J, Ackermann G, Marotz C, Huttley G, Knight R. Redbiom: a rapid sample discovery and feature characterization system. mSystems 2019;4.

32. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, et al. Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems 2017;2(2):e00191–16.
33. Gonzalez A, Navas-Molina JA, Kosciolek T, McDonald D, Vázquez-Baeza Y, Ackermann G, DeReus J, Janssen S, Swafford AD, Orchanian SB, et al. Qiita: rapid, web-enabled microbiome meta-analysis. Nat Methods 2018;15(10):796–8.
34. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 2019;37(8):852–7.
35. Amir A, McDonald D, Navas-Molina JA, Debelius J, Morton JT, Hyde E, Robbins-Pianka A, Knight R. Correcting for microbial blooms in fecal samples during room-temperature shipping. mSystems 2017;2(2).
36. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. An improved greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J 2012;6(3):610–8.
37. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, et al. Vegan: community ecology package[Internet]. 2019. Available from: https://CRAN.R-project.org/package=vegan.
38. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15(12):550.
39. Morton JT, Marotz C, Washburne A, Silverman J, Zaramela LS, Edlund A, Zengler K, Knight R. Establishing microbial composition measurement standards with reference frames. Nat Commun 2019;10(1):2719.
40. Subirana I, Sanz H, Vila J. Building bivariate tables: the compareGroups package for R. J Stat Software 2014;57(12):1–16.
41. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Series B Stat Methodol 1995;57(1):289–300.
42. R Core Team. R: a language and environment for statistical computing. [Internet]. Vienna (Austria): R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/.
43. Villarroel MA, Blackwell DL, Jen A. Tables of summary health statistics for U.S. adults: 2018 national health interview survey. [Internet]. National Center for Health Statistics; 2019. NCHS, National Health Interview Survey, 2018. Available from: http://www.cdc.gov/nchs/nhis/SHS/tables.htm.
44. Dahlhamer JM. Prevalence of inflammatory bowel disease among adults aged ≥18 years—United States, 2015. MMWR Morb Mortal Wkly Rep 2016;65–9.
45. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol 2002;13(1):3–9.
46. Maskarinec G, Hullar MAJ, Monroe KR, Shepherd JA, Hunt J, Randolph TW, Wilkens LR, Boushey CJ, Le Marchand L, Lim U, et al. Fecal microbial diversity and structure are associated with diet quality in the multiethnic cohort adiposity phenotype study. J Nutr 2019;149(9):1575–84.
47. Menni C, Jackson MA, Pallister T, Steves CJ, Spector TD, Valdes AM. Gut microbiome diversity and high-fibre intake are related to lower long-term weight gain. Int J Obes 2017;41(7):1099–105.
48. Devlin UM, McNulty BA, Nugent AP, Gibney MJ. The use of cluster analysis to derive dietary patterns: methodological considerations, reproducibility, validity and the effect of energy mis-reporting. Proc Nutr Soc 2012;71(4):599–609.
49. Hearty AP, Gibney MJ. Comparison of cluster and principal component analysis techniques to derive dietary patterns in Irish adults. Br J Nutr 2009;101(4):598–608.
50. Solans M, Coenders G, Marcos-Gragera R, Castelló A, Gràcia-Lavedan E, Benavente Y, Moreno V, Pérez-Gómez B, Amiano P, Fernández-Villa T, et al. Compositional analysis of dietary patterns. Stat Methods Med Res 2019;28(9):2834–47.
51. Costea PI, Hildebrand F, Arumugam M, Bäckhed F, Blaser MJ, Bushman FD, de Vos WM, Ehrlich SD, Fraser CM, Hattori M, et al. Enterotypes in the landscape of gut microbial community composition. Nat Microbiol 2018;3(1):8–16.
52. Walsh EI, Jacka FN, Butterworth P, Anstey KJ, Cherbuin N. The association between Western and Prudent dietary patterns and fasting blood glucose levels in type 2 diabetes and normal glucose metabolism in older Australian adults. Heliyon 2017;3(6):e00315.
53. Strate LL, Keeley BR, Cao Y, Wu K, Giovannucci EL, Chan AT. Western dietary pattern increases, whereas Prudent dietary pattern decreases, risk of incident diverticulitis in a prospective cohort study. Gastroenterology 2017;152(5):1023–30.e2.
54. Stricker MD, Onland-Moret NC, Boer JMA, van der Schouw YT, Verschuren WMM, May AM, Peeters PHM, Beulens JWJ. Dietary patterns derived from principal component- and k-means cluster analysis: long-term association with coronary heart disease and stroke. Nutr Metab Cardiovasc Dis 2013;23(3):250–6.
55. Van Horn L, Tian L, Neuhouser ML, Howard BV, Eaton CB, Snetselaar L, Matthan NR, Lichtenstein AH. Dietary patterns are associated with disease risk among participants in the Women's Health Initiative observational study. J Nutr 2012;142(2):284–91.
56. Jang HB, Choi M-K, Kang JH, Park SI, Lee H-J. Association of dietary patterns with the fecal microbiota in Korean adolescents. BMC Nutrition 2017;3(1):20.
57. Holscher HD. Dietary fiber and prebiotics and the gastrointestinal microbiota. Gut Microbes 2017;8(2):172–84.
58. Singh RK, Chang H-W, Yan D, Lee KM, Ucmak D, Wong K, Abrouk M, Farahnik B, Nakamura M, Zhu TH, et al. Influence of diet on the gut microbiome and implications for human health. J Transl Med 2017;15(1):73.
59. Ruengsomwong S, La-Ongkham O, Jiang J, Wannissorn B, Nakayama J, Nitisinprasert S. Microbial community of healthy Thai vegetarians and non-vegetarians, their core gut microbiota, and pathogen risk. J Microbiol Biotechnol 2016;26(10):1723–35.
60. Martínez I, Kim J, Duffy PR, Schlegel VL, Walter J. Resistant starches types 2 and 4 have differential effects on the composition of the fecal microbiota in human subjects. PLoS One 2010;5(11):e15046.
61. Garcia-Mantrana I, Selma-Royo M, Alcantara C, Collado MC. Shifts on gut microbiota associated to Mediterranean diet adherence and specific dietary intakes on general adult population. Front Microbiol 2018;9:890.
62. Meslier V, Laiola M, Roager HM, De Filippis F, Roume H, Quinquis B, Giacco R, Mennella I, Ferracane R, Pons N, et al. Mediterranean diet intervention in overweight and obese subjects lowers plasma cholesterol and causes changes in the gut microbiome and metabolome independently of energy intake. Gut 2020;69(7):1258–68.
63. Dhillon J, Li Z, Ortiz RM. Almond snacking for 8 wk increases alpha-diversity of the gastrointestinal microbiome and decreases bacteroides fragilis abundance compared with an isocaloric snack in college freshmen. Curr Dev Nutr 2019;3(8):nzz079. doi: 10.1093/cdn/nzz079.
64. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van Treuren W, Knight R, Bell JT, et al. Human genetics shape the gut microbiome. Cell 2014;159(4):789–99.
65. Reddel S, Putignani L, Del Chierico F. The impact of low-fodmaps, gluten-free, and ketogenic diets on gut microbiota modulation in pathological conditions. Nutrients 2019;11(2):373.
66. Świątecka D, Dominika Ś, Narbad A, Arjan N, Ridgway KP, Karyn RP, Kostyra H, Henryk K. The study on the impact of glycated pea proteins on human intestinal bacteria. Int J Food Microbiol 2011;145(1):267–72.
67. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, Mujagic Z, Vila AV, Falony G, Vieira-Silva S, et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science 2016;352(6285):565–9.
68. Day N, McKeown N, Wong M, Welch A, Bingham S. Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. Int J Epidemiol 2001;30(2):309–17.
69. Naska A, Lagiou A, Lagiou P. Dietary assessment methods in epidemiological research: current state of the art and future prospects. F1000Res 2017;6:926.