Consumption of Fermented Foods Is Associated with Systematic Differences in the Gut Microbiome and Metabolome

Bryn C Taylor, Franck Lejzerowicz, Marion Poirel, Justin P Shaffer, Lingjing Jiang, Alexander Aksenov, Nicole Litwin, Gregory Humphrey, Cameron Martino, Sandrine Miller-Montgomery, et al.

Consumption of Fermented Foods Is Associated with Systematic Differences in the Gut Microbiome and Metabolome

Bryn C. Taylor,[a] Franck Lejzerowicz,[b] Marion Poirel,[f] Justin P. Shaffer,[b] Lingjing Jiang,[c] Alexander Aksenov,[d] Nicole Litwin,[i] Gregory Humphrey,[b] Cameron Martino,[e] Sandrine Miller-Montgomery,[h,i] Pieter C. Dorrestein,[b,d,i] Patrick Veiga,[g] Se Jin Song,[i] Daniel McDonald,[b] Muriel Derrien,[g] Rob Knight[b,h,i,j]

aBiomedical Sciences Graduate Program, University of California San Diego, La Jolla, California, USA

bDepartment of Pediatrics, School of Medicine, University of California San Diego, La Jolla, California, USA

cDivision of Biostatistics, University of California San Diego, La Jolla, California, USA

dSkaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California, USA

eBioinformatics and Systems Biology Program, University of California San Diego, La Jolla, California, USA

fIT&M Innovation on behalf of Danone Nutricia Research, Palaiseau, France

gDanone Nutricia Research, Palaiseau, France

hDepartment of Bioengineering, Jacobs School of Engineering, University of California San Diego, La Jolla, California, USA

iCenter for Microbiome Innovation, University of California San Diego, La Jolla, California, USA

jDepartment of Computer Science and Engineering, University of California San Diego, La Jolla, California, USA

ABSTRACT Lifestyle factors, such as diet, strongly influence the structure, diversity, and composition of the microbiome. While we have witnessed over the last several years a resurgence of interest in fermented foods, no study has specifically explored the effects of their consumption on gut microbiota in large cohorts. To assess whether the consumption of fermented foods is associated with a systematic signal in the gut microbiome and metabolome, we used a multi-omic approach (16S rRNA amplicon sequencing, metagenomic sequencing, and untargeted mass spectrometry) to analyze stool samples from 6,811 individuals from the American Gut Project, including 115 individuals specifically recruited for their frequency of fermented food consumption for a targeted 4-week longitudinal study. We observed subtle but statistically significant differences between consumers and nonconsumers in beta diversity as well as differential taxa between the two groups. We found that the metabolome of fermented food consumers was enriched with conjugated linoleic acid (CLA), a putatively health-promoting molecule. Cross-omic analyses between metagenomic sequencing and mass spectrometry suggest that CLA may be driven by taxa associated with fermented food consumers. Collectively, we found modest yet persistent signatures associated with fermented food consumption that appear present in multiple -omic types which motivate further investigation of how different types of fermented food impact the gut microbiome and overall health.

IMPORTANCE Public interest in the effects of fermented food on the human gut microbiome is high, but limited studies have explored the association between fermented food consumption and the gut microbiome in large cohorts. Here, we used a combination of omics-based analyses to study the relationship between the microbiome and fermented food consumption in thousands of people using both cross-sectional and longitudinal data. We found that fermented food consumers have subtle differences in their gut microbiota structure, which is enriched in conjugated linoleic acid, thought to be beneficial. The results suggest that further studies of specific kinds of fermented food and their impacts on the microbiome and health will be useful.

Fermentation is an ancient process of food preparation dating from the introduction of agriculture and animal husbandry during the Neolithic period approximately 10,000 years ago. Advantages of food fermentation include improvements in food preservation, food safety, nutritional value, and organoleptic quality resulting from the activity of microbial ecosystems (bacteria and yeast) (1). Fermentation can be applied to a range of food types, including meat, fish, milk, vegetables, beans, cereals, and fruits, and occurs spontaneously from the original ingredients or environment or is controlled by the addition of specific starters such as lactic acid bacteria (LAB) (2). These bacteria are commonly detected in fermented food, mostly including *Lactobacillus*, *Streptococcus*, *Lactococcus*, and *Leuconostoc*, but other bacteria as well as yeast and fungi are also involved in food fermentations (3). In addition to microbial diversity, the number of microorganisms present in fermented foods varies between food type, process, and storage. A survey of diverse fermented food products suggested that the count of viable lactic acid bacteria usually reaches at least 10^6 cells/ml (4). Recovery of viable bacterial and fungal species ingested through fermented food has been observed in subjects who consume an animal-based diet (5). Moreover, metabolites generated from fermentation, including lactic acid, vitamins, and exopolysaccharides, are thought to exert health benefits (6). A recent study reported that D-phenyllactic acid, produced by LAB, interacts with the human host through the activation of hydroxycarboxylic acid receptor 3 (HCA3) and is involved in the regulation of immune functions and energy homeostasis under changing metabolic and dietary conditions (7).

Due to their supposed health benefits (6), there has been a resurgence of interest in consumption of fermented foods in Western society. To date, many of the studies focused on the health benefits of fermented food intake have been mostly focused on yogurt, consumption of which is associated with better metabolic parameters in large American cohorts (8, 9). Similarly, high intake of fermented foods has been associated with a lower prevalence of atopic dermatitis in a Korean population (10), and another study found consumption of miso and natto to be inversely associated with high blood pressure in a Japanese population (11).

While we know that both short- and long-term dietary intake affects the structure, function, and activity of the human gut microbiome (5, 12–16), and a few studies have explored the response of gut microbiota to a single type of fermented food (recently reviewed in reference 17), no study has explored the functional capacity of the gut microbiota of fermented food consumers. Intervention studies, which are often underpowered for analysis of the gut microbiome response, are complemented by studies of population-based cohorts, which due to large sample sizes have the advantage of capturing large amounts of microbial variation and enable us to disentangle the contributions of host and environmental factors such as diet (18–21).

To address the hypothesis that fermented food consumption is associated with compositional or functional changes in the human gut microbiome, we analyzed a subset of the American Gut Project (AGP) cohort based on self-reported consumption of fermented foods, and in particular, fermented plants. We also explored the longitudinal stability and function of the gut microbiota using untargeted high-performance liquid chromatography–tandem mass spectrometry (HPLC-MS/MS) and 16S rRNA amplicon sequencing, as well as shotgun sequencing on a subset of subjects at a single time point.

RESULTS

Demographic and dietary assessments of fermented plant consumers and nonconsumers. To explore the differences in the gut microbiome between fermented food consumers and nonconsumers, we analyzed 16S rRNA sequencing data from 28,114 samples from 21,464 individuals in the AGP (Fig. 1a). After filtering (see Materials and Methods), 6,811 participants were retained, and here are referred to as the
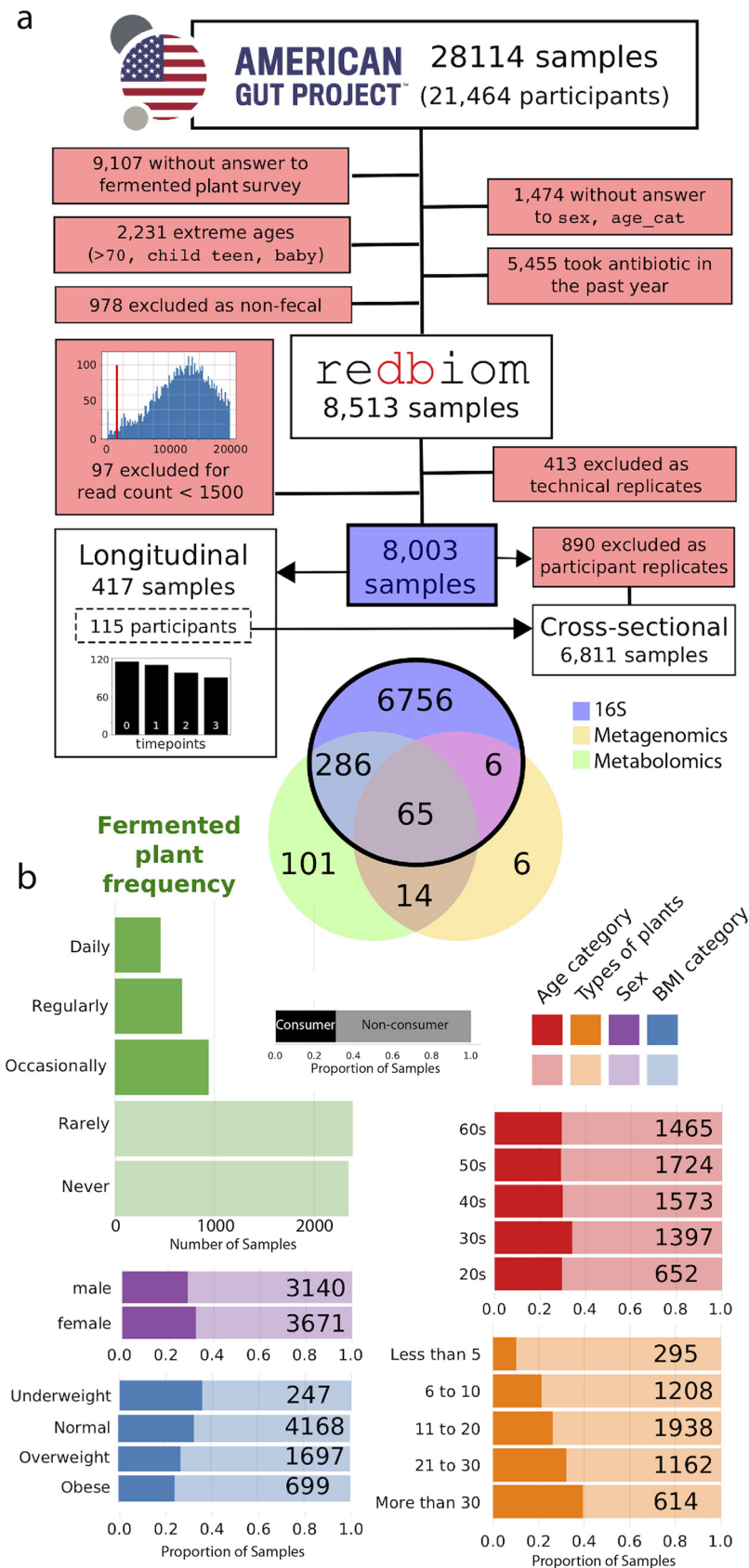
FIG 1 Cohort overview, sample filtering, and metadata exploration. (a) Data filtering process and the number of samples analyzed by 16S rRNA gene sequencing, metabolomics, and shotgun metagenomics

(Continued on next page)

cross-sectional cohort (Fig. 1a). One hundred fifteen of these participants were initially recruited for a concurrent longitudinal assessment which is discussed in detail below. Participants were identified as "consumers" or "nonconsumers" depending on the frequency of fermented plants that they reported consuming. The fermented plant frequency question is in the standard AGP questionnaire that every participant answered, and while the language may not have allowed for the capture of all fermented foods, this represented the most efficient way to delineate consumers and nonconsumers. We considered consumers to be those who reported eating fermented plants "daily," "regularly (3 to 5 times/week)," or "occasionally (1 to 2 times/week)" and nonconsumers to be those who reported eating fermented plants either "rarely (less than once/week)" or "never" (Fig. 1b). A 30.5% proportion of participants were considered consumers, of which most (45.3%) were occasional consumers. Consumer and nonconsumer cohorts were composed of slightly differing demographic groups. For example, while consumers were significantly younger than nonconsumers, the difference was modest (47 versus 47.61 years, respectively), with a higher proportion of participants in their 30s (23.0% versus 19.4%; chi-square test = 11.08, P = 0.03) (Fig. 1b). Similarly, the consumer group was composed of a modestly higher proportion of females (56.8% versus 52.6%; chi-square test = 9.60, P = 0.002) and a higher proportion of participants with a normal body mass index (BMI) between 18.5 and 25 (65.6% versus 59.3%; chi-square test = 35.93, $P \ll 0.001$), with an average BMI of 23.9 and 24.8, respectively. Consumers also reported eating a greater diversity of plants (>20) (29.7% versus 24.5%; chi-square test = 126.96, $P \ll 0.001$). In addition, because alcohol may be an end product of a fermentation process and might be a confounding factor associated with gut microbiota variation, we verified that alcohol consumption was not associated with fermented plant consumption (81.7% versus 82.6%, chi-square test = 0.76, P value = 0.38).

Statistically significant differences in mean total carbohydrate and fat intake (grams/day and percentage of energy) and percentage of energy from protein, as estimated by the food frequency questionnaire (FFQ), were observed between fermented plant consumers and nonconsumers, while total energy (kilocalories/day), dietary fiber (grams/day), and protein (grams/day) intake did not differ (see Table S1 in the supplemental material). There was no significant difference in overall diet quality observed, as assessed by the Healthy Eating Index (HEI-2010; Mann-Whitney U = 223409, P value = 0.094; Fig. S1A), despite the differences in the consumption of fermented plants and number of plant types between consumers and nonconsumers; this nonsignificant difference in total HEI-2010 scores between consumers and nonconsumers (71.29 versus 71.53, respectively) suggests similar intake of dietary patterns relatively high in quality. It should be noted that the mean total HEI-2010 score for both consumers and nonconsumers is above the national average (58.27) for U.S. adults aged 18 to 64 years based on 2011–2012 National Health and Nutrition Examination (NHANES) data (22). This suggests that the cohort in our study has a diet pattern that better aligns to the Dietary Guidelines for Americans than that of average American adults. Additionally, it has been shown that higher HEI scores are associated with higher income and education levels (23, 24), thereby suggesting that the higher total HEI scores observed in this AGP cohort may reflect higher-than-average socioeconomic status and education level as previously observed (25).

Gut microbiome composition in fermented plant consumers and nonconsumers. Examining unweighted UniFrac distances (26), we observed a statistically significant difference in the overall gut microbial communities between consumers and nonconsumers (Fig. S1B, permutational multivariate analysis of variance [PERMANOVA]

FIG 1 Legend (Continued)
and the resulting number of samples in the cross-sectional and longitudinal cohorts. (b) Distribution of some metadata categories (demographic and diet) in the cross-sectional cohort between consumers and nonconsumers. Darker colors denote consumers; lighter colors denote nonconsumers. The consumer and nonconsumer groups were defined by the "fermented plant frequency" questionnaire.

pseudo-F-statistic = 3.677, P = 0.001). The comparison of nonconsumers with occasional consumers results in a weaker group separation (F-statistic = 2.233, P value = 0.001) than with regular or daily consumers (F-statistics = 3.512 and 3.246, respectively; P values = 0.001), suggesting a dose dependence for the frequency of fermented plant consumption on the gut microbiome. However, there was no dose dependence with frequency of types of plants between consumers and nonconsumers (unweighted UniFrac distances between consumers and nonconsumers versus the frequency number of types of plants, R^2 = 0.0065). There was no difference in alpha diversity between the two groups (Faith's phylogenetic diversity [PD], Shannon diversity, nor observed operational taxonomic unit [OTU] richness; Fig. S1B) and also no difference when groups were stratified by consumption frequency (Table S2).

Next, we used Songbird (27) to identify specific microbes that were associated with consumers or nonconsumers. Songbird is a compositionally aware differential abundance method which provides rankings of features (suboperational taxonomic units [sOTUs]) based on their log fold change with respect to covariates of interest. In this case, the formula we used described whether the subject consumed fermented plants or not. We selected the 20 highest ("set 1," Table S3)- and 20 lowest ("set 2," Table S3)-ranked sOTUs associated with fermented plant consumption and used Qurro (28) to compute the log ratio of these sets of taxa (Fig. S1C). Comparing the ratios of taxa in this way mitigates bias from the unknown total microbial load in each sample, and taking the log of this ratio gives equal weight to relative increases and decreases of taxa (27). Evaluation of the Songbird model for fermented plant consumption against a baseline model obtained a Q^2 value of -5.4249, suggesting possible overfitting related to the subtlety of the differences between fermented plant consumption groups. In order to verify the log ratios chosen by Songbird ranks, we performed a permutation test by taking 1,000 random permutations of log ratios with 20 nonoverlapping features in the numerator and denominator. The rank order, compared to the random permutation, was 16, corresponding to a P value of 0.0159 (Fig. S2A), suggesting that the log ratio based on the Songbird ranks is nonrandom. We found that consumers have a significantly higher log ratio of set 1 to set 2 than nonconsumers (t test, P = 0.00065, t = 3.6367), suggesting that they are associated with *Bacteroides* spp., *Pseudomonas* spp., *Dorea* spp., *Lachnospiraceae*, *Prevotella* spp., *Alistipes putredinis*, *Oscillospira* spp., *Enterobacteriaceae*, *Fusobacterium* spp., *Actinomyces* spp., *Achromobacter* spp., *Clostridium clostridioforme*, *Faecalibacterium prausnitzii*, *Bacteroides uniformis*, *Clostridiales*, and *Delftia* spp.

Gut microbiome composition in frequent and rare fermented food consumers. One hundred fifteen participants were recruited for a longitudinal study in order to assess the gut microbiome over time and at a finer resolution by using untargeted mass spectrometry in addition to 16S rRNA sequencing (Fig. 1a). We targeted participants who self-identified as frequent consumers or very rare consumers. Consumers were identified using the same definition as in the cross-sectional cohort: consumers ate fermented plants "daily," "regularly (3 to 5 times/week)," or "occasionally (1 to 2 times/week)"; nonconsumers ate fermented plants "rarely (less than once/week)" or "never" (Fig. S3). The longitudinal cohort was designed to have a higher proportion of consumers who reported eating fermented plants "daily" and "regularly" versus "occasionally" than the cross-sectional cohort (Fig. S4). Similarly, the nonconsumer group in the longitudinal cohort had a higher proportion of participants who reported eating them "never" and "rarely" (Fig. S4) than did nonconsumers in the cross-sectional study.

A separate fermented food questionnaire was provided to these 115 participants to characterize additional types of fermented food consumed and to evaluate the proxy of fermented plant consumption for general fermented food consumption. Briefly, the major fermented foods consumed were beer, kimchi, kombucha, pickled vegetables, sauerkraut, and yogurt. More consumers reported eating fermented foods than did nonconsumers (Fig. S3B). Only 7.0% of participants (8/115) who stated that they never consumed fermented plants reported consuming another type of fermented food. Of these eight participants, two reported that they consumed wine or beer; one partici-

pant reported consuming yogurt, cider, wine, and beer; and five participants reported consuming unspecified fermented foods. We also observed that fermented plant consumers more frequently ate fermented dairy products (yogurt, sour cream/crème fraiche, kefir milk, and cottage cheese) than did nonconsumers (Fig. S3B). Therefore, we further identified them as "fermented food consumers," in contrast to the cross-sectional cohort.

Within the 16S data, we did not observe a difference in alpha diversity (Shannon's index [29]) and Faith's phylogenetic diversity (30) between consumers and nonconsumers (Fig. S1B). We further applied a sparse functional principal-component analysis (31), which explicitly factors in the longitudinal component, and did not observe a significant difference in alpha diversity (Shannon's index, Wilcoxon $P = 0.20$), suggesting that the stability of alpha diversity in the microbiome over 4 weeks is consistent for consumers and nonconsumers.

A subset of 100 samples were sequenced by shotgun metagenomics to provide a finer resolution of the taxonomic differences between the two groups. First, we verified whether the gut microbiota of self-reported fermented food consumers was associated with fermented food-associated species. We computed a log ratio using Qurro (28) of fermented food-associated taxa according to the work of Marco et al. (6) ("set 3," Table S3) compared to a set of taxa that were present across all samples ("set 4," Table S3) (Fig. 2b). Eight species were detected in our data set and were used to compute this log ratio: *Lactobacillus acidophilus*, *Lactobacillus brevis*, *Lactobacillus fermentum*, *Lactococcus lactis*, *Leuconostoc mesenteroides*, *Lactobacillus paracasei*, *Lactobacillus plantarum*, and *Lactobacillus rhamnosus* (Fig. 2a). We found that consumers had a significantly higher log ratio of set 3 to set 4 than nonconsumers (t test, P value = 0.0001838, $t = 3.9386$, Cohen's $D = 0.851$), suggesting that consumers were associated with some taxa derived from fermented foods.

We then used Songbird (27) to test whether there was a broader set of microbial features associated with consumers or nonconsumers. We selected the 40 highest-ranked ("set 5," Table S3) and 40 lowest-ranked ("set 6," Table S3) microbes associated with fermented plant consumption and used Qurro to compute the log ratio of these sets of taxa (Fig. 2c); these were the smallest sets of features that provided meaningful differences between consumers and nonconsumers. Again, because evaluation of the Songbird models for fermented plant consumption against a baseline model suggested overfitting (Q^2 value of -0.12), we further verified the log ratios chosen by Songbird ranks by performing a permutation test of taking 1,000 random permutations of log ratios with 20 nonoverlapping features in the numerator and denominator. The rank order, compared to the random permutation, was 2, corresponding to a P value of 0.0019 (Fig. S2B), suggesting that the log ratio based on the Songbird ranks is nonrandom. This analysis at the species level showed that consumers have a significantly higher log ratio of set 5 to set 6 than nonconsumers (t test $P = 0.0024$, $t = 3.15$, Cohen's $D = 0.692$).

Several microbes of relevance to fermented foods were also associated with consumers, including *Lactobacillus acidophilus*, *Lactobacillus brevis*, *Lactobacillus kefiranofaciens*, *Lactobacillus parabuchneri*, *Lactobacillus helveticus*, and *Lactobacillus sakei* (6, 32–35) (Fig. 2a). Consumers were also associated with several other microbes unrelated to fermented foods, including *Streptococcus dysgalactiae*, *Prevotella melaninogenica*, *Enorma massiliensis*, *Prevotella multiformis*, *Enterococcus cecorum*, and *Bacteroides paurosaccharolyticus*. The microbes that distinguish consumers and nonconsumers in the cross-sectional and longitudinal data sets may not fully overlap because the longitudinal cohort was intentionally composed of participants in the more "extreme" ends of consumption (individuals who consume "daily" and "regularly" versus individuals who "never" consume fermented plants), because the cohorts were analyzed using different sequencing methods (16S versus metagenomics), or because of a combination of these aspects.

The functional profile of the gut microbiome differs with consumption of fermented food. To assess the functional profile of the gut microbiome of specifically
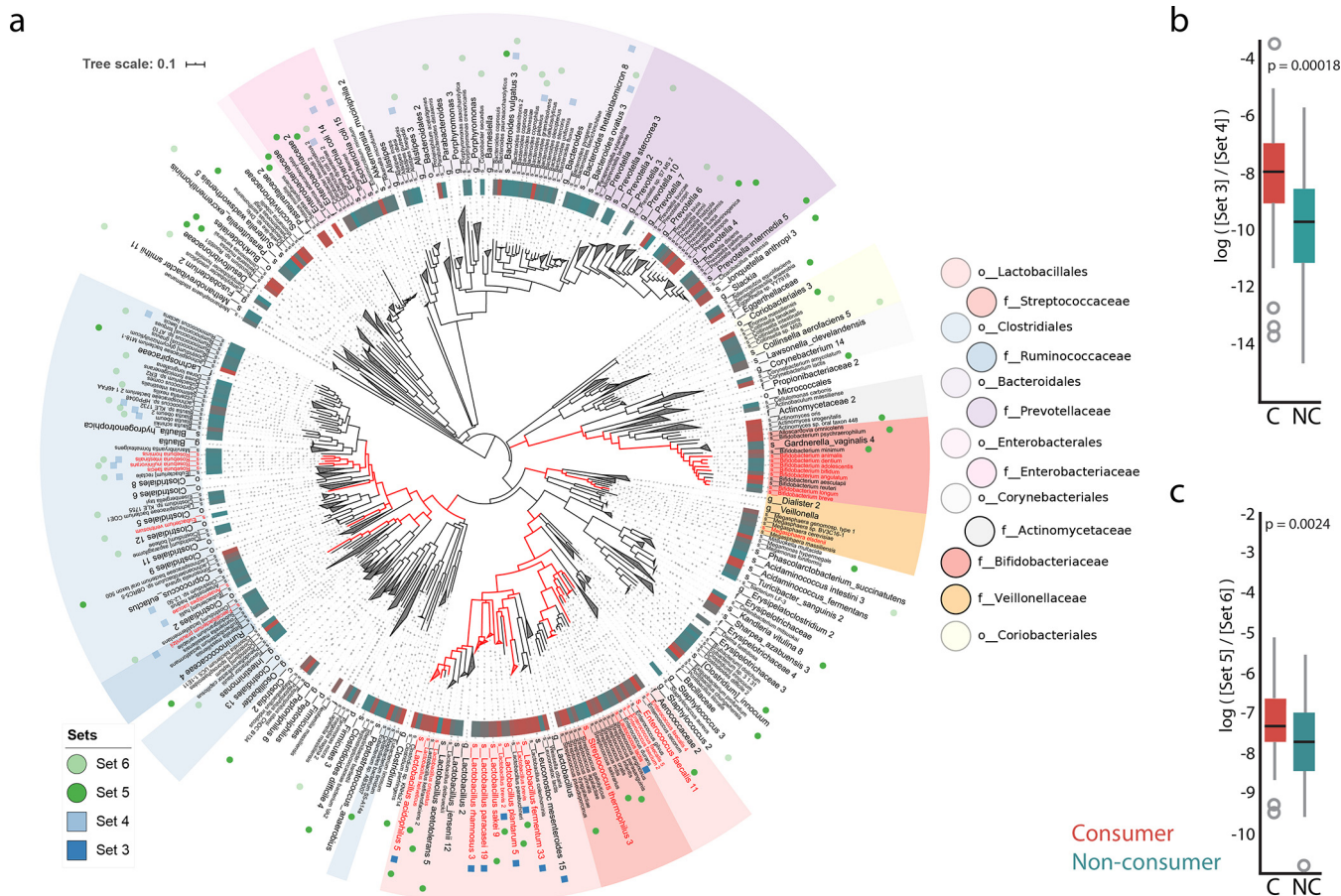
FIG 2 Phylogenetic and log ratio differences between consumers and nonconsumers in gut metagenomes. (a) Phylogenetic diversity captured in the metagenomes. The species taxon hits to the rep82 database were mapped to the species taxa of the Web of Life database (74) (97.8% mapped) in order to represent the phylogenetic distances computed from full genomes. The species known to produce CLA are indicated in red font. The species sets used for log ratio calculations are labeled using opaque (numerators) and transparent (denominators) colors. Set 3 is composed of microbes identified from reference 6, and set 4 contains the most prevalent microbes across all samples (blue squares). Set 5 and set 6 are derived from Songbird (green circles). (b) Consumers have a significantly higher ratio of set 3 to set 4 (*t* test, *P* value = 0.0001838, *t* = 3.9386, Cohen's *D* = 0.851). (c) The log ratio of set 5 to set 6 is significantly different (box plot, *t* test *P* = 0.0024, *t* = 3.15, Cohen's *D* = 0.692). The lists of microbes in each set are available in Table S3.

recruited fermented food consumers and nonconsumers, we performed untargeted HPLC-MS/MS analysis on all longitudinal samples (115 subjects, 417 samples, with up to 4 samples per subject, collected weekly for 4 weeks) (Fig. 1a). We explored the longitudinal stability using both the 16S and mass spectrometry data and found that the taxa and metabolites remained stable (Spearman's rho ranging from 0.42 to 0.68; *P* < 0.001) between time points within both consumers and nonconsumers (Fig. S5). The correlation coefficients for metabolites tended to be lower than for the taxa, suggesting more volatility in the observed metabolic features. This is expected since the metabolome is driven in large part by the diet, which changes day to day.

Using partial least squares discriminatory analysis (PLS-DA), we found that notable differences exist between consumers and nonconsumers when all time points were taken into account (Fig. 3a; Fig. S6A). The majority of the top discriminating features appeared to be lipids, several of which have broad natural distributions and thus are likely common. In particular, one compound was identified as octadecadienoic acid and then determined specifically to be an isomer of conjugated linoleic acid (CLA). At a single time point, we found that this isomer of CLA (designated "CLA4"; the exact configuration is unknown) was enriched in consumers (Wilcoxon test, *P* value = 0.04) whereas the unconjugated linoleic acid (LA) was not significantly different between the two groups (Wilcoxon test, *P* value = 0.52) (Fig. 3b). As CLA has also been found as one
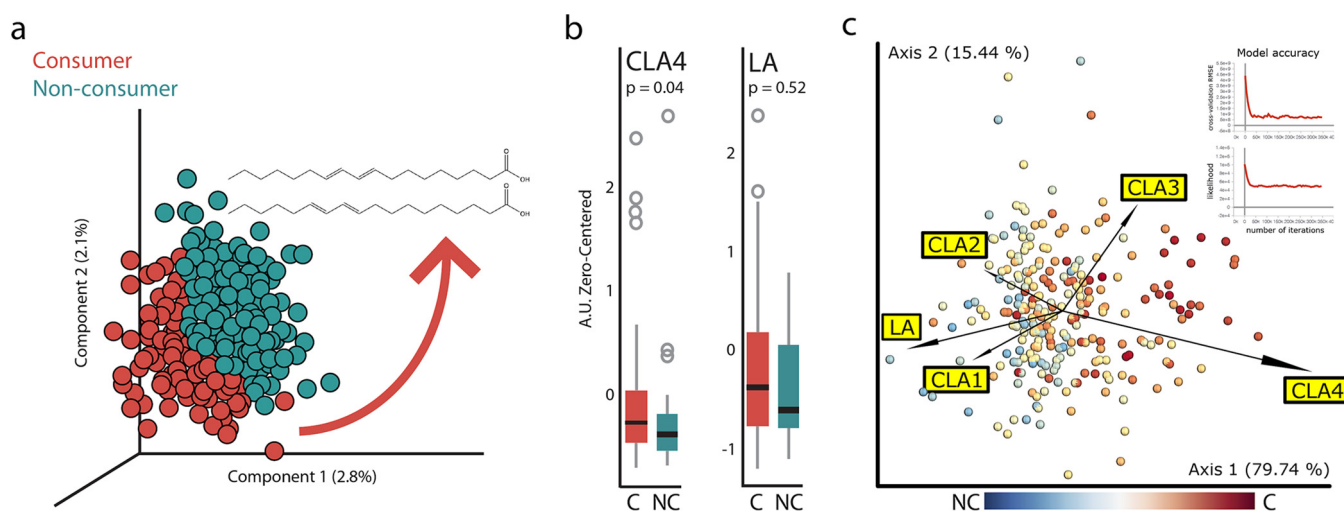
FIG 3 Conjugated linoleic acid is significantly higher in consumers than nonconsumers. (a) Partial least squares discriminatory analysis (PLS-DA) of untargeted mass spectrometry data identified CLA as one of the discriminating features in fermented food consumer samples. (b) Zero-centered counts of MS1 features annotated with a CLA isomer (designated "CLA4") and with the unconjugated linoleic acid (LA) between consumers and nonconsumers. "CLA4" is enriched in the consumer group (Wilcoxon test, P value = 0.04) but not LA (Wilcoxon test, P value = 0.52). (c) Integrative analysis of metagenomics and mass spectrometry data sets using mmvec. Genome features (dots) are labeled according to their strength of change with respect to fermented food consumption (red is associated with consumption, blue with nonconsumption). The metabolites are represented by arrows indicating their cooccurrences with the genomes.

of the discriminating features in samples from subjects who consume a large number of types of plants (25), it might suggest that the difference between consumers and nonconsumers could be partly explained by the number of types of plants consumed. However, in this study CLA abundances were not significantly different between the two extreme groups of types of plant consumption: fewer than 10 types of plants versus more than 30 types of plants (Wilcoxon rank sum test, P value = 0.98). From the food frequency questionnaire, we found that dietary consumption of total LA (18:2 n-6; g/day) and total CLA (g/day) did not differ significantly between consumers and nonconsumers (Fig. S6B), suggesting that the elevated levels of CLA in the fecal samples of consumers are likely derived from an endogenous process or microbial origin.

A total of 79 samples were analyzed using both metagenomic sequencing and mass spectrometry (Fig. 1a). We used mmvec (36) to integrate these data to assess cooccurrence patterns between genomic features (species) and the LA and CLA metabolites. We found that "CLA4," which was significantly enriched in consumers, cooccurs with the species (previously identified using Songbird) that were most strongly associated with consumers. Additionally, we found that linoleic acid (LA) cooccurs with the microbes that are most strongly associated with nonconsumers (Fig. 3c). Of the top 50 taxa that had the highest probability of cooccurring with "CLA," 14 are known CLA producers. These include *Eubacterium rectale*, *Faecalibacterium prausnitzii*, *Eubacterium siraeum*, *Eubacterium hallii*, *Bifidobacterium adolescentis*, and genera *Roseburia*, *Anaerostipes*, *Eubacterium*, *Ruminococcus*, and *Clostridium* (Table S4) (37–40). Forty-eight out of these top 50 taxa were more abundant in consumers than nonconsumers (Table S4).

DISCUSSION

In this study, we explored the gut microbiome of fermented plant consumers and nonconsumers in the American Gut Project (25), an extensive collection of sample contributions from tens of thousands of citizen scientists. Gut microbiome profiles, but not overall microbial diversity, differed slightly between the groups, suggesting that small but systematic compositional differences may occur based on a dietary choice to consume fermented plants. In a concurrent targeted longitudinal study, we found that fermented-food related taxa as well as a putatively health-associated molecule were associated with consumers. Several microbes that were found to be associated with

fermented consumers include microbes known to be derived from fermented foods, including fermented milk products (*Lactobacillus acidophilus* [6], *Lactobacillus brevis* [6], *Lactobacillus kefiranofaciens* [32], *Lactobacillus parabuchneri* [33], and *Lactobacillus helveticus* [34]) and fermented meat (*Lactobacillus sakei* [35]). This is consistent with other metagenomic studies from population-based cohorts that detected species related to starters such as *Leuconostoc mesenteroides* and *Lactococcus lactis* in subjects who consumed a specific fermented milk product (buttermilk) in the Dutch cohort Lifeline DEEP (20).

Analysis of the metabolomics data using PLS-DA found that shifts in lipid metabolism were associated with consumption of fermented plants, since the majority of the top discriminating metabolites appeared to be lipids. Of those that could be identified, CLA was particularly notable. The abundance of the CLA isomer "CLA4"is significantly increased in consumers over nonconsumers. CLA is known to be produced during ruminal bacterial fermentation and impacts the fatty acid composition of meat and dairy products from ruminants that represent the major dietary sources of CLA in humans (40). Due to its possible health benefits (41, 42), CLA is also often consumed as a nutritional supplement. However, CLA fecal recovery did not correlate with dietary CLA intake derived mainly from meat, full-fat dairy, and egg sources as determined by the food frequency questionnaire (FFQ). Moreover, dietary consumption of total CLA (grams/day) did not differ between consumers and nonconsumers. Thus, it is possible that CLA is being produced by resident or transient bacteria derived from fermented foods.

Indeed, diet-related bacteria, such as LAB, bifidobacteria, and propionibacteria, have been shown previously to produce CLA (39). Intestinal bacteria belonging to the families *Lachnospiraceae* and *Ruminococcaceae* have also been shown to metabolize LA into products that can be precursors of CLA (37), and two of these *Lachnospiraceae* were also found to be associated with consumers. The order *Lactobacillales* includes the largest diversity of previously reported CLA producers, and notably, seven out of the eight species previously identified as associated with fermented foods (set 3) are CLA-producing *Lactobacillus* species that we found to be associated with fermented food consumers: *L. acidophilus*, *L. brevis*, *L. fermentum*, *L. helveticus*, *L. paracasei*, *L. plantarum,* and *L. sakei* (for reviews, see references 38 to 40, 43). However, increased CLA in consumers cannot be fully attributed to production by fermented food-associated bacteria. For example, some members of the order *Clostridiales* previously reported to produce CLA in human feces (including four *Roseburia* species: *R. inulinivorans*, *R. hominis*, *R. intestinalis*, and *R. faecis* [37]) were found to be associated with nonconsumers, along with *Anaerostipes caccae*, *Eubacterium ventriosum* (L2-12), and *Faecalibacterium prausnitzii*, which are also known to metabolize LA.

We detected seven *Bifidobacterium* species previously reported to produce CLA using LA as a precursor (38, 39), including *Bifidobacterium animalis*, *B. longum* (44), and *B. breve*, which has been considered for CLA enrichment in commercial foods such as yogurt due to its CLA-producing ability (45). Yet, none of these were found to be associated with the fermented food consumers. Rather, two other *Bifidobacterium* species not known to produce CLA (*B. aesculapii* and *B. reuteri*) were found to be associated with fermented food consumers, with *B. reuteri* growth actually inhibited at high concentrations of LA precursor (46). Moreover, of the top 50 taxa that were identified as having the highest probability of cooccurring with "CLA4," only 14 were known CLA producers (see Table S4 in the supplemental material). Future investigation into metabolic pathways in larger data sets would allow the identification of species that explain the higher abundance of "CLA4" in consumers than in nonconsumers.

This is to our knowledge the largest study of the association between fermented food (specifically, fermented plant) consumption and the human gut microbiome, with nearly 7,000 individuals at one time point and over 100 individuals across 4 weeks of sampling. We took a multi-omic approach—a combination of 16S rRNA sequencing, shotgun metagenomics, and mass spectrometry—coupled with state-of-the-art tools to evaluate the data. We find that the consumption of fermented plants and, more

broadly, fermented foods is associated with quite subtle microbiome variation in healthy individuals. While this explorative study provides the foundation for more-directed research, such as randomized placebo-controlled studies, it has some limitations, particularly that consumers were categorized according to self-reported frequency of fermented plant consumption. First, self-reported dietary information can be flawed with measurement errors (47). Second, although our data suggest that fermented plant consumption may be a reasonable proxy for consumption of fermented food more generally, they do not explicitly take into account other food types, such as fermented dairy products. Additionally, this study is mostly limited to participants living in the United States, who may consume a lower diversity of fermented foods than populations living in other countries; expanding this study to a wider range of populations would allow us to capture a greater diversity of fermented food types and associated microbial communities. Due to a combination of these factors, we may be underestimating the potential effects of fermented food consumption on the gut microbiome. Yet notably, the recovery of LAB and fermented-food-derived microbes in the stool of self-reported consumers suggests that data from stool may be used to help verify the reliability of self-reported dietary information. It would therefore be of great relevance to evaluate not only the associations between specific types of food and the microbiome but also our ability to detect consumption of specific fermented foods in future studies.

MATERIALS AND METHODS

Participant recruitment, sample processing, and sample selection. This research was performed in accordance with the University of Colorado Boulder's Institutional Review Board protocol number 12-0582 and the University of California San Diego's Human Research Protection Program protocol number 141853. In order to investigate the effect of fermented plant and food consumption on the gut microbiome, a retrospective analysis was performed on the American Gut Project data set (25). An additional cohort of 115 subjects was recruited to explore the effect of fermented food consumption or nonconsumption over a period of 4 weeks; the samples from the longitudinal cohort were processed and sequenced in accordance with AGP protocol and integrated into the AGP data set. The time point with the highest read count from each of the 115 recruited individuals was added to the concurrent cross-sectional assessment. The longitudinal cohort also responded to a specific fermented food questionnaire.

The entire AGP data set was subset using the metadata version accessed 8 August 2019 for stool samples from adult participants (age >19 and <70 years) who answered the "fermented plant frequency" question from the AGP questionnaire. Participants were excluded if they took antibiotics in the last year or if they had outlier values for their body mass index (<15 or >50), height (<48 cm or >210 cm), or weight (<2.5 kg or >200 kg). If biological replicates were present, the replicate with the lower number of reads was removed (with the exception of the 115 participants who constitute the longitudinal cohort). Based on the AGP questionnaire, participants were considered consumers if they reported "daily," "frequent" and "occasional" fermented plant consumption (i.e., >1 to 2 times per week) and nonconsumers if they reported "rarely" and "never."

Diet quality and intake assessment. Overall diet quality was assessed by the Healthy Eating Index 2010 (HEI-2010) as described elsewhere (48). Briefly, the HEI-2010 is a valid, reliable measure of diet quality that assesses how an individual's diet pattern adheres to the 2010 to 2015 Dietary Guidelines for Americans (DGA). HEI-2010 includes 12 dietary components, nine of which are classified as "adequacy" components that should be included regularly in the diet (total fruit, whole fruit, total vegetables, greens and beans, whole grains, dairy, total protein foods, seafood and plant proteins, and fatty acids), and 3 "moderation" components (refined grains, sodium, and empty calories) that should be limited in the diet. Individual dietary components are scored from 0 to 5, 10, or 20 points with maximum points indicating higher consumption of adequacy components and lower consumption of moderation components. Total HEI-2010 scores (range: 0 to 100) were calculated as the sum of the 12 components with a higher total score indicating better/optimal diet quality and greater adherence to the DGA. HEI-2010 scores, as well as total energy, carbohydrate, fat, protein, and fiber intake, were calculated from individuals in the AGP cohort who completed the VioScreen food frequency questionnaire (FFQ). We compared the total HEI score and mean nutrient intakes between consumers and nonconsumers using the Mann-Whitney U test.

Daily total consumption of CLA and LA (grams/day) was estimated from the VioScreen FFQ reports. Total CLA consumption was deduced from the following food sources: beef and other meat such as fish and turkey, full-fat dairy products (e.g., milk, butter, cheese, and yogurt), and eggs. Total LA consumption was obtained from the following reported foods: vegetable oil (e.g., canola and olive), salad dressings containing vegetable oils, butter, eggs, meat (beef, chicken, turkey, and pork), potatoes (e.g., French fries/fried white potatoes, and potato chips), nuts, nut butters and seeds, mixed Mexican dishes, and meat dishes such as stews and casseroles.

16S rRNA gene sequencing. DNA extraction and 16S rRNA amplicon sequencing were done using Earth Microbiome Project (EMP) standard protocols (http://www.earthmicrobiome.org/protocols-and -standards/16s). DNA was extracted with the Qiagen MagAttract PowerSoil DNA kit as previously

described (49). Amplicon PCR was performed on the V4 region of the 16S rRNA gene using the primer pair 515f-806r with Golay error-correcting barcodes on the reverse primer. Amplicons were barcoded and pooled in equal concentrations for sequencing. The amplicon pool was purified with the Mo Bio UltraClean PCR cleanup kit and sequenced on the Illumina MiSeq sequencing platform. Based on the filtering noted above, a feature table representing the 16S V4 rRNA gene sequence data was obtained from Qiita (50) using redbiom (51) from the Deblur-Illumina-16S-V4-150nt-780653 context. This table was composed of 8,513 samples. Prior to extraction from Qiita, the AGP data had been trimmed to 150 bases and processed using Deblur v1.0.4 (52) using the Qiita default parameters (i.e., setting –min-reads 1) to generate sOTUs. Technical replicates of samples were excluded in order to keep only the most-sequenced version of each sample. After previously recognized bloom sequences were removed (53), samples with fewer than 1,500 reads were omitted. Taxonomies for sOTUs were assigned using the sklearn-based taxonomy classifier trained on the Greengenes reference database 13_8 (54) clustered at 99% similarity (feature classifier plug-in of QIIME 2 v2019.1 [55]). The sOTU table was rarefied to a depth of 1,500 sequences/sample to control for sequencing effort (56) and sOTUs totaling 5 reads across samples. The deblurred sequence fragments were inserted into the Greengenes 13_8 phylogenetic tree using SATé-enabled phylogenetic placement (57, 58).

16S marker gene data analysis. QIIME 2 v2019.1 (55) was used to generate pairwise unweighted and weighted UniFrac distances (51, 59). Between-group differences based on these distances were tested using PERMANOVA (60) and permuted t tests in QIIME 2. Alpha diversity (Faith's PD [30], Shannon diversity [29], and observed OTU richness) between consumers and nonconsumers (as a whole and when stratified by consumption frequency) was generated using QIIME 2 (55) and compared with a Kruskal-Wallis test. Wilcoxon signed-rank (61) and Mann-Whitney U tests were used to assess alpha diversity between successive time points within consumers and nonconsumers and within time point between consumers and nonconsumers in the longitudinal cohort, respectively. Songbird v1.0.1 (27) was used to identify feature ranks corresponding to consumers and nonconsumers (parameters: –epochs 5000 –batch-size 5 –learning-rate 1e−4 –min-sample-count 1000 –min-feature-count 0 –num-random-test-examples 10), and Qurro v0.4.0 (28) was used to compute log ratios of these ranked features. t tests and Cohen's D were calculated to assess the significance (alpha = 0.05) and effect size of the log ratios. The stability of the participants' microbiomes was assessed by comparing sample log ratios in consecutive time points, for both the 16S and metabolomic data sets. The 40 highest- and lowest-ranked features were used in order to compute enough log ratios for Spearman's rank correlation coefficients across all samples and for ordinary least squares regression (Fig. S5).

LC-MS/MS data acquisition. The untargeted metabolomics analysis using high-performance liquid chromatography–tandem mass spectrometry (HPLC-MS) was carried out as described previously (25). The chromatography was performed on a Dionex UltiMate 3000 Thermo Fisher Scientific high-performance liquid chromatography system (Thermo Fisher Scientific, Waltham, MA) coupled to a Bruker Impact HD quadrupole time of flight (qTOF) mass spectrometer. The chromatographic separation was carried out on a reverse-phase (RP) Kinetex C_{18} 1.7-μm, 100-Å ultrahigh-performance liquid chromatography (UHPLC) column (50 mm by 2.1 mm) (Phenomenex, Torrance, CA), held at 40°C during analysis. A total of 5 μl of each sample was injected. Mobile phase A was water, and mobile phase B was acetonitrile, both with added 0.1% (vol/vol) formic acid. The solvent gradient table was set as follows: initial mobile phase composition was 5% B for 1 min, increased to 40% B over 1 min and then to 100% B over 6 min, held at 100% B for 1 min, and decreased back to 5% B in 0.1 min, followed by a washout cycle and equilibration for a total analysis time of 13 min. The scanned m/z range was 80 to 2,000, the capillary voltage was 4,500 V, the nebulizer gas pressure was 2×10^5 Pa, the drying gas flow rate was 9 liters/min, and the temperature was 200°C. Each full MS scan was followed by tandem MS (MS/MS) using collision-induced dissociation (CID) fragmentation of the seven most abundant ions in the spectrum. For MS/MS, the collision cell collision energy was set at 3 eV and the collision energy was stepped 50%, 75%, 150%, and 200% to obtain optimal fragmentation for differentially sized ions. The scan rate was 3 Hz. An HP-921 lock mass compound was infused during the analysis to carry out postprocessing mass correction. To determine the specific isomer of the annotations for octadecadienoic acid isomers, authentic standards for linoleic acid (LA; Spectrum Laboratory Products, Inc., USA) and conjugated linoleic acid (CLA; mixture of 4 isomers: 9,11 and 10,12 isomers, E and Z) (Sigma-Aldrich, USA) were compared by retention times (RTs) and MS/MS spectra. This brings these annotations to the level 1 identifications (authentic compound was analyzed under identical experimental conditions with orthogonal physical property compared).

LC-MS/MS data analysis. The collected data were processed as described in reference 62. Briefly, the feature tables were obtained using MZmine2 (63). The collected HPLC-MS raw data were converted from Bruker's .d to .mzXML format. The data were then batch processed with the following settings for each step: (i) mass detection, noise level of 1,000, chromatogram builder, minimum time span of 0.01 min, minimum peak height of 3,000, and m/z tolerance of 0.1 m/z or 20 ppm; (ii) chromatogram deconvolution—baseline cutoff, minimum peak height of 3,000, peak duration range of 0.01 to 3.00 min, and baseline level of 300; (iii) deisotopization—isotopic peak grouper, m/z tolerance of 0.1 m/z or 20 ppm, RT tolerance of 0.1 min, and maximum charge of 4; (iv) peak alignment—join aligner, m/z tolerance of 0.1 m/z or 20 ppm, weight for m/z 75, weight for RT 25, and RT tolerance of 0.1 min; and (v) peak filtering—peak list raw filter, minimum peak in a row of 3 and minimum peak in an isotope pattern of 2.

The metadata were added into the resulting extracted feature table and used as input for the MetaboAnalyst software (64, 65). The feature tables were filtered with interquantile ranges to remove outliers, and the data were imputed, normalized by the quantile normalization, and autoscaled (mean centering and dividing by the standard deviation for each feature). Partial least-squares discriminant

analysis (PLS-DA) was used to explore and visualize variance within data and differences among experimental categories. The CLA and LA metabolite features were identified manually based on GNPS (66) and MZmine 2 (63) processing pipelines (see link below to feature-based molecular networking). The Wilcoxon rank sum test (Mann-Whitney U test) was used to assess the significance of difference between the consumers and nonconsumers for the levels of identified CLA and LA metabolites (alpha = 0.05).

The annotations and visualizations of chemical distributions were explored on GNPS using molecular networking (66) as follows. MS/MS spectra were window filtered by choosing only the top 6 peaks in the 50-Da window throughout the spectrum. The MS spectra were then clustered with a parent mass tolerance of 0.02 Da and an MS/MS fragment ion tolerance of 0.02; consensus spectra that contained fewer than 4 spectra were discarded. The network was created with edges filtered to have a cosine score above 0.65 and more than 5 matched peaks. The edges between two nodes are kept in the network if and only if each of the nodes appears in each other's respective top 10 most similar nodes. The required library matches were set to have a score above 0.7 and at least 6 matched peaks when searching the spectra in the network against GNPS spectral libraries. All resulting annotations are at level 2/3 according to the proposed minimum standards in metabolomics (67). The GNPS results are located at https://gnps .ucsd.edu/ProteoSAFe/status.jsp?task=420a545b5b164d10a20f62c0ec0ce7e7. Feature-based molecular network (68) results can be found at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=9ce1517e83a94 d9a8cd9d79f3e16eea0. The CLA and LA metabolite features were initially identified based on GNPS library search (66), and then their annotation was further confirmed via use of authentic standards. The Wilcoxon rank sum test was used to assess the significance of difference between the consumers and nonconsumers for the levels of identified CLA and LA metabolites (alpha = 0.05).

Metagenomic sequencing. Extracted DNA was quantified with the PicoGreen double-stranded DNA (dsDNA) assay kit, and 5 ng of input, or a maximum of 3.5 μl, genomic DNA (gDNA) was used in a 1:10 miniaturized Kapa HyperPlus protocol. Per-sample libraries were quantified and pooled at equal nano-molar concentrations. The pooled library was cleaned with the QIAquick PCR purification kit and size selected for fragments between 300 and 700 bp on the Sage Science PippinHT. The pooled library was sequenced as a paired-end 150-cycle run on an Illumina HiSeq2500 v2 in Rapid Run mode at the UCSD IGM Genomics Center, with a target depth of ca. 20 million reads per sample. The sequencing adapter and short reads were first removed using Atropos v1.1.21 (-q 15 –minimum-length 100 –pair-filter any) as well as reads aligning to the human genome using bowtie2 (–very-sensitive). The pass-filter reads were then concatenated per sample, excluding 1 biological duplicate and 8 samples from participants exposed to antibiotics, in order to obtain 91 pairs of fastq files.

Metagenomic data analysis. On each separate sample fastq file, paired-end reads were merged using FLASH v1.2.11 (69) and then processed for taxonomic profiling using SHOGUN v1.0.6 (70) with Bowtie 2 v2.3.4.3 (71) to align reads to the 85,626 prokaryotic genomes covering 12,977 species from the NCBI RefSeq database release 82 (72). The read counts for the genome features identified in each sample were merged into one genome-per-sample table that was then filtered to keep genomes with a per-sample relative mapped read abundance of at least 0.01%. The features labeled at the subspecies level were sum collapsed at the species level; taxonomy was used as a proxy for a phylogeny. As with the 16S cross-sectional data, Songbird (Songbird v1.0.1 [27]) was used for regression modeling on our binary fermented consumption variable to identify features associated with consumption and nonconsumption (parameters as above). Qurro v0.4.0 (28) was used to compute log ratios of these ranked features. t tests and Cohen's D were calculated to assess the significance (alpha = 0.05) and effect size of the log ratios.

Multi-omics data analysis. In order to identify microbial features associated with fermented food consumption and the metabolites they might be producing, we measured probabilities of cooccurrence between observed species (based on metagenomic data) and either all metabolites, or a set of five linoleic and isomers of conjugated linoleic acids discernible in the data (as informed by the metabolomic analysis). For this analysis, we used mmvec v1.0.2 (36), a neural network solution inspired from natural language processing, to build a log-transformed conditional probability matrix from each cross-omics feature pair and apply singular value decomposition in order to represent cooccurrence in the form of biplots. We chose the model where accuracy was highest for different initialization conditions for the gradient descent algorithm (–batch-size of 1,000, 2,500, and 5,000 and –learn-rate of 1e−4 and 1e−5), with low cross-validation error and model likelihood. To evaluate the fitness of the mmvec microbe-metabolite interactions, we compared the latent representation to the observed Songbird differentials. The relationship between the microbial first principal component learned from mmvec and the log fold change of the microbes between fermented food consumption was significantly negatively correlated (Pearson's $r = -0.651$, $P = 4.63e-22$, $n = 249$ microbes; Fig. S2C), suggesting that the mmvec microbe-metabolite relationship to fermented food consumption is a valid comparison. We used EMPeror v2019.1.0 (73) to visualize feature-feature biplots along with overlying genomedifferential abundance ranks for our fermented food consumption model.

Data availability. The data generated in this study are available publicly in Qiita under the study ID 10317. Sequence data associated with this study can be found under EBI accession ERP012803. The metabolomics analysis is available at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task= 420a545b5b164d10a20f62c0ec0ce7e7 (classical molecular networking) and https://gnps.ucsd.edu/ ProteoSAFe/status.jsp?task=9ce1517e83a94d9a8cd9d79f3e16eea0 (feature-based molecular network-ing). All of the raw data are publicly available at the UCSD Center for Computational Mass Spectrometry (data set ID MassIVE MSV000081171, https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=9996246aab 414427a80bb5a451ec3c3d).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 1.9 MB.

FIG S2, PDF file, 0.5 MB.

FIG S3, PDF file, 1 MB.

FIG S4, PDF file, 0.5 MB.

FIG S5, PDF file, 1.1 MB.

FIG S6, PDF file, 0.8 MB.

TABLE S1, PDF file, 0.1 MB.

TABLE S2, PDF file, 0.02 MB.

TABLE S3, PDF file, 0.1 MB.

TABLE S4, PDF file, 0.1 MB.

REFERENCES

1. Bourdichon F, Casaregola S, Farrokh C, Frisvad JC, Gerds ML, Hammes WP, Harnett J, Huys G, Laulund S, Ouwehand A, Powell IB, Prajapati JB, Seto Y, Ter Schure E, Van Boven A, Vankerckhoven V, Zgoda A, Tuijtelaars S, Hansen EB. 2012. Food fermentations: microorganisms with technological beneficial use. Int J Food Microbiol 154:87–97. https://doi.org/10.1016/j.ijfoodmicro.2011.12.030.

2. Tamang JP, Cotter PD, Endo A, Han NS, Kort R, Liu SQ, Mayo B, Westerik N, Hutkins R. 2020. Fermented foods in a global age: East meets West. Compr Rev Food Sci Food Saf 19:184–217. https://doi.org/10.1111/1541-4337.12520.

3. Tamang JP, Watanabe K, Holzapfel WH. 2016. Review: diversity of microorganisms in global fermented foods and beverages. Front Microbiol 7:377. https://doi.org/10.3389/fmicb.2016.00377.

4. Rezac S, Kok CR, Heermann M, Hutkins R. 2018. Fermented foods as a dietary source of live organisms. Front Microbiol 9:1785. https://doi.org/10.3389/fmicb.2018.01785.

5. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, Biddinger SB, Dutton RJ, Turnbaugh PJ. 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559–563. https://doi.org/10.1038/nature12820.

6. Marco ML, Heeney D, Binda S, Cifelli CJ, Cotter PD, Foligné B, Gänzle M, Kort R, Pasin G, Pihlanto A, Smid EJ, Hutkins R. 2017. Health benefits of fermented foods: microbiota and beyond. Curr Opin Biotechnol 44:94–102. https://doi.org/10.1016/j.copbio.2016.11.010.

7. Peters A, Krumbholz P, Jäger E, Heintz-Buschart A, Çakir MV, Rothemund S, Gaudl A, Ceglarek U, Schöneberg T, Stäubert C. 2019. Metabolites of lactic acid bacteria present in fermented foods are highly potent agonists of human hydroxycarboxylic acid receptor 3. PLoS Genet 15: e1008145. https://doi.org/10.1371/journal.pgen.1008283.

8. Mozaffarian D, Hao T, Rimm EB, Willett WC, Hu FB. 2011. Changes in diet and lifestyle and long-term weight gain in women and men. N Engl J Med 364:2392–2404. https://doi.org/10.1056/NEJMoa1014296.

9. Chen M, Sun Q, Giovannucci E, Mozaffarian D, Manson JAE, Willett WC, Hu FB. 2014. Dairy consumption and risk of type 2 diabetes: 3 cohorts of US adults and an updated meta-analysis. BMC Med 12:215. https://doi.org/10.1186/s12916-014-0215-1.

10. Park S, Bae JH. 2016. Fermented food intake is associated with a reduced likelihood of atopic dermatitis in an adult population (Korean National Health and Nutrition Examination Survey 2012-2013). Nutr Res 36:125–133. https://doi.org/10.1016/j.nutres.2015.11.011.

11. Nozue M, Shimazu T, Sasazuki S, Charvat H, Mori N, Mutoh M, Sawada N, Iwasaki M, Yamaji T, Inoue M, Kokubo Y, Yamagishi K, Iso H, Tsugane S. 2017. Fermented soy product intake is inversely associated with the development of high blood pressure: the Japan Public Health Center-Based Prospective Study. J Nutr 14:1749–1756. https://doi.org/10.3945/jn.117.250282.

12. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L, Li H, Bushman FD, Lewis JD. 2011. Linking long-term dietary patterns with gut microbial enterotypes. Science 334:105–108. https://doi.org/10.1126/science.1208344.

13. Muegge BD, Kuczynski J, Knights D, Clemente JC, González A, Fontana L, Henrissat B, Knight R, Gordon JI. 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science 332:970–974. https://doi.org/10.1126/science.1198719.

14. Duncan SH, Belenguer A, Holtrop G, Johnstone AM, Flint HJ, Lobley GE. 2007. Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces. Appl Environ Microbiol 73:1073–1078. https://doi.org/10.1128/AEM.02340-06.

15. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. 2006. Microbial ecology: human gut microbes associated with obesity. Nature 444:1022–1023. https://doi.org/10.1038/4441022a.

16. Walker AW, Ince J, Duncan SH, Webster LM, Holtrop G, Ze X, Brown D, Stares MD, Scott P, Bergerat A, Louis P, McIntosh F, Johnstone AM, Lobley GE, Parkhill J, Flint HJ. 2011. Dominant and diet-responsive groups of bacteria within the human colonic microbiota. ISME J 5:220–230. https://doi.org/10.1038/ismej.2010.118.

17. Dimidi E, Cox SR, Rossi M, Whelan K. 2019. Fermented foods: definitions and characteristics, impact on the gut microbiota and effects on gastrointestinal health and disease. Nutrients 11:E1806. https://doi.org/10.3390/nu11081806.

18. Grieneisen LE, Blekhman R. 2018. Crowdsourcing our national gut. mSystems 3:e00060-18. https://doi.org/10.1128/mSystems.00060-18.

19. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, Kurilshikov A, Bonder MJ, Valles-Colomer M, Vandeputte D, Tito RY, Chaffron S, Rymenans L, Verspecht C, Sutter LD, Lima-Mendez G, D'hoe K, Jonckheere K, Homola D, Garcia R, Tigchelaar EF, Eeckhaudt L, Fu J, Henckaerts L, Zhernakova A, Wijmenga C, Raes J. 2016. Population-level analysis of gut microbiome variation. Science 352:560–564. https://doi.org/10.1126/science.aad3503.

20. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, Mujagic Z, Vila AV, Falony G, Vieira-Silva S, Wang J, Imhann F, Brandsma E, Jankipersadsing SA, Joossens M, Cenit MC, Deelen P, Swertz MA, Weersma RK, Feskens EJM, Netea MG, Gevers D, Jonkers D, Franke L, Aulchenko YS, Huttenhower C, Raes J, Hofker MH, Xavier RJ, Wijmenga C, Fu J. 2016. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science 352:565–569. https://doi.org/10.1126/science.aad3369.

21. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, Costea PI, Godneva A, Kalka IN, Bar N, Shilo S, Lador D, Vila AV, Zmora N, Pevsner-Fischer M, Israeli D, Kosower N, Malka G, Wolf BC, Avnit-Sagi T, Lotan-Pompan M, Weinberger A, Halpern Z, Carmi S, Fu J, Wijmenga C, Zhernakova A, Elinav E, Segal E. 2018. Environment dominates over host genetics in shaping human gut microbiota. Nature 555:210–215. https://doi.org/10.1038/nature25973.

22. Centers for Disease Control and Prevention. 2018. National Health and Nutrition Examination Survey data, 2011–2012. Centers for Disease Control and Prevention, Atlanta, GA.

23. Wang DD, Leung CW, Li Y, Ding EL, Chiuve SE, Hu FB, Willett WC. 2014. Trends in dietary quality among adults in the United States, 1999 through 2010. JAMA Intern Med 174:1587–1595. https://doi.org/10.1001/jamainternmed.2014.3422.

24. Drewnowski A, Aggarwal A, Cook A, Stewart O, Moudon AV. 2016. Geographic disparities in Healthy Eating Index scores (HEI-2005 and 2010) by residential property values: findings from Seattle Obesity Study (SOS). Prev Med 83:46–55. https://doi.org/10.1016/j.ypmed.2015.11.021.

25. McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, Aksenov AA, Behsaz B, Brennan C, Chen Y, DeRight Goldasich L, Dorrestein PC, Dunn RR, Fahimipour AK, Gaffney J, Gilbert JA, Gogul G, Green JL, Hugenholtz P, Humphrey G, Huttenhower C, Jackson MA, Janssen S, Jeste DV, Jiang L, Kelley ST, Knights D, Kosciolek T, Ladau J, Leach J, Marotz C, Meleshko D, Melnik AV, Metcalf JL, Mohimani H, Montassier E, Navas-Molina J, Nguyen TT, Peddada S, Pevzner P, Pollard KS, Rahnavard G, Robbins-Pianka A, Sangwan N, Shorenstein J, Smarr L, Song SJ, Spector T, Swafford AD, Thackray VG, Thompson LR, Tripathi A, Vázquez-Baeza Y, Vrbanac A, Wischmeyer P, Wolfe E, Zhu Q, Knight R, Mann AE, Amir A, Frazier A, Martino C, Lebrilla C, Lozupone C, Lewis CM, Raison C, Zhang C, Lauber CL, Warinner C, Lowry CA, Callewaert C, Bloss C, Willner D, Galzerani DD, Gonzalez DJ, Mills DA, Chopra D, Gevers D, Berg-Lyons D, Sears DD, Wendel D, Lovelace E, Pierce E, TerAvest E, Bolyen E, Bushman FD, Wu GD, Church GM, Saxe G, Holscher HD, Ugrina I, German JB, Caporaso JG, Wozniak JM, Kerr J, Ravel J, Lewis JD, Suchodolski JS, Jansson JK, Hampton-Marcell JT, Bobe J, Raes J, Chase JH, Eisen JA, Monk J, Clemente JC, Petrosino J, Goodrich J, Gauglitz J, Jacobs J, Zengler K, Swanson KS, Lewis K, Mayer K, Bittinger K, Dillon L, Zaramela LS, Schriml LM, Dominguez-Bello MG, Jankowska MM, Blaser M, Pirrung M, Minson M, Kurisu M, Ajami N, Gottel NR, Chia N, Fierer N, White O, Cani PD, Gajer P, Strandwitz P, Kashyap P, Dutton R, Park RS, Xavier RJ, Mills RH, Krajmalnik-Brown R, Ley R, Owens SM, Klemmer S, Matamoros S, Mirarab S, Moorman S, Holmes S, Schwartz T, Eshoo-Anton TW, Vigers T, Pandey V, Treuren WV, Fang X, Zech Xu Z, Jarmusch A, Geier J, Reeve N, Silva R, Kopylova E, Nguyen D, Sanders K, Salido Benitez RA, Heale AC, Abramson M, Waldispühl J, Butyaev A, Drogaris C, Nazarova E, Ball M, Gunderson B. 2018. American Gut: an open platform for citizen science microbiome research. mSystems 3:e00031-18. https://doi.org/10.1128/mSystems.00031-18.

26. Lozupone C, Hamady M, Knight R. 2006. UniFrac—an online tool for comparing microbial community diversity in a phylogenetic context. BMC Bioinformatics 7:371. https://doi.org/10.1186/1471-2105-7-371.

27. Morton JT, Marotz C, Washburne A, Silverman J, Zaramela LS, Edlund A, Zengler K, Knight R. 2019. Establishing microbial composition measurement standards with reference frames. Nat Commun 10:2719. https://doi.org/10.1038/s41467-019-10656-5.

28. Fedarko MW, Martino C, Morton JT, Marotz CA, Minich JJ, Allen EE, Knight R. 2019. Qurro. Zenodo https://doi.org/10.5281/zenodo.3369454.

29. Shannon CE. 1948. A mathematical theory of communication. Bell Syst Tech J 27:379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

30. Faith DP. 1992. Conservation evaluation and phylogenetic diversity. Biol Conserv 61:1–10. https://doi.org/10.1016/0006-3207(92)91201-3.

31. Jiang L, Vazquez-Baeza Y, Gonzalez A, Natarajan L, Knight R, Thompson WK. 2019. Bayesian sparse functional principal components analysis models dynamic temporal changes in longitudinal microbiome studies, 1836–1853. In JSM Proceedings, Statistics in Epidemiology Section. American Statistical Association, Alexandria, VA. https://ww2.amstat.org/membersonly/proceedings/2019/data/assets/pdf/1199578.pdf.

32. Wang X, Xiao J, Jia Y, Pan Y, Wang Y. 2018. Lactobacillus kefiranofaciens, the sole dominant and stable bacterial species, exhibits distinct morphotypes upon colonization in Tibetan kefir grains. Heliyon 4:e00649. https://doi.org/10.1016/j.heliyon.2018.e00649.

33. Fröhlich-Wyder MT, Guggisberg D, Badertscher R, Wechsler D, Wittwer A, Irmler S. 2013. The effect of Lactobacillus buchneri and Lactobacillus parabuchneri on the eye formation of semi-hard cheese. Int Dairy J 33:120–128. https://doi.org/10.1016/j.idairyj.2013.03.004.

34. Giraffa G. 2014. Lactobacillus helveticus: importance in food and health. Front Microbiol 5:338. https://doi.org/10.3389/fmicb.2014.00338.

35. McLeod A, Zagorec M, Champomier-Vergès MC, Naterstad K, Axelsson L. 2010. Primary metabolism in Lactobacillus sakei food isolates by proteomic analysis. BMC Microbiol 10:120. https://doi.org/10.1186/1471-2180-10-120.

36. Morton JT, Aksenov AA, Nothias LF, Foulds JR, Quinn RA, Badri MH, Swenson TL, Van Goethem MW, Northen TR, Vazquez-Baeza Y, Wang M, Bokulich NA, Watters A, Song SJ, Bonneau R, Dorrestein PC, Knight R. 2019. Learning representations of microbe–metabolite interactions. Nat Methods 16:1306–1314. https://doi.org/10.1038/s41592-019-0616-3.

37. Devillard E, McIntosh FM, Duncan SH, Wallace RJ. 2007. Metabolism of linoleic acid by human gut bacteria: different routes for biosynthesis of conjugated linoleic acid. J Bacteriol 189:2566–2570. https://doi.org/10.1128/JB.01359-06.

38. Sieber R, Collomb M, Aeschlimann A, Jelen P, Eyer H. 2004. Impact of microbial cultures on conjugated linoleic acid in dairy products—a review. Int Dairy J 14:1–15. https://doi.org/10.1016/S0958-6946(03)00151-1.

39. Yang B, Gao H, Stanton C, Ross RP, Zhang H, Chen YQ, Chen H, Chen W. 2017. Bacterial conjugated linoleic acid production and their applications. Prog Lipid Res 68:26–36. https://doi.org/10.1016/j.plipres.2017.09.002.

40. Van Nieuwenhove CP, Teran V, Gonzalez SN. 2012. Conjugated linoleic and linolenic acid production by bacteria: development of functional foods. In Rigobelo E (ed), Probiotics. IntechOpen, London, United Kingdom.

41. Koba K, Yanagita T. 2014. Health benefits of conjugated linoleic acid (CLA). Obes Res Clin Pract 8:e525–532. https://doi.org/10.1016/j.orcp.2013.10.001.

42. Dilzer A, Park Y. 2012. Implication of conjugated linoleic acid (CLA) in human health. Crit Rev Food Sci Nutr 52:488–513. https://doi.org/10.1080/10408398.2010.501409.

43. Kishino S, Takeuchi M, Park SB, Hirata A, Kitamura N, Kunisawa J, Kiyono H, Iwamoto R, Isobe Y, Arita M, Arai H, Ueda K, Shima J, Takahashi S, Yokozeki K, Shimizu S, Ogawa J. 2013. Polyunsaturated fatty acid saturation by gut lactic acid bacteria affecting host lipid composition. Proc Natl Acad Sci U S A 110:17808–17813. https://doi.org/10.1073/pnas.1312937110.

44. Terán V, Pizarro PL, Zacarías MF, Vinderola G, Medina R, Van Nieuwenhove C. 2015. Production of conjugated dienoic and trienoic fatty acids by lactic acid bacteria and bifidobacteria. J Funct Foods 19(Part A):417–425. https://doi.org/10.1016/j.jff.2015.09.046.

45. Chung SH, Kim IH, Park HG, Kang HS, Yoon CS, Jeong HY, Choi NJ, Kwon EG, Kim YJ. 2008. Synthesis of conjugated linoleic acid by human-derived Bifidobacterium breve LMC 017: utilization as a functional starter culture for milk fermentation. J Agric Food Chem 56:3311–3316. https://doi.org/10.1021/jf0730789.

46. Coakley M, Ross RP, Nordgren M, Fitzgerald G, Devery R, Stanton C. 2003. Conjugated linoleic acid biosynthesis by human-derived Bifidobacterium species. J Appl Microbiol 94:138–145. https://doi.org/10.1046/j.1365-2672.2003.01814.x.

47. Subar AF, Freedman LS, Tooze JA, Kirkpatrick SI, Boushey C, Neuhouser ML, Thompson FE, Potischman N, Guenther PM, Tarasuk V, Reedy J, Krebs-Smith SM. 2015. Addressing current criticism regarding the value of self-report dietary data. J Nutr 145:2639–2645. https://doi.org/10.3945/jn.115.219634.

48. Guenther PM, Kirkpatrick SI, Reedy J, Krebs-Smith SM, Buckman DW,

Dodd KW, Casavale KO, Carroll RJ. 2014. The Healthy Eating Index-2010 Is a valid and reliable measure of diet quality according to the 2010 Dietary Guidelines for Americans. J Nutr 144:399–407. https://doi.org/10.3945/jn.113.183079.

49. Marotz C, Amir A, Humphrey G, Gaffney J, Gogul G, Knight R. 2017. DNA extraction for streamlined metagenomics of diverse environmental samples. Biotechniques 62:290–293. https://doi.org/10.2144/000114559.

50. Gonzalez A, Navas-Molina JA, Kosciolek T, McDonald D, Vázquez-Baeza Y, Ackermann G, DeReus J, Janssen S, Swafford AD, Orchanian SB, Sanders JG, Shorenstein J, Holste H, Petrus S, Robbins-Pianka A, Brislawn CJ, Wang M, Rideout JR, Bolyen E, Dillon M, Caporaso JG, Dorrestein PC, Knight R. 2018. Qiita: rapid, web-enabled microbiome meta-analysis. Nat Methods 15:796–798. https://doi.org/10.1038/s41592-018-0141-9.

51. McDonald D, Kaehler B, Gonzalez A, DeReus J, Ackermann G, Marotz C, Huttley G, Knight R. 2019. redbiom: a rapid sample discovery and feature characterization system. mSystems 4:e00215-19. https://doi.org/10.1128/mSystems.00215-19.

52. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. 2017. Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems 2:e00191-16. https://doi.org/10.1128/mSystems.00191-16.

53. Amir A, McDonald D, Navas-Molina JA, Debelius J, Morton JT, Hyde E, Robbins-Pianka A, Knight R. 2017. Correcting for microbial blooms in fecal samples during room-temperature shipping. mSystems 2:e00199-16. https://doi.org/10.1128/mSystems.00199-16.

54. McDonald D, Price MN, Goodrich J, Nawrocki EP, Desantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J 6:610–618. https://doi.org/10.1038/ismej.2011.139.

55. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu YX, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver J, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–857. https://doi.org/10.1038/s41587-019-0209-9.

56. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, Lozupone C, Zaneveld JR, Vázquez-Baeza Y, Birmingham A, Hyde ER, Knight R. 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome 5:27. https://doi.org/10.1186/s40168-017-0237-y.

57. Mirarab S, Nguyen N, Warnow T. 2012. SEPP: SATé-enabled phylogenetic placement. Pac Symp Biocomput 2012:247–258. https://doi.org/10.1142/9789814366496_0024.

58. Janssen S, McDonald D, Gonzalez A, Navas-Molina JA, Jiang L, Xu ZZ, Winker K, Kado DM, Orwoll E, Manary M, Mirarab S, Knight R. 2018. Phylogenetic placement of exact amplicon sequences improves associations with clinical information. mSystems 3:e00021-18. https://doi.org/10.1128/mSystems.00021-18.

59. Lozupone C, Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 71:8228–8235. https://doi.org/10.1128/AEM.71.12.8228-8235.2005.

60. Anderson MJ. 2005. PERMANOVA. Permutational multivariate analysis of variance. University of Auckland, Auckland, New Zealand.

61. Wilcoxon F. 1945. Individual comparisons by ranking methods. Biom Bull 1(6):80–83. https://doi.org/10.2307/3001968.

62. Dührkop K, Fleischauer M, Ludwig M, Aksenov AA, Melnik AV, Meusel M, Dorrestein PC, Rousu J, Böcker S. 2019. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. Nat Methods 16:299–302. https://doi.org/10.1038/s41592-019-0344-8.

63. Pluskal T, Castillo S, Villar-Briones A, Orešič M. 2010. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics 11:395. https://doi.org/10.1186/1471-2105-11-395.

64. Xia J, Sinelnikov IV, Han B, Wishart DS. 2015. MetaboAnalyst 3.0—making metabolomics more meaningful. Nucleic Acids Res 43:W251–W257. https://doi.org/10.1093/nar/gkv380.

65. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, Wishart DS, Xia J. 2018. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Res 46(W1):W486–W494. https://doi.org/10.1093/nar/gky310.

66. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, Boya CAP, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard PM, Phapale P, Nothias LF, Alexandrov T, Litaudon M, Wolfender JL, Kyle JE, Metz TO, Peryea T, Nguyen DT, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson B, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol 34:828–837. https://doi.org/10.1038/nbt.3597.

67. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TWM, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reily MD, Thaden JJ, Viant MR. 2007. Proposed minimum reporting standards for chemical analysis. Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics 3:211–221. https://doi.org/10.1007/s11306-007-0082-2.

68. Nothias LF, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A, Protsyuk I, Ernst M, Tsugawa H, Fleischauer M, Aicheler F, Aksenov A, Alka O, Allard PM, Barsch A, Cachet X, Caraballo M, Silva RRD, Dang T, Garg N, Gauglitz JM, Gurevich A, Isaac G, Jarmusch AK, Kameník Z, Kang KB, Kessler N, Koester I, Korf A, Gouellec AL, Ludwig M, Christian MH, McCall LI, McSayles J, Meyer SW, Mohimani H, Morsy M, Moyne O, Neumann S, Neuweger H, Nguyen NH, Nothias-Esposito M, Paolini J, Phelan VV, Pluskal T, Quinn RA, Rogers S, Shrestha B, Tripathi A, van der Hooft JJJ, Vargas F, Weldon KC, Witting M, Yang H, Zhang Z, Zubeil F, Kohlbacher O, Böcker S, Alexandrov T, Bandeira N, Wang M, Dorrestein PC. 2019. Feature-based molecular networking in the GNPS analysis environment. bioRxiv https://doi.org/10.1101/812404 https://www.biorxiv.org/content/10.1101/812404v1.

69. Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957–2963. https://doi.org/10.1093/bioinformatics/btr507.

70. Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Gohl DM, Beckman KB, Knight R, Knights D. 2018. Evaluating the information content of shallow shotgun metagenomics. mSystems 3:e00069-18. https://doi.org/10.1128/mSystems.00069-18.

71. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.

72. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P,

Murphy TD, Pruitt KD. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44(D1):D733–D745. https://doi.org/10.1093/nar/gkv 1189.

73. Vázquez-Baeza Y, Pirrung M, Gonzalez A, Knight R. 2013. EMPeror: a tool for visualizing high-throughput microbial community data. Gigascience 2:16. https://doi.org/10.1186/2047-217X-2-16.

74. Zhu Q, Mai U, Pfeiffer W, Janssen S, Asnicar F, Sanders JG, Belda-Ferre P, Al-Ghalith GA, Kopylova E, McDonald D, Kosciolek T, Yin JB, Huang S, Salam N, Jiao JY, Wu Z, Xu ZZ, Cantrell K, Yang Y, Sayyari E, Rabiee M, Morton JT, Podell S, Knights D, Li WJ, Huttenhower C, Segata N, Smarr L, Mirarab S, Knight R. 2019. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. Nat Commun 10:5477. https://doi.org/10.1038/s41467-019-13443-4.