



Differences in Gut Metabolites and Microbial Composition and Functions between Egyptian and U.S. Children Are Consistent with Their Diets

V. Shankar,^a  M. Gouda,^b J. Moncivaiz,^a A. Gordon,^a N. V. Reo,^a L. Hussein,^b O. Paliy^a

Department of Biochemistry and Molecular Biology, Boonshoft School of Medicine, Wright State University, Dayton, Ohio, USA^a; Department of Human Nutrition, National Research Centre, Giza, Egypt^b

ABSTRACT Previous studies indicated that populations consuming a Mediterranean diet rich in fiber, vegetables, and fruits have a significantly lower risk of cardiovascular and metabolic diseases than populations of industrialized societies consuming diets enriched in processed carbohydrates, animal proteins, and fats. To explore the potential contributions of gut microbiota to the observed diet-related metabolic effects, we conducted an integrative analysis of distal gut microbiota composition and functions and intestinal metabolites in Egyptian and U.S. teenagers. All Egyptian gut microbial communities belonged to the *Prevotella* enterotype, whereas all but one of the U.S. samples were of the *Bacteroides* enterotype. The intestinal environment of Egyptians was characterized by higher levels of short-chain fatty acids, a higher prevalence of microbial polysaccharide degradation-encoding genes, and a higher proportion of several polysaccharide-degrading genera. Egyptian gut microbiota also appeared to be under heavier bacteriophage pressure. In contrast, the gut environment of U.S. children was rich in amino acids and lipid metabolism-associated compounds; contained more microbial genes encoding protein degradation, vitamin biosynthesis, and iron acquisition pathways; and was enriched in several protein- and starch-degrading genera. Levels of 1-methylhistamine, a biomarker of allergic response, were elevated in U.S. guts, as were the abundances of members of *Faecalibacterium* and *Akkermansia*, two genera with recognized anti-inflammatory effects. The revealed corroborating differences in fecal microbiota structure and functions and metabolite profiles between Egyptian and U.S. teenagers are consistent with the nutrient variation between Mediterranean and Western diets.

IMPORTANCE The human gastrointestinal microbiota functions as an important mediator of diet for host metabolism. To evaluate how consumed diets influence the gut environment, we carried out simultaneous interrogations of distal gut microbiota and metabolites in samples from healthy children in Egypt and the United States. While Egyptian children consumed a Mediterranean diet rich in plant foods, U.S. children consumed a Western diet high in animal protein, fats, and highly processed carbohydrates. Consistent with the consumed diets, Egyptian gut samples were enriched in polysaccharide-degrading microbes and end products of polysaccharide fermentation, and U.S. gut samples were enriched in proteolytic microbes and end products of protein and fat metabolism. Thus, the intestinal microbiota might be selected on the basis of the diets that we consume, which can open opportunities to affect gut health through modulation of gut microbiota with dietary supplementations.

KEYWORDS children, diet, metabolomics, metagenomics, microbiota, nutrition

Received 29 October 2016 Accepted 12 January 2017 Published 7 February 2017

Citation Shankar V, Gouda M, Moncivaiz J, Gordon A, Reo NV, Hussein L, Paliy O. 2017. Differences in gut metabolites and microbial composition and functions between Egyptian and U.S. children are consistent with their diets. mSystems 2:e00169-16. <https://doi.org/10.1128/mSystems.00169-16>.

Editor Chaysavanh Manichanh, Vall d'Hebron Research Institute

Copyright © 2017 Shankar et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to L. Hussein, dr.lailahussein@yahoo.com, or O. Paliy, oleg.paliy@wright.edu.

Throughout human evolution, dramatic shifts in the lifestyle and geographical distribution of the human species have led to periodic changes in consumed diets and nutritional intakes. When the original human societies were primarily hunter-gatherers, their diet was rich in lean wild animal meat and seafood as well as in what we now call dietary fiber from roots, tubers, fruits, and leafy vegetables (1, 2). The advent of agriculture about 12,000 years ago and the accompanying change from a nomadic to a sedentary lifestyle (3) led to a dramatic shift toward the consumption of large quantities of refined grains, seeds, and, eventually, simple sugars such as sucrose and fructose (2, 4). Over time, animal husbandry also developed and became an economically viable source of food, such that the diet of the modern industrialized societies now contains significant amounts of animal proteins and fats.

These shifts in dietary habits in different geographical regions have historically given rise to several different diet types. The typical “Western” diet consumed by the majority of populations in most industrialized countries is rich in animal proteins and fats, dairy products, and refined, starch-enriched grains, cereals, flour, and sugars (4–6). Modern agricultural and husbandry methods coupled with the ability to preserve perishable foods made these products profitable to produce on a large scale. The level of consumption of fruits, nuts, and vegetables in the Western diet is generally low, and intake of dietary fiber is well below recommended levels (7, 8). This diet, enriched in refined carbohydrates, animal fat, and protein, has been postulated to be one of the primary causes for the rising number of metabolic diseases in the industrialized countries (4, 9). In contrast, the “Mediterranean” diet is considered the standard of healthy nutrition and has been shown to be associated with an increased life span and a low incidence of metabolic and cardiovascular diseases (10, 11). Consumed by populations of the Mediterranean Sea region, it is high in fruits, vegetables, whole grain, beans, nuts, and plant fats, with a low fraction of meats and sweets (12, 13).

In addition to the well-recognized direct effects of consumed nutrients on human physiology, diet is also considered to be one of the main determinants of gut microbiota composition and diversity (14, 15). Switching from a diet rich in animal products to one high in fiber and plant foods rapidly changes gut microbiota in humans and animals (16, 17). Reciprocally, the gut microbiota plays a large and vital role in the biotransformation of consumed foods. A significant proportion of ingested foods escapes digestion and absorption in the small intestine and reaches the colon. These include dietary fiber (nonstarch polysaccharides), resistant starch, small amounts of simpler carbohydrates, and some proteins and fats, as well as bile acids and enzymes released in the small intestine (18, 19). Most of these unabsorbed compounds are fermented in the colon by gut microbes. The end products of microbial metabolism have been shown to have many positive (short-chain fatty acids [SCFAs]) as well as negative (trimethylamine, ammonia, hydrogen sulfide) effects on the host health (19, 20). Alterations of the gut microbial populations have been associated with the development of metabolic disorders such as obesity and type 2 diabetes (21, 22), and gut microbiota changes during childhood can have lifelong effects (23, 24).

In this study, we aimed to discover possible relationships between human gut microbiota and consumed diets. To achieve that goal, we compared fecal microbiota structures and functions as well as fecal metabolites in two cohorts of children: teenagers from the United States consuming a typical Western diet and population group age-matched Egyptians consuming a Mediterranean-type diet. Simultaneous analyses of microbial community membership and functional gene pool data combined with the quantification of metabolites in the same samples allowed integrative analysis of these data sets and revealed links between the gut microbiota and the intestinal environment.

RESULTS

Gut microbiota compositions differ between healthy U.S. and Egyptian teenagers. High-throughput 16S rRNA gene amplicon sequencing was used to analyze microbial composition in fresh fecal samples collected from 28 Egyptian teenagers and

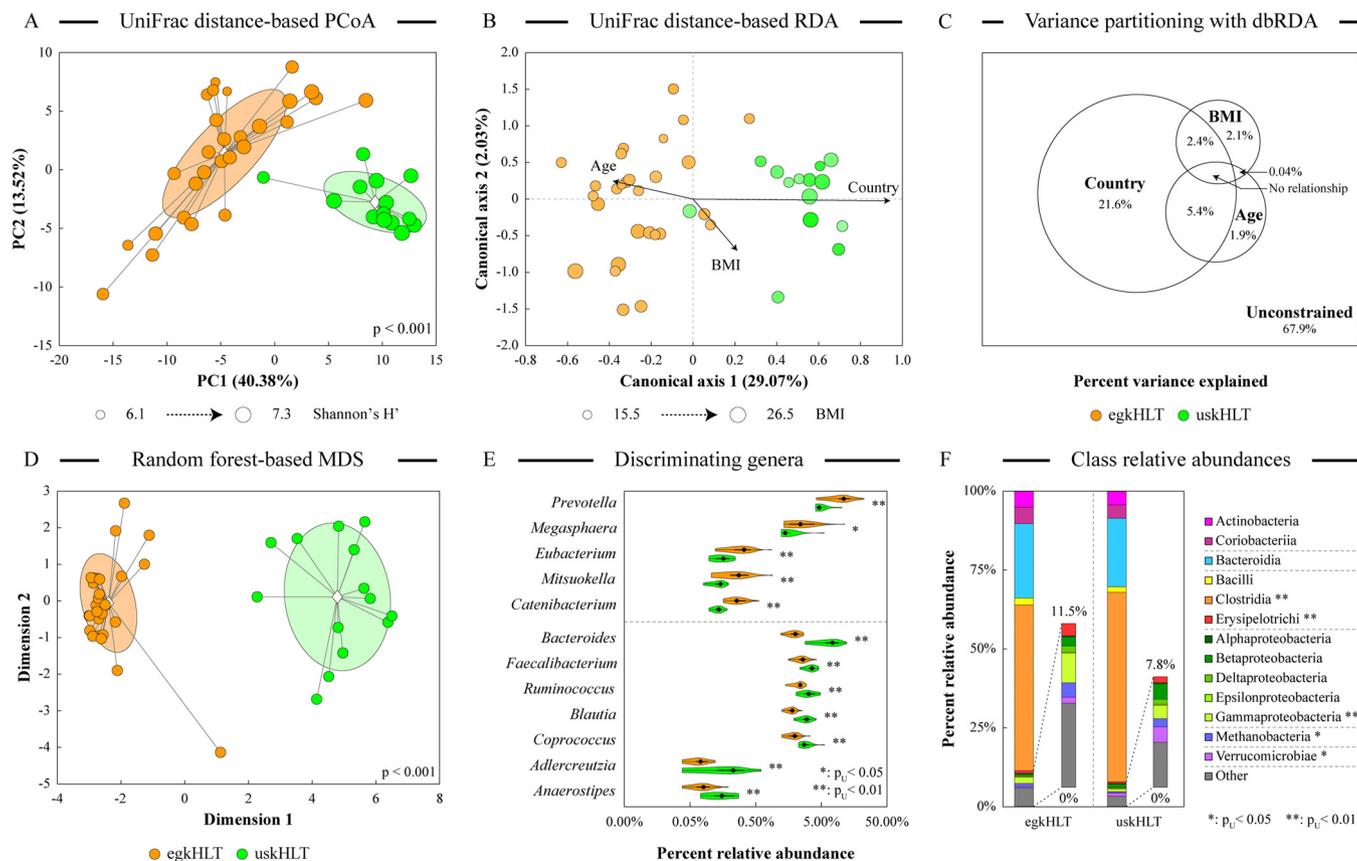


FIG 1 Comparison of distal gut microbiota composition between Egyptian and U.S. groups. (A and B) Sample similarity was assessed by unconstrained weighted UniFrac-based principal-coordinate analysis (A) and constrained weighted UniFrac-based redundancy analysis (B) run on the chord-transformed genus abundance data set. The statistical significance of group separation in PCoA is based on the Davies-Bouldin index. Group clouds represent areas of three standard errors around the group centroid (diamond), and dot sizes in PCoA are proportional to Shannon’s H’ alpha diversity values for that sample. Distance-based RDA used three explanatory variables (group, age, and body mass index). The age and the BMI of the subjects are represented by color gradients and the size of each dot, respectively. Arrows in the db-RDA biplot denote the magnitudes and directions of the effects of explanatory variables. (C) A variation partitioning diagram depicts the relative contributions of explanatory variables to the overall variability in the data set. (D) Results of random forest discriminant analysis of chord-transformed genus abundances were visualized through multidimensional scaling of the sample proximity matrix. The statistical significance of group separation is based on the Davies-Bouldin index. Group clouds represent areas of three standard errors around the group centroid (diamond). MDS, multidimensional scaling. (E) The relative abundances of the top 12 RF discriminatory genera are depicted on a violin plot. Each violin shows the density distribution of genus abundances among all samples in the group. (F) Community structure at the class level. Classes are ordered according to the phylum. Where shown, single asterisks (*) and double asterisks (**) indicate statistical significance of taxon abundance differences between two groups ($p_u < 0.05$ and $p_u < 0.01$, respectively) based on the FDR-adjusted Mann-Whitney *U* test.

14 age-matched teenagers from the Midwest region of United States. Exploratory principal-coordinate analysis (PCoA) using phylogenetically defined weighted UniFrac distance measure (25) distributed samples in the ordination space based largely on their group identity. Figure 1A shows PCoA results based on the analysis of the genus abundance data set; Fig. S1 in the supplemental material displays PCoA ordination results for the phylotype data set. We also calculated Bray-Curtis (BC) beta diversity distances among all samples. Consistent with the PCoA findings, average intersample BC distances between samples from healthy preadolescent and adolescent male volunteers from Giza, Egypt (designated egkHLT), and from Dayton, OH (designated uskHLT), were significantly larger than the intragroup distances (0.640 versus 0.569 and 0.576; $p_u < 0.001$). Using sample group, age, and body mass index (BMI) values as explanatory variables, we then conducted constrained ordination analysis using distance-based redundancy analysis (db-RDA) and a weighted UniFrac distance matrix (Fig. 1B). The three explanatory variables together explained 32.1% of the overall variability in the microbiota composition. Variation partitioning and sample dispersal in the constrained ordination space indicated that sample group assignment was the dominant gradient in the data set, accounting for 29.3% of the overall variability. In

contrast, age and BMI displayed smaller effects (Fig. 1C). A large portion of the overall microbiota composition variation was not accounted for by the available explanatory variables. This finding is consistent with the previously observed high interpersonal variability in human gut microbiota composition (26, 27).

Specific microbial genera are differentially abundant between Egyptian and U.S. teenagers. To identify specific genera that contributed to the observed separation of samples between groups in ordination analyses, random forest (RF) discriminant analysis was performed on the genus abundance data set. The RF model clearly separated the two groups of samples, as shown in Fig. 1D, and variable importance scores (calculated as the increase in model error due to the permutation of the variable) were used to define the major discriminatory genera (see Table S4 in the supplemental material) (28). Figure 1E displays the distribution of abundances of the top 12 discriminating genera among egkHLT and uskHLT samples; a list of all genus abundances together with the *P* values of Mann-Whitney *U* test for significant differences between groups is provided in Table S1. While many differences were observed at the genus level, such distinctions between groups were less pronounced at the class level. Levels of *Gammaproteobacteria* and *Methanobacteria* were statistically significantly higher in egkHLT, while levels of *Clostridia* and *Verrucomicrobia* were higher in uskHLT (see Fig. 1F and Table S2). Considering the differentially abundant genera, the higher abundance of *Prevotella* in egkHLT samples and the reciprocal higher abundance of *Bacteroides* in the uskHLT samples (both are members of class *Bacteroidia*) are consistent with several previous studies that indicated a higher prevalence of members of *Bacteroides* in samples from the United States, western Europe, and industrialized Asian countries than in samples from less-industrialized and more-rural populations (29–33). Many *Bacteroides* members can utilize proteins for growth (34, 35), which might explain their prevalence in subjects from developed countries consuming typical Western diets heavy in animal fats and proteins. On the other hand, *Prevotella* spp. are known degraders of xylan and other fibrous polysaccharides (36), which is consistent with their presence in ethnic groups where a large fraction of the diet is comprised of vegetables and grains (29, 30). Other genera enriched in the adolescent Egyptian gut included polysaccharide-degrading *Megasphaera*, *Eubacterium*, *Mitsuokella*, and *Catenibacterium* (37–39). *Catenibacterium* and *Mitsuokella* were also found in the stool of Bangladeshi children but not in the samples from the cohort of U.S. kids (31), and both *Catenibacterium* and *Eubacterium* were previously linked to the abundances of *Prevotella* (26, 32). The gut microbiota of Egyptian children was also enriched in several genera typically associated with pathogenicity and infections, including *Succinivibrio* and *Treponema* (see Table S1) (29, 30). The presence of *Treponema* might relate to the ability of some members of this genus to degrade xylan and cellulose (29, 33).

In comparison, species of several known starch-degrading genera, namely, *Ruminococcus*, *Coprococcus*, and *Blautia* (40), were more abundant in the stools of U.S. children, possibly due to the high prevalence of starch as a dietary polysaccharide in the Western diet (41). In addition, abundances of *Bilophila*, a genus that is associated with bile acids and high-fat diets (42), mucin-degrading genus *Akkermansia*, and clostridial genus *Faecalibacterium* were 1.5-fold to 4-fold higher in the uskHLT samples (see Table S1). *Akkermansia* and *Faecalibacterium* were previously shown to exert anti-inflammatory effects on the intestinal mucosa and adipose tissues (43, 44). Considering the significantly higher incidence of autoimmune diseases in Western populations (45), the higher abundance of this genus in the gut of U.S. teenagers might indicate a host-driven adaptation to elevated inflammatory levels.

Egyptian and U.S. gut microbiotas belong to different enterotypes. Following the recently described discovery of different enterotypes of human distal gut microbiota (26, 46, 47), we assessed if any such enterotypes could be revealed in the microbial communities profiled in the egkHLT and uskHLT samples. Calinski-Harabasz (CH) index values (Fig. S5) indicated that microbial data set can be optimally distributed into two enterotypes, and samples were separated into these enterotypes utilizing the parti-

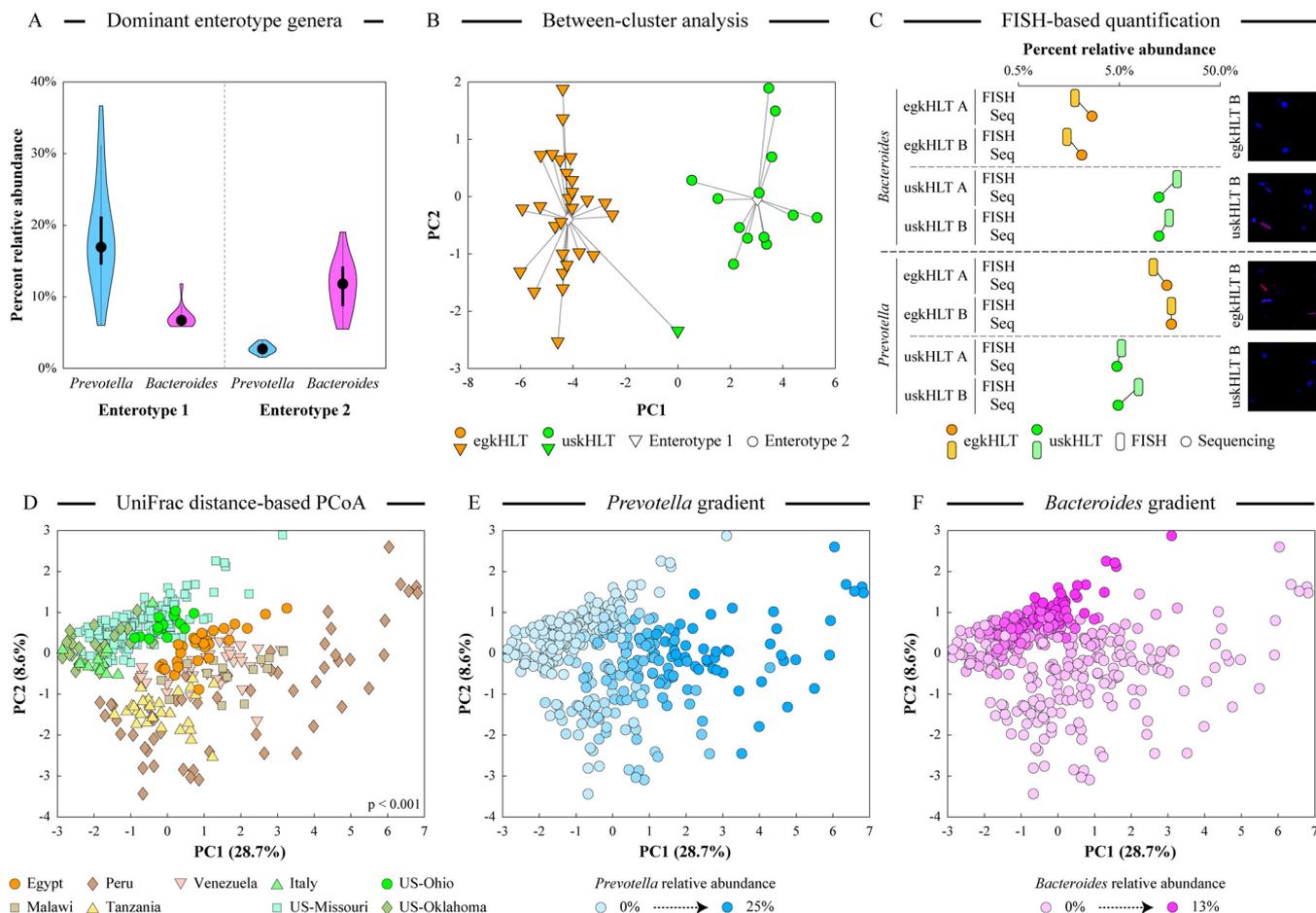


FIG 2 Enterotypes of the distal gut microbial profiles. (A) Relative abundances of the main enterotype drivers, *Prevotella* and *Bacteroides*, among the two identified clusters. Each violin shows the density distribution of genus abundance among all samples within the cluster (thickness of the violin), median value (black dot), and 25% to 75% range of values (black bar). (B) Sample clustering into different enterotypes based on the between-cluster ordination analysis. (C) Comparison of *Prevotella* and *Bacteroides* relative abundances in four samples based on sequencing (Seq) and fluorescent *in situ* hybridization (FISH) results. Representative fluorescent images are shown on the right; pink coloring corresponds to genus-specific fluorescent probes, and blue coloring represents DAPI DNA staining. (D) Results of weighted UniFrac distance-based PCoA performed using OTU abundances showing separation of fecal samples between industrialized and nonindustrialized countries. The statistical significance of group separation is based on the Davies-Bouldin index. Panels E and F display the same PCoA space, with relative abundances of *Prevotella* (E) and *Bacteroides* (F) overlaid as dot color gradients.

tioning around the medoid (PAM) clustering algorithm (48). With the exception of a single uskHLT sample which had an unusually large representation of *Prevotella* (15.8% compared with 7.7% average for uskHLT), all uskHLT samples clustered into the “*Bacteroides*” enterotype, and all egkHLT samples clustered into the “*Prevotella*” enterotype (see Fig. 2A and B). Previously, in a diet assessment study, the *Bacteroides* enterotype was strongly associated with protein and fat consumption, whereas the *Prevotella* enterotype was associated with carbohydrate consumption (26). These associations are consistent with the differences between the typical U.S. diet (Western diet; high in animal protein and fats) and the typical Egyptian diet (Mediterranean diet; high in plant polysaccharides). To confirm our sequencing results, we utilized fluorescent *in situ* hybridization (FISH) to visualize *Bacteroides* and *Prevotella* cells in select egkHLT and uskHLT samples. As shown in Fig. 2C, a good concordance between sequencing and FISH results was observed, lending additional support to the idea of the significance of the observed *Bacteroides-Prevotella* reciprocity between the uskHLT and egkHLT groups.

Distal gut microbiotas separate subjects from Western and developing countries. Extending our observation of the similarities of the gut microbiota differences between U.S. and Egyptian children to the results of several previous studies, we sought

to carry out a cumulative ordination analysis of these data sets. Because it was shown previously that the choice of the interrogated 16S rRNA gene variable region has an impact on the estimates of microbial composition (49), we limited our comparison to studies that used the same V4 variable region of prokaryotic 16S rRNA gene to obtain a phylogenetic profile of gut microbiota. We combined our high-throughput amplicon sequencing data set (Illumina MiSeq platform) with a U.S.-versus-Malawi-versus-Venezuela subject comparison (Illumina HiSeq platform) (50), a Tanzania-versus-Italy subject comparison (Roche 454 FLX titanium platform) (30), and a U.S.-versus-Peru subject comparison (Illumina HiSeq platform) (33). UniFrac distance-based ordination PCoA of the combined data set revealed a statistically significant separation of fecal samples between industrialized countries (Italy and United States) and developing countries (Egypt, Malawi, Venezuela, Tanzania, and Peru) (Fig. 2D). Discriminatory random forest analysis and orthogonal projections to latent structures discriminant analysis (OPLS-DA) separated these two groups of samples well and designated *Prevotella* and *Bacteroides* the top separating genera (see Fig. S2; note that the random forest proximity matrix projection also separated our study samples from the others, which was likely due to technical variations among studies). The abundance gradients of these genera aligned along the primary axes of variability in the cumulative PCoA (Fig. 2E and F). Thus, we can speculate that there is a general dissimilarity of human distal gut microbiotas between industrialized populations and “developing” societies illustrated by the *Prevotella-Bacteroides* dichotomy.

Differences in distal gut metabolites reflect dietary preferences. To assess whether the differences in microbiota composition and consumed diets between Egyptian and U.S. teenagers can alter luminal environment, we employed proton nuclear magnetic resonance (NMR) to obtain metabolite profiles from all collected stool samples. ¹H NMR spectra were recorded for all water-soluble fecal extracts, and a dynamic binning algorithm was utilized to digitize spectral data (overlaid ¹H NMR spectra are shown in Fig. S3). Exploratory principal-component analysis (PCA), as well as discriminatory RF analysis and ordination OPLS-DA, separated the samples with statistical significance according to the group assignment (see Fig. S4). A spectral deconvolution algorithm was then utilized to robustly quantify the levels of 32 metabolites in all interrogated samples. Statistically significant differences between egkHLT and uskHLT groups were observed for the fractional abundances of many metabolites (Fig. 3A and Table S3). The three major short-chain fatty acids (SCFAs), acetate, butyrate, and propionate, were the most abundant metabolites in every fecal sample and overall showed significantly higher levels in the gut of Egyptian teenagers. Because SCFAs are the end products of the fermentation of complex polysaccharides, the latter finding is consistent with the higher fraction of dietary fiber in Mediterranean diet (51). On the other hand, levels of seven of nine measured amino acids were higher in U.S. children (see Fig. 3A), consistent with the higher protein consumption in subjects consuming a Western diet (4, 52). One of the largest differences was observed for lysine, likely because the levels of this indispensable amino acid are low in plant protein products (53). Interestingly, tryptophan and glycine were somewhat more abundant in Egyptian children. While levels of tryptophan are higher in seeds and nuts, which are consumed more frequently in Mediterranean countries (53), this amino acid is not easily accessible from many products such as cereals (54). Elevated levels of fecal glycine were previously noted after dietary supplementation of fructo-oligosaccharides (55), which are found at high levels in fruits. Similarly to the majority of amino acids, we found that the levels of metabolites related to lipid metabolism, including bile acids, taurine (which can be derived through deconjugation of primary bile acids), and choline, were all higher in the American volunteers. Because release of bile acids into the small intestine is increased in subjects on a high-fat diet (56), more bile acids likely reach the colon in subjects consuming fat-rich diets, which aligns with our observations. Levels of several central metabolism and fermentation intermediates (ethanol, lactate, malate, pyruvate, and succinate) were also higher in U.S. samples, possibly indicating incomplete

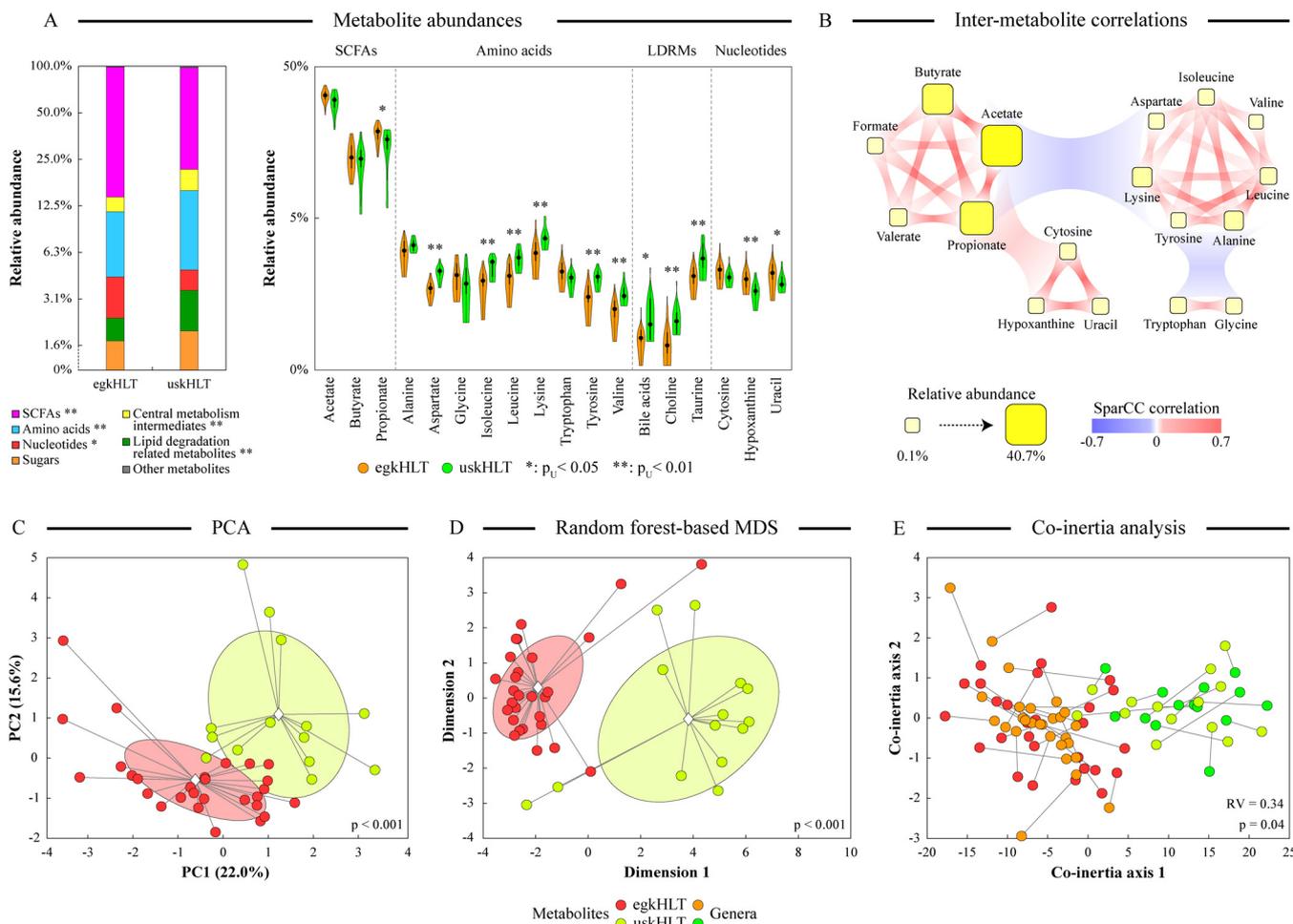


FIG 3 Comparison of distal gut metabolite profiles between Egyptian and U.S. groups. (A) Overall distribution of metabolite categories among egkHLT and uskHLT samples shown on a stacked-column graph. The relative abundances of individual measured metabolites are represented as violin plots. Where shown, single asterisks (*) and double asterisks (**) indicate statistical significance of taxon abundance differences between two groups ($p_u < 0.05$ and $p_u < 0.01$, respectively) based on the FDR-adjusted Mann-Whitney *U* test. LDRMs, lipid degradation-related metabolites. (B) Statistically significant SparCC-based correlations among individual metabolites and metabolite groups. The associations between metabolite categories represent the median of pairwise correlations among individual metabolites from different categories. (C and D) Sample distribution in ordination space based on the centered log-ratio transformed metabolite relative abundances from principal-component analysis (C) and random forest discriminant analysis (D). *P* values indicate statistical significance of separation of sample groups based on the Davies-Bouldin index. Group clouds represent three standard errors around the group centroids (diamond). MDS, multidimensional scaling. (E) Coinertia analysis showing congruency of sample dispersal in ordination space based on metabolite and genus abundance profiles. The distance between the positions of each sample on two ordination plots is indicated by a connecting line. Shorter lines represent similar sample positions in the plots. Statistical significance and the relative fit of the ordinations were assessed by *P* value and *RV* coefficient, respectively.

fermentation of complex polysaccharides in the guts of these teenagers (57). Intriguingly, the abundance of 1-methylhistamine, a fecal biomarker of allergic response (58), was also higher in uskHLT stools, consistent with the well-established higher prevalence of allergic diseases in industrialized countries (59).

These differences in measured metabolite levels between studied cohorts were sufficiently consistent within each group to distinguish the egkHLT and uskHLT samples. Exploratory PCA, as well as discriminant RF analysis, separated Egyptian and U.S. samples in the ordination space based on the relative abundances of 32 measured metabolites (Fig. 3C and D). Coinertia analysis (CIA) indicated that such sample separation was congruent with the dispersion of the same samples in the genus-based ordination analysis (Fig. 3E).

Utilizing the SparCC algorithm to account for the limitations of compositional data (28), we also generated networks of intermetabolite correlations shared among all samples (Fig. 3B). Not surprisingly, levels of SCFAs, nucleotide metabolites, and amino acids all correlated strongly within each group. An overall negative correlation was

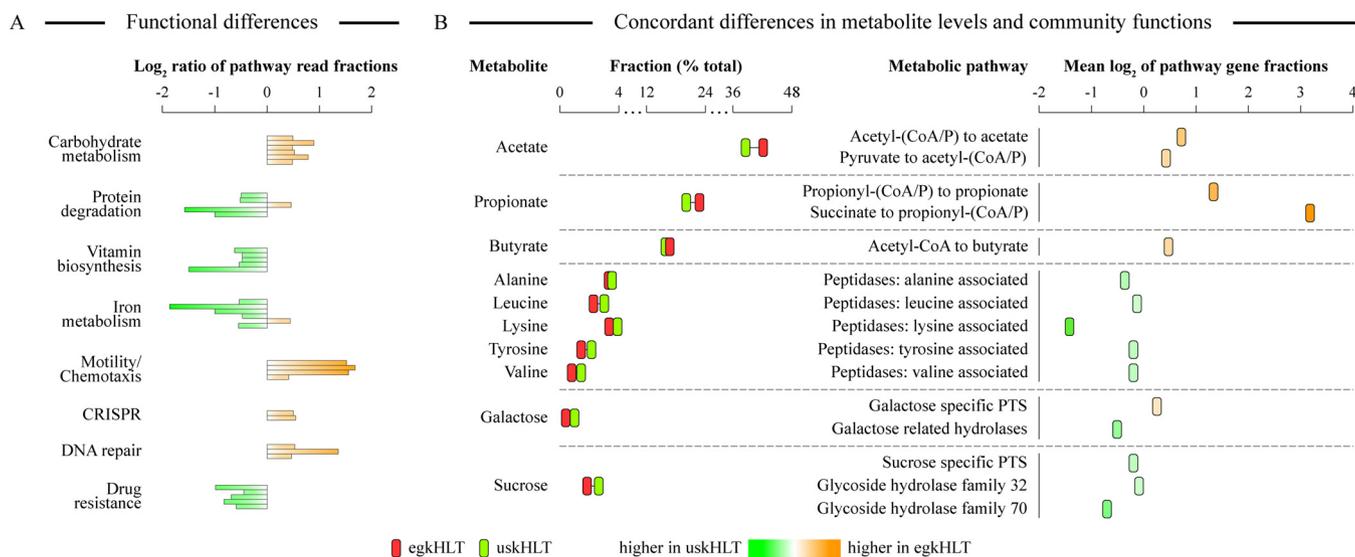


FIG 4 Functional analysis of the distal gut microbiota metagenomes. (A) Differences in specific functional categories between Egyptian and U.S. gut microbiota. Each category comprises several SEED level 3 entries, each represented as individual bars with results calculated as log₂-transformed ratios of reads between egkHLT and uskHLT metagenomes. All displayed entries are statistically significant, with FDR-corrected *P* of <0.01. (B) Relative abundances of specific metabolites and the mean log₂ ratios of reads of metabolic enzymes that produce and/or utilize the metabolites are compared. CoA/P, coenzyme A/phosphate.

observed between levels of SCFAs and amino acids (excluding tryptophan and glycine), consistent with nutritional differences between Mediterranean and Western diets.

Abundances of gut microbial functions are consistent with the prevalent Egyptian and U.S. diets. To establish a link between microbial composition in the gut and the levels of luminal metabolites, we carried out functional metagenomic profiling of both sets of stool samples. Statistically significant differences in the abundances of many functional genes were observed between the egkHLT and uskHLT cohorts (the full table of functional annotations is provided in Data Set S1 in the supplemental material). Specifically, concordant with the high consumption of cereals by Egyptians, many carbohydrate utilization pathways were more abundant in the egkHLT stools (see Fig. 4A). These pathways included cellulosome complex; catabolism of D-galactarate, D-glucarate, D-glycerate, and D-gluconate; mannitol and melibiose utilization; and enzymes of the pentose phosphate pathway. In contrast, many protein degradation pathways, including general protein degradation modules, general aminopeptidases, and degradation enzymes for specific amino acids such as lysine and histidine, were more abundant in the gut microbiota of U.S. children. Interestingly, and in concordance with our metabolite quantification, tryptophan catabolism-related genes were more prevalent in the Egyptian fecal samples. Biosynthetic pathways for several vitamins, including biotin, cobalamin, and vitamin K, were significantly more abundant in the guts of the U.S. children (average ratio, 1.68), likely because consumption of many refined products such as sugars, cereals, and vegetable oils, which are low in micro-nutrients and vitamins, leads to lower vitamin consumption (60). Several iron acquisition systems were also more prevalent in the guts of U.S. teenagers (see Fig. 4A and Data Set S1). It was shown previously that the intestinal barrier function in Egyptian children is occasionally compromised and thus that more iron can leak from mucosal tissues into intestinal lumen (61). At the same time, U.S. teenagers underconsume foods rich in this metal (60).

The overall abundance of motility and chemotaxis operons and quorum-sensing genes was higher in the Egyptian samples; this is consistent with the observed differential abundances of members of phylum *Proteobacteria*, which contains many known human gut pathogens (3.3% and 2.5% weighted mean relative abundances among egkHLT and uskHLT samples, respectively). Clustered regularly interspaced short palindromic repeat (CRISPR), restriction-modification, and DNA repair systems

were also more abundant in the bacterial genomes in egkHLT samples (Fig. 4A), indicating that the Egyptian gut microbiota is under heavier bacteriophage pressure (62, 63). In contrast, levels of genes in the category of “resistance to antibiotics and toxic compounds” (including resistance to vancomycin and β -lactam antibiotics) were significantly higher in U.S. samples, a finding which is likely explained by the widespread use of antibiotic treatments in the United States, resulting in selective advantages of microbial genome-carried antibiotic resistance genes.

To further validate the observed links between the fecal microbiota and metabolites, we also linked the differences in individual fatty acid, sugar, and amino acid levels between egkHLT and uskHLT cohorts to the levels of enzymes producing or utilizing these metabolites (see Fig. 4B). The abundances of all SCFA-producing fermentation pathway enzymes were higher in the Egyptian samples, in agreement with higher SCFA abundances in these samples. Interestingly, the abundances of sucrose- and galactose-specific glucoside hydrolase enzymes as well as of sucrose- and fructose-specific phosphotransferase systems (PTS) were higher in the uskHLT samples, matching the higher fecal sucrose and galactose metabolite levels. Peptidases specific to the release of alanine, leucine, lysine, tyrosine, and valine from proteins and peptides were also more prevalent in the U.S. gut microbiota, concordant with the higher amino acid levels in uskHLT stools (Fig. 4B).

DISCUSSION

The human physiological state is currently viewed as an interaction between a person’s genotype and the environment. This might explain in part why gene variants identified in many recent genome-wide studies can account for only a small proportion of the heritability of most complex diseases (64). Similarly, a recent alarming increase in the rates of obesity and excessive weight cannot be attributed to any genetic changes occurring on such short evolutionary time frame, and environmental factors such as diet, levels of physical activity, and compromised immune systems all contribute substantially to this obesity epidemic (65). The historic rate of dietary changes during the development of human species far exceeded that of possible genotype alterations (2). Thus, our genotype is presumably not well adapted to the current abundance of refined grains and sugars, dairy products, and animal fats in the diets of industrialized populations (1, 4, 66). These considerations are used to explain the observations that a Mediterranean diet, which is rich in plant products and low in animal fats, is associated with lower risk of cardiovascular diseases (11). The human gut microbiota serves as an important bridge connecting diet to human metabolism, since emerging evidence points to the metabolic mediation by the microbiota of both harmful and beneficial effects of dietary nutrients on human health (67, 68).

In this study, we explored the microbial and metabolic differences in the gut environment between two groups of adolescents—Egyptians consuming a Mediterranean diet and U.S. teenagers fed a typical Western diet. Integrative analysis of microbiota composition and functional capacity coupled with quantitative measurements of intestinal metabolites provided strong matching evidence of the differences between these populations. It appears that the gut microbiota in each population has adapted, at least in part, to the host’s prevalent diet. Thus, the Egyptian gut microbiota was enriched in polysaccharide-degrading members and genome-encoded enzymatic functions, whereas microbial communities in U.S. teenagers had higher counts of protein-degrading microbes and were enriched in protein and fat utilization pathways as well as in biosynthesis of the vitamins that are often found at low levels in Western diets. Microbiota adaptation to each gut environment was also evident from the overabundance of iron scavenging genes in the gut of U.S. children, consistent with the reports of insufficient iron consumption in subjects on a Western diet (60). Such differences in microbiota structure and function were reflected in the differences in the intestinal metabolites. While the gut environment of Egyptian teenagers was characterized by an abundance of short-chain fatty acids, intestines of U.S. children had increased amino acid content, higher levels of lipid metabolism-associated compounds, and elevated

concentrations of 1-methylhistamine, a biomarker for allergic reactions. SCFAs, especially butyrate, inhibit inflammation and protect against obesity (69, 70), whereas products of protein and lipid degradation are associated with a risk of developing atherosclerosis and colon cancer (67, 71). Thus, the observed differences in these metabolites between our cohorts are consistent with the epidemiological data showing higher rates of cardiovascular disease, metabolic syndrome, colorectal cancer, and autoimmune and allergic diseases in industrialized populations (72).

While it is tempting to assume that gut microbiota transformation of dietary nutrients plays a central role in the development of these diseases, other environmental, cultural, and genetic contributions, as well as study limitations, should also be taken into account. Factors such as health care, hygiene practices, cultural variation, and environmental exposures to toxins and pathogens are all likely to exert selection pressures on the gut microbiome and metabolome. For example, while the diet appears to be “healthier” in the Egyptian population, the rate of obesity is actually higher in Egypt among the members of the adult population, especially women, than in the United States. This seeming inconsistency can be potentially explained by several factors: the recent proliferation of “junk food” outlets in Egyptian cities (all our Egyptian teenagers were provided prepared meals and thus were not exposed to these sources of foods); increasingly sedentary lifestyles; and the lack of opportunities to play sports and to exercise (73). In this study, only male preadolescent and adolescent subjects were recruited; the gut environment in female teenagers was not profiled. Children in many developing countries, including Egypt, also often suffer from the environmental enteric dysfunction that can lead to nutrient malabsorption, altered immunity, and changes in the gut microbiota. In many cases, this enteric dysfunction might be caused by altered barrier function and reduced absorptive surface of the intestinal epithelium, and it is often associated with poorer hygiene and a higher prevalence of pathogens in the environment (74). Finally, acquisition of more-detailed dietary data in future studies should provide additional insights into associations of specific microbes and metabolites with particular types of foods. Nevertheless, because microbes can evolve and adapt to environmental changes much more rapidly than humans can, dietary modifications are likely to be among the most efficient and, at the same time, low-cost options for prevention and treatment of metabolic and immune diseases (75). Thus, modulation of human gut microbiota with prebiotic, probiotic, and synbiotic dietary supplementations, or through microbiota transplantation, can provide new approaches to control the diet-microbiota-human health interactions in the near future.

MATERIALS AND METHODS

Study subjects. Fresh fecal samples were collected in sterile containers from healthy preadolescent and adolescent male volunteers from Giza, Egypt (designated “egkHLT”; $n = 28$, average age = 13.9 ± 0.6 years; average body mass index [BMI] = 18.9 ± 2.5 kg/m²), and from Dayton, OH (designated “uskHLT”; $n = 14$, average age = 12.9 ± 2.8 years; average BMI = 21.2 ± 3.4 kg/m²). Both cohorts were living in the urban setting. Fresh fecal samples were homogenized immediately after collection and were frozen within 0 to 2 h after defecation as described previously (76, 77). The subject enrollment was limited to males to take advantage of the availability of a teenage male cohort in a welfare institution in Cairo, Egypt. Healthy volunteers did not have any gastrointestinal symptoms and had not consumed antibiotics or probiotics for at least 3 months prior to sample collection. For each volunteer, age and BMI values were collected and used in data interpretation.

Isolation of gDNA and high-throughput DNA sequencing. Total genomic DNA (gDNA) was isolated from 150 mg of fecal material using a ZR fecal DNA isolation kit (Zymo Research Corporation) according to manufacturer’s protocol. For the interrogation of microbial composition, the V4 variable region of the 16S ribosomal RNA gene was amplified using the universal primers 515F (5′-GTGCCAGCMGCCGCGGTAA) and 806R (5′-GGACTACHVGGGTWTCTAAT). The forward primers contained an 8-nucleotide barcode to permit sample pooling. PCR amplifications were performed in a 25- μ l volume with 25 ng of genomic DNA and 28 cycles of amplification. PCR products were cleaned and purified using calibrated AMPure XP beads (Beckman Coulter, Inc.). Amplicons were equimolarly pooled and processed using the Illumina TruSeq DNA library preparation protocol. Sequencing was performed on an Illumina MiSeq platform using the 2 \times 250-nucleotide-paired-end sequencing protocol following the manufacturer’s guidelines. An average of $72,039 \pm 34,259$ reads were obtained per sample. Paired-end sequence reads were joined together. Low-quality (average Q , <25) and short (<150-bp) reads were removed from the data set. High-quality reads were analyzed in QIIME using the default pipeline (78). Operational taxonomic units (OTUs) were defined by clustering at 97% sequence similarity.

Taxonomic annotation of OTUs was performed with the UCLUST algorithm (79) against curated Green-Genes database v13.8 (80). All OTU counts were adjusted to calculated taxon 16S rRNA gene copy numbers using the rrnDB library (81) in order to represent true relative abundances (76, 77). Finally, the adjusted reads from all samples were subsampled at the read level (rarefied) such that all samples were represented by the same number of counts. This final data set was used for all multivariate analyses.

To analyze community functional capacity, shotgun metagenomic sequencing was employed. Whole-community genomic DNAs from individual samples were equimolarly pooled within each group. Each pooled DNA sample was fragmented and processed using the Illumina TruSeq DNA library preparation protocol. Sequencing was performed on an Illumina MiSeq platform using the 2× 150-bp-paired-end sequencing protocol. Totals of 17,655,028 and 14,534,774 reads were obtained for egkHLT and uskHLT samples, respectively. All reads were uploaded into the MG-RAST analysis server (82). Paired reads were combined and subjected to quality filtering, and host sequences were depleted. The data set was then processed using the default MG-RAST analysis pipeline. The functional annotation was based on the SEED hierarchical system (83). STAMP statistical software was used for visualization and statistical hypothesis testing based on the two-sample Fisher exact test with *P* values adjusted for multiple-hypothesis testing using the Benjamini-Hochberg false-discovery-rate (FDR) algorithm (84). The following filters were used for the selection of level 3 entries from the SEED hierarchical annotation for in-depth analysis: (i) entries with greater than 500 reads; (ii) entries with a value for fold change between sample groups of at least 1.3; (iii) entries with consistent changes among functional genes within the same group.

In order to determine the abundances of genes encoding carbohydrate active enzymes, the protein sequences of all available glycoside hydrolases (GH), glycosyl transferases (GT), polysaccharide lyases (PL), and carbohydrate esterases (CE) were downloaded from the CAZy database (85). A BLASTP search was then used to match metagenomic sequence reads against this custom annotation database. Individual families clustered in a manner depending on the enzyme class substrate specificity and function. A similar approach was used to annotate gene reads encoding galactose- and-sucrose specific phosphotransferase systems (PTS).

Fluorescent *in situ* hybridization (FISH). FISH was carried out based on the methods of Zhu and Joerger (86). *Bacteroides* and *Prevotella* were visualized using newly designed fluorescein isothiocyanate (FITC)-labeled probes Bfra602 (5'-GAGCCGCAAACCTTCACAA) and Prev743 (5'-AATCCTGTCGATACCCG CA). The probes were designed to be specific to each genus with no cross-hybridizing to any other genera. The ability of each probe to detect members of the corresponding genus was checked via the probe match function of the Ribosomal Database Project's 16S rRNA gene database, and the correct hybridization was validated using pure cultures of *Bacteroides fragilis* and *Prevotella oralis*. To carry out FISH, bacterial cells were isolated from 100 mg of fecal material with phosphate-buffered saline (PBS)–0.1% SDS buffer and were fixed overnight with a 4% paraformaldehyde–PBS solution. Fixed cells were treated with lysozyme and then hybridized with the appropriate fluorescent probe at 46°C for 16 h using a probe-specific hybridization solution (86, 87). Fluorescent images of DAPI (4',6-diamidino-2-phenylindole)-stained and FITC-stained cells were captured through a 100× oil immersion objective using Image-Pro 6.2 software. Eight fields were imaged per hybridization. Total cell counts were obtained from images of DAPI-stained cells, and the ratio of probe-hybridized cells to total cells was determined using FITC images.

Preparation of fecal water extracts and proton nuclear magnetic resonance (NMR) analysis. A total of 250 mg of homogenized stool was used to prepare a metabolite water extract in phosphate buffer following our previously described protocol (57). A 550- μ l aliquot of the prepared fecal extract sample was transferred to a 5-mm-inner-diameter NMR tube together with 150 μ l of 9 mM trimethylsilylpropionic-2,2,3,3- d_4 acid (TSP) in D_2O . Proton (1H) NMR spectra were acquired at 25°C using a Varian Inova instrument operating at 600 MHz (14.1 Tesla) and a previously described procedure (57). TSP served as a chemical shift reference and quantification standard, and D_2O provided a field-frequency lock for NMR acquisition. Data were signal averaged over 400 transients using a 4.0-s acquisition time and an interpulse delay of 11.05 s. Spectral processing included removal of the residual water signal, chemical shift referencing, and sum normalization. For multivariate data analyses, spectra were binned to reduce the dimensionality and mitigate peak misalignment, and signal intensities were autoscaled (88). Quantification of specific metabolite resonances was accomplished using an interactive spectral deconvolution algorithm in MatLab as previously described (57). The deconvolution tool fits a defined spectral region using a combination of tunable baseline shapes (spline, v-shaped, linear, or constant) and a Gauss-Lorentz peak-fitting function. All metabolite peak intensities were corrected for equivalent numbers of protons and normalized to the TSP signal intensity. Peak metabolite assignments were either taken from our previous study (57) or confirmed in additional metabolite spike-in experiments. In the latter cases, each metabolite compound was added as a spike-in into a baseline fecal extract at a final concentration of 1.5 mM, the NMR spectrum was acquired, and the position of the metabolite-identifying peak was confirmed. There was substantial variability in the total sum of 32 metabolites among samples; the variability was not a function of sample group or sample water content. To increase the robustness of downstream analyses, all metabolite values were converted into fractions of the total.

Statistical data analyses. Statistical procedures were carried out in R, SPSS v19 (SPSS, Inc.), and MatLab (the MathWorks, Inc.). Weighted mean values around the median were calculated to obtain sample group averages as previously described (77). Weighted mean values reduce the effect of outliers on the mean estimate. The statistical significance of observed differences in the values of any quantitative variables between sample groups was assessed by the Mann-Whitney *U* test (reported as p_U values after FDR adjustment) (84). Beta diversity was measured with ecological Bray-Curtis and phylogenetic

UniFrac distances (25, 89). Multivariate ordination analyses were carried out to assess sample dispersal as a function of microbial and metabolite profiles. Prior to the analyses, genus and phylotype relative abundance data sets were subjected to chord transformation to account for many zero values, and the metabolite abundance data set was subjected to centered log-ratio transformation to correct for data compositionality (28). Principal-component analysis (PCA), principal-coordinate analysis (PCoA), UniFrac distance-based redundancy analysis (db-RDA), db-RDA-based variation partitioning, random forest (RF) analysis, and orthogonal projections to latent structures discriminant analysis (OPLS-DA) were run in MatLab and R. A Venn diagram of variation partitioning was constructed with eulerAPE (90). The statistical significance of group separation in PCA and PCoA was tested using the permutation of the Davies-Bouldin index measure (91). Performance of discrimination models was assessed based on the out-of-bag error rates. Identification of microbial enterotypes within the chord-transformed genus abundance data set was achieved by partitioning around the medoid (PAM) analysis (48) and between-cluster analysis (BCA) in R as previously described (46). The Calinski-Harabasz (CH) index was calculated for different numbers of clusters (between 2 and 20 clusters) to determine the number of clusters that provided the optimal sample distribution. The separation of samples into two clusters provided the best CH index value. This finding was further validated by calculating Silhouette scores in R.

To test the sample distribution congruency between the microbiota and metabolite data sets, transformed genus and metabolite relative abundance data sets were subjected to coinertia analysis (CIA) (28). CIA was performed in R, and permutation of the RV coefficient was used to test the significance of the congruency. To assess putative associations among the quantified metabolites in all samples, the metabolite fractional abundance data set was analyzed in SparCC (92). The statistical significance of observed correlations was calculated through comparison to null distributions generated by permutation and renormalization of data.

Data accessibility. Sequence data sets from 16S ribosomal RNA gene sequencing and metagenomic sequencing supporting the conclusions of this article are available in the Sequence Read Archive Repository (BioProject identifier [ID] [PRJNA314988](https://doi.org/10.1093/bioinformatics/btt111)) and the MG-RAST analysis server (accession IDs: for egkHLT, 4552772.3 [<http://metagenomics.anl.gov/linkin.cgi?metagenome=mgm4552772.3>] and 4552773.3 [<http://metagenomics.anl.gov/linkin.cgi?metagenome=mgm4552773.3>]; for uskHLT, 4552774.3 [<http://metagenomics.anl.gov/linkin.cgi?metagenome=mgm4552774.3>] and 4552775.3 [<http://metagenomics.anl.gov/linkin.cgi?metagenome=mgm4552775.3>]).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mSystems.00169-16>.

FIG S1, TIF file, 0.3 MB.

FIG S2, TIF file, 0.6 MB.

FIG S3, TIF file, 2.1 MB.

FIG S4, TIF file, 0.6 MB.

FIG S5, TIF file, 0.2 MB.

TABLE S1, XLSX file, 0.2 MB.

TABLE S2, XLSX file, 0.2 MB.

TABLE S3, XLSX file, 0.04 MB.

TABLE S4, XLSX file, 0.01 MB.

DATA SET S1, XLSX file, 0.6 MB.

ACKNOWLEDGMENTS

O.P. and L.H. designed the study; M.G., L.H., and O.P. collected samples from volunteers; V.S., M.G., J.M., A.G., and N.V.R. processed the samples and performed the experiments. O.P., N.V.R., L.H., and V.S. wrote the paper.

We are thankful to Andrew Neuforth for participation in NMR data processing and to Michael Leffak, Richard Agans, Denise Kramer, and Ghaith Fallata for helpful comments.

REFERENCES

- Cordain L, Miller JB, Eaton SB, Mann N, Holt SH, Speth JD. 2000. Plant-animal subsistence ratios and macronutrient energy estimations in worldwide hunter-gatherer diets. *Am J Clin Nutr* 71:682–692.
- Jew S, AbuMweis SS, Jones PJ. 2009. Evolution of the human diet: linking our ancestral diet to modern functional foods as a means of chronic disease prevention. *J Med Food* 12:925–934. <https://doi.org/10.1089/jmf.2008.0268>.
- Salamini F, Ozkan H, Brandolini A, Schäfer-Pregl R, Martin W. 2002. Genetics and geography of wild cereal domestication in the near east. *Nat Rev Genet* 3:429–441. <https://doi.org/10.1038/nrg817>.
- Cordain L, Eaton SB, Sebastian A, Mann N, Lindeberg S, Watkins BA, O'Keefe JH, Brand-Miller J. 2005. Origins and evolution of the Western diet: health implications for the 21st century. *Am J Clin Nutr* 81: 341–354.
- Reedy J, Krebs-Smith SM. 2010. Dietary sources of energy, solid fats, and added sugars among children and adolescents in the United States. *J Am Diet Assoc* 110:1477–1484. <https://doi.org/10.1016/j.jada.2010.07.010>.
- Steyn NP, Mann J, Bennett PH, Temple N, Zimmet P, Tuomilehto J, Lindström J, Louheranta A. 2004. Diet, nutrition and the prevention of

- type 2 diabetes. *Publ Health Nutr* 7:147–165. <https://doi.org/10.1079/PHN2003586>.
7. Grotto D, Zied E. 2010. The standard American diet and its relationship to the health status of Americans. *Nutr Clin Pract* 25:603–612. <https://doi.org/10.1177/0884533610386234>.
 8. Krebs-Smith SM, Cook A, Subar AF, Cleveland L, Friday J, Kahle LL. 1996. Fruit and vegetable intakes of children and adolescents in the United States. *Arch Pediatr Adolesc Med* 150:81–86. <https://doi.org/10.1001/archpedi.1996.02170260085014>.
 9. Van Horn LV. 2010. Report of the dietary guidelines advisory committee on the dietary guidelines for Americans. US Department of Agriculture, Washington, DC.
 10. Papandreou C, Tuomilehto H. 2014. Coronary heart disease mortality in relation to dietary, lifestyle and biochemical risk factors in the countries of the Seven Countries Study: a secondary dataset analysis. *J Hum Nutr Diet* 27:168–175. <https://doi.org/10.1111/jhn.12187>.
 11. de Lorgeril M, Salen P. 2001. Mediterranean type of diet for the prevention of coronary heart disease. A global perspective from the seven countries study to the most recent dietary trials. *Int J Vitam Nutr Res* 71:166–172. <https://doi.org/10.1024/0300-9831.71.3.166>.
 12. Karamanos B, Thanopoulou A, Angelico F, Assaad-Khalil S, Barbato A, Del Ben M, Dimitrijevic-Sreckovic V, Djordjevic P, Gallotti C, Katsilambros N, Migdalis I, Mrabet M, Petkova M, Roussi D, Tenconi MT. 2002. Nutritional habits in the Mediterranean Basin. The macronutrient composition of diet and its relation with the traditional Mediterranean diet. Multi-centre study of the Mediterranean Group for the Study of Diabetes (MGSD). *Eur J Clin Nutr* 56:983–991. <https://doi.org/10.1038/sj.ejcn.1601413>.
 13. Dernini S, Berry EM. 2015. Mediterranean diet: from a healthy diet to a sustainable dietary pattern. *Front Nutr* 2:15. <https://doi.org/10.3389/fnut.2015.00015>.
 14. 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>.
 15. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R, Gordon JI. 2008. Evolution of mammals and their gut microbes. *Science* 320:1647–1651. <https://doi.org/10.1126/science.1155725>.
 16. 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>.
 17. Serino M, Luche E, Gres S, Baylac A, Bergé M, Cenac C, Waget A, Klopp P, Iacovoni J, Klopp C, Mariette J, Bouche O, Lluch J, Ouarne F, Monsan P, Valet P, Roques C, Amar J, Bouloumié A, Théodorou V, Burcelin R. 2012. Metabolic adaptation to a high-fat diet is associated with a change in the gut microbiota. *Gut* 61:543–553. <https://doi.org/10.1136/gutjnl-2011-301012>.
 18. Cummings JH, Englyst HN. 1987. Fermentation in the human large intestine and the available substrates. *Am J Clin Nutr* 45:1243–1255.
 19. Flint HJ, Scott KP, Louis P, Duncan SH. 2012. The role of the gut microbiota in nutrition and health. *Nat Rev Gastroenterol Hepatol* 9:577–589. <https://doi.org/10.1038/nrgastro.2012.156>.
 20. Davis CD, Milner JA. 2009. Gastrointestinal microflora, food components and colon cancer prevention. *J Nutr Biochem* 20:743–752. <https://doi.org/10.1016/j.jnutbio.2009.06.001>.
 21. Larsen N, Vogensen FK, van den Berg FW, Nielsen DS, Andreasen AS, Pedersen BK, Al-Soud WA, Sørensen SJ, Hansen LH, Jakobsen M. 2010. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS One* 5:e9085. <https://doi.org/10.1371/journal.pone.0009085>.
 22. Karlsson F, Tremaroli V, Nielsen J, Bäckhed F. 2013. Assessing the human gut microbiota in metabolic diseases. *Diabetes* 62:3341–3349. <https://doi.org/10.2337/db13-0844>.
 23. Paliy O, Piyathilake CJ, Kozyrskyj A, Celep G, Marotta F, Rastmanesh R. 2014. Excess body weight during pregnancy and offspring obesity: potential mechanisms. *Nutrition* 30:245–251. <https://doi.org/10.1016/j.nut.2013.05.011>.
 24. Trasande L, Blustein J, Liu M, Corwin E, Cox LM, Blaser MJ. 2013. Infant antibiotic exposures and early-life body mass. *Int J Obes (Lond)* 37: 16–23. <https://doi.org/10.1038/ijo.2012.132>.
 25. 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>.
 26. 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, Baldasano 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>.
 27. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. 2009. Bacterial community variation in human body habitats across space and time. *Science* 326:1694–1697. <https://doi.org/10.1126/science.1177486>.
 28. Paliy O, Shankar V. 2016. Application of multivariate statistical techniques in microbial ecology. *Mol Ecol* 25:1032–1057. <https://doi.org/10.1111/mec.13536>.
 29. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, Collini S, Pieraccini G, Lionetti P. 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A* 107:14691–14696. <https://doi.org/10.1073/pnas.1005963107>.
 30. Schnorr SL, Candela M, Rampelli S, Centanni M, Consolandi C, Basaglia G, Turroni S, Biagi E, Peano C, Severgnini M, Fiori J, Gotti R, De Bellis G, Luiselli D, Brigidi P, Mabulla A, Marlowe F, Henry AG, Crittenden AN. 2014. Gut microbiome of the Hadza hunter-gatherers. *Nat Commun* 5:3654. <https://doi.org/10.1038/ncomms4654>.
 31. Lin A, Bik EM, Costello EK, Dethlefsen L, Haque R, Relman DA, Singh U. 2013. Distinct distal gut microbiome diversity and composition in healthy children from Bangladesh and the United States. *PLoS One* 8:e53838. <https://doi.org/10.1371/journal.pone.0053838>.
 32. Nakayama J, Watanabe K, Jiang J, Matsuda K, Chao SH, Haryono P, La-Ongkham O, Sarwoko MA, Sujaya IN, Zhao L, Chen KT, Chen YP, Chiu HH, Hidaka T, Huang NX, Kiyohara C, Kurakawa T, Sakamoto N, Sonomoto K, Tashiro K, Tsuji H, Chen MJ, Leelavatcharamas V, Liao CC, Nitisinprasert S, Rahayu ES, Ren FZ, Tsai YC, Lee YK. 2015. Diversity in gut bacterial community of school-age children in Asia. *Sci Rep* 5:8397. <https://doi.org/10.1038/srep08397>.
 33. Obregon-Tito AJ, Tito RY, Metcalf J, Sankaranarayanan K, Clemente JC, Ursell LK, Zech Xu Z, Van Treuren W, Knight R, Gaffney PM, Spicer P, Lawson P, Marin-Reyes L, Trujillo-Villarroel O, Foster M, Gujja-Poma E, Troncoso-Corzo L, Warinner C, Ozga AT, Lewis CM. 2015. Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat Commun* 6:6505. <https://doi.org/10.1038/ncomms7505>.
 34. Elhenawy W, Debely MO, Feldman MF. 2014. Preferential packing of acidic glycosidases and proteases into Bacteroides outer membrane vesicles. *mBio* 5:e00909-14. <https://doi.org/10.1128/mBio.00909-14>.
 35. Gibson SA, Macfarlane GT. 1988. Studies on the proteolytic activity of Bacteroides fragilis. *J Gen Microbiol* 134:19–27. <https://doi.org/10.1099/00221287-134-1-19>.
 36. Dodd D, Mackie RI, Cann IK. 2011. Xylan degradation, a metabolic property shared by rumen and human colonic Bacteroidetes. *Mol Microbiol* 79:292–304. <https://doi.org/10.1111/j.1365-2958.2010.07473.x>.
 37. Shetty SA, Marathe NP, Lanjekar V, Ranade D, Shouche YS. 2013. Comparative genome analysis of Megasphaera sp. reveals niche specialization and its potential role in the human gut. *PLoS One* 8:e79353. <https://doi.org/10.1371/journal.pone.0079353>.
 38. Ordiz MI, May TD, Mihindukulasuriya K, Martin J, Crowley J, Tarr PI, Ryan K, Mortimer E, Gopalsamy G, Maleta K, Mitreva M, Young G, Manary MJ. 2015. The effect of dietary resistant starch type 2 on the microbiota and markers of gut inflammation in rural Malawi children. *Microbiome* 3:37. <https://doi.org/10.1186/s40168-015-0102-9>.
 39. Kageyama A, Benno Y. 2000. Catenibacterium mitsuokai gen. nov., sp. nov., a Gram-positive anaerobic bacterium isolated from human faeces. *Int J Syst Evol Microbiol* 50:1595–1599. <https://doi.org/10.1099/00207713-50-4-1595>.
 40. Sun Y, Zhou L, Fang L, Su Y, Zhu W. 2015. Responses in colonic microbial community and gene expression of pigs to a long-term high resistant starch diet. *Front Microbiol* 6:877. <https://doi.org/10.3389/fmicb.2015.00877>.
 41. Lingström P, van Houte J, Kashket S. 2000. Food starches and dental caries. *Crit Rev Oral Biol Med* 11:366–380. <https://doi.org/10.1177/10454411000110030601>.
 42. Devkota S, Wang Y, Musch MW, Leone V, Fehlner-Peach H, Nadimpalli A, Antonopoulos DA, Jabri B, Chang EB. 2012. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in IL10^{-/-} mice. *Nature* 487:104–108. <https://doi.org/10.1038/nature11225>.
 43. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermúdez-Humarán LG, Gratadoux JJ, Blugeon S, Bridonneau C, Furet JP, Corthier G, Grangette C, Vasquez N, Pochart P, Trugnan G, Thomas G, Blottière HM, Doré J,

- Marteau P, Seksik P, Langella P. 2008. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* 105: 16731–16736. <https://doi.org/10.1073/pnas.0804812105>.
44. Everard A, Belzer C, Geurts L, Ouwerkerk JP, Druart C, Bindels LB, Guoit Y, Derrien M, Muccioli GG, Delzenne NM, de Vos WM, Cani PD. 2013. Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. *Proc Natl Acad Sci U S A* 110:9066–9071. <https://doi.org/10.1073/pnas.1219451110>.
 45. Cooper GS, Bynum ML, Somers EC. 2009. Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. *J Autoimmun* 33:197–207. <https://doi.org/10.1016/j.jaut.2009.09.008>.
 46. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandez GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Doré J; MetaHIT Consortium, Antolin M, Artiguenave F, Blottiere HM, Almeida M, Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariac G, Dervyn R, et al. 2011. Enterotypes of the human gut microbiome. *Nature* 473:174–180. <https://doi.org/10.1038/nature09944>.
 47. Roager HM, Licht TR, Poulsen SK, Larsen M, Bahl MI. 2014. Microbial enterotypes, inferred by the Prevotella-to-Bacteroides ratio, remained stable during a 6-month randomized controlled diet intervention with the new Nordic diet. *Appl Environ Microbiol* 80:1142–1149. <https://doi.org/10.1128/AEM.03549-13>.
 48. Kaufman L, Rousseeuw P. 1990. Partitioning around medoids (program pam), finding groups in data: an introduction to cluster analysis. Wiley & Sons, New York.
 49. Guo F, Ju F, Cai L, Zhang T. 2013. Taxonomic precision of different hypervariable regions of 16S rRNA gene and annotation methods for functional bacterial groups in biological wastewater treatment. *PLoS One* 8:e76185. <https://doi.org/10.1371/journal.pone.0076185>.
 50. Yatsunenkov T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JL. 2012. Human gut microbiome viewed across age and geography. *Nature* 486:222–227. <https://doi.org/10.1038/nature11053>.
 51. Kushi LH, Lenart EB, Willett WC. 1995. Health implications of Mediterranean diets in light of contemporary knowledge. 1. Plant foods and dairy products. *Am J Clin Nutr* 61:1407S–1415S.
 52. Dai ZL, Wu G, Zhu WY. 2011. Amino acid metabolism in intestinal bacteria: links between gut ecology and host health. *Front Biosci* 16: 1768–1786. <https://doi.org/10.2741/3820>.
 53. Young VR, Pellett PL. 1994. Plant proteins in relation to human protein and amino acid nutrition. *Am J Clin Nutr* 59:1203S–1212S.
 54. Charalampopoulos D, Wang R, Pandiella SS, Webb C. 2002. Application of cereals and cereal components in functional foods: a review. *Int J Food Microbiol* 79:131–141. [https://doi.org/10.1016/S0168-1605\(02\)00187-3](https://doi.org/10.1016/S0168-1605(02)00187-3).
 55. Martin FP, Sprenger N, Yap IK, Wang Y, Bibiloni R, Rochat F, Rezzi S, Cherbut C, Kochhar S, Lindon JC, Holmes E, Nicholson JK. 2009. Panorganismal gut microbiome-host metabolic crosstalk. *J Proteome Res* 8:2090–2105. <https://doi.org/10.1021/pr801068x>.
 56. Reddy BS. 1981. Diet and excretion of bile acids. *Cancer Res* 41: 3766–3768.
 57. Shankar V, Homer D, Rigsbee L, Khamis HJ, Michail S, Raymer M, Reo NV, Paliy O. 2015. The networks of human gut microbe-metabolite associations are different between health and irritable bowel syndrome. *ISME J* 9:1899–1903. <https://doi.org/10.1038/ismej.2014.258>.
 58. Bischoff SC, Grabowsky J, Manns MP. 1997. Quantification of inflammatory mediators in stool samples of patients with inflammatory bowel disorders and controls. *Dig Dis Sci* 42:394–403. <https://doi.org/10.1023/A:1018886423475>.
 59. Okada H, Kuhn C, Feillet H, Bach JF. 2010. The “hygiene hypothesis” for autoimmune and allergic diseases: an update. *Clin Exp Immunol* 160: 1–9. <https://doi.org/10.1111/j.1365-2249.2010.04139.x>.
 60. U.S. Departments of Agriculture and Health and Human Services. 2015. Scientific report of the 2015 Dietary Guidelines Advisory Committee. <https://health.gov/dietaryguidelines/2015-scientific-report/pdfs/scientific-report-of-the-2015-dietary-guidelines-advisory-committee.pdf>.
 61. Mohammad M, Hussein L, Yamamah G, Rawi S. 2007. The impact of probiotic and/or honey supplements on gut permeability among Egyptian children. *J Nutr Environ Med* 16:10–15. <https://doi.org/10.1080/13590840601016387>.
 62. Szczepankowska A. 2012. Role of CRISPR/cas system in the development of bacteriophage resistance. *Adv Virus Res* 82:289–338. <https://doi.org/10.1016/B978-0-12-394621-8.00011-X>.
 63. Rohwer F, Azam F. 2000. Detection of DNA damage in prokaryotes by terminal deoxyribonucleotide transferase-mediated dUTP nick end labeling. *Appl Environ Microbiol* 66:1001–1006. <https://doi.org/10.1128/AEM.66.3.1001-1006.2000>.
 64. Thomas D. 2010. Gene-environment-wide association studies: emerging approaches. *Nat Rev Genet* 11:259–272. <https://doi.org/10.1038/nrg2764>.
 65. Sellayah D, Cagampang FR, Cox RD. 2014. On the evolutionary origins of obesity: a new hypothesis. *Endocrinology* 155:1573–1588. <https://doi.org/10.1210/en.2013-2103>.
 66. Zucoloto FS. 2011. Evolution of the human feeding behavior. *Psychol Neurosci* 4:131–141. <https://doi.org/10.3922/j.psns.2011.1.015>.
 67. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, Feldstein AE, Britt EB, Fu X, Chung YM, Wu Y, Schauer P, Smith JD, Allayee H, Tang WH, DiDonato JA, Lusis AJ, Hazen SL. 2011. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472:57–63. <https://doi.org/10.1038/nature09922>.
 68. Moco S, Martin FP, Rezzi S. 2012. Metabolomics view on gut microbiome modulation by polyphenol-rich foods. *J Proteome Res* 11:4781–4790. <https://doi.org/10.1021/pr300581r>.
 69. Guilloteau P, Martin L, Eeckhaut V, Ducatelle R, Zabielski R, Van Immerseel F. 2010. From the gut to the peripheral tissues: the multiple effects of butyrate. *Nutr Res Rev* 23:366–384. <https://doi.org/10.1017/S0954422410000247>.
 70. Zimmerman MA, Singh N, Martin PM, Thangaraju M, Ganapathy V, Waller JL, Shi H, Robertson KD, Munn DH, Liu K. 2012. Butyrate suppresses colonic inflammation through HDAC1-dependent Fas upregulation and Fas-mediated apoptosis of T cells. *Am J Physiol Gastrointest Liver Physiol* 302:G1405–G1415. <https://doi.org/10.1152/ajpgi.00543.2011>.
 71. O’Keefe SJ, Ou J, Aufreiter S, O’Connor D, Sharma S, Sepulveda J, Fukuwatari T, Shibata K, Mawhinney T. 2009. Products of the colonic microbiota mediate the effects of diet on colon cancer risk. *J Nutr* 139:2044–2048. <https://doi.org/10.3945/jn.109.104380>.
 72. World Health Organization. 2011. Global status report on noncommunicable diseases, p 1–162. World Health Organization, Geneva, Switzerland.
 73. Galal OM. 2002. The nutrition transition in Egypt: obesity, undernutrition and the food consumption context. *Publ Health Nutr* 5:141–148. <https://doi.org/10.1079/PHN2001286>.
 74. Crane RJ, Jones KD, Berkley JA. 2015. Environmental enteric dysfunction: an overview. *Food Nutr Bull* 36:576–587. <https://doi.org/10.1177/156482651503615113>.
 75. Bifulco M. 2015. Mediterranean diet: the missing link between gut microbiota and inflammatory diseases. *Eur J Clin Nutr* 69:1078. <https://doi.org/10.1038/ejcn.2015.81>.
 76. Rigsbee L, Agans R, Foy BD, Paliy O. 2011. Optimizing the analysis of human intestinal microbiota with phylogenetic microarray. *FEMS Microbiol Ecol* 75:332–342. <https://doi.org/10.1111/j.1574-6941.2010.01009.x>.
 77. Agans R, Rigsbee L, Kenche H, Michail S, Khamis HJ, Paliy O. 2011. Distal gut microbiota of adolescent children is different from that of adults. *FEMS Microbiol Ecol* 77:404–412. <https://doi.org/10.1111/j.1574-6941.2011.01120.x>.
 78. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JL, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenkov T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336. <https://doi.org/10.1038/nmeth.f303>.
 79. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
 80. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72:5069–5072. <https://doi.org/10.1128/AEM.03006-05>.
 81. Stoddard SF, Smith BJ, Hein R, Roller BR, Schmidt TM. 2015. rrnDB:

- improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res* 43:D593–D598. <https://doi.org/10.1093/nar/gku1201>.
82. Glass EM, Wilkening J, Wilke A, Antonopoulos D, Meyer F. 2010. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb Protoc* 2010:pdb.prot5368. <https://doi.org/10.1101/pdb.prot5368>.
83. Mitra S, Rupek P, Richter DC, Urich T, Gilbert JA, Meyer F, Wilke A, Huson DH. 2011. Functional analysis of metagenomes and metatranscriptomes using SEED and KEGG. *BMC Bioinformatics* 12(Suppl 1):S21. <https://doi.org/10.1186/1471-2105-12-S1-S21>.
84. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. 2014. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30:3123–3124. <https://doi.org/10.1093/bioinformatics/btu494>.
85. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42:D490–D495. <https://doi.org/10.1093/nar/gkt1178>.
86. Zhu XY, Joerger RD. 2003. Composition of microbiota in content and mucus from caecae of broiler chickens as measured by fluorescent in situ hybridization with group-specific, 16S rRNA-targeted oligonucleotide probes. *Poult Sci* 82:1242–1249. <https://doi.org/10.1093/ps/82.8.1242>.
87. Manz W, Amann R, Ludwig W, Vancanneyt M, Schleifer KH. 1996. Application of a suite of 16S rRNA-specific oligonucleotide probes designed to investigate bacteria of the phylum Cytophaga-Flavobacter-Bacteroides in the natural environment. *Microbiology* 142:1097–1106. <https://doi.org/10.1099/13500872-142-5-1097>.
88. Anderson PE, Mahle DA, Doom TE, Reo NV, DelRaso NJ, Raymer ML. 2011. Dynamic adaptive binning: an improved quantification technique for NMR spectroscopic data. *Metabolomics* 7:179–190. <https://doi.org/10.1007/s11306-010-0242-7>.
89. Bray JR, Curtis JT. 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr* 27:325–349. <https://doi.org/10.2307/1942268>.
90. Micallef L, Rodgers P. 2014. eulerAPE: drawing area-proportional 3-Venn diagrams using ellipses. *PLoS One* 9:e101717. <https://doi.org/10.1371/journal.pone.0101717>.
91. Davies DL, Bouldin DW. 1979. A cluster separation measure. *IEEE Transactions Pattern Anal Mach Intell* 1:224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>.
92. Friedman J, Alm EJ. 2012. Inferring correlation networks from genomic survey data. *PLoS Comput Biol* 8:e1002687. <https://doi.org/10.1371/journal.pcbi.1002687>.