Sciencexpress Report

Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes

Gary D. Wu,[1]* Jun Chen,[2,3] Christian Hoffmann,[4,5] Kyle Bittinger,[4] Ying-Yu Chen,[1] Sue A. Keilbaugh,[1] Meenakshi Bewtra,[1,2] Dan Knights,[6] William A. Walters,[7] Rob Knight,[8,9] Rohini Sinha,[4] Erin Gilroy,[2] Kernika Gupta,[10] Robert Baldassano,[10] Lisa Nessel,[2] Hongzhe Li,[2,3] Frederic D. Bushman,[4]* James D. Lewis[1,2,3]*

[1]Division of Gastroenterology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [2]Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [3]Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [4]Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [5]Instituto de Ciências Biológicas, Universidade Federal de Goiás Goiania, GO, 74001-970, Brazil. [6]Department of Computer Science, University of Colorado, Boulder, CO 80309, USA. [7]Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309, USA. [8]Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA. [9]Howard Hughes Medical Institute, University of Colorado, Boulder, CO 80309, USA. [10]Division of Gastroenterology, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.

*To whom correspondence should be addressed. E-mail: gdwu@mail.med.upenn.edu (G.D.W.); lewisjd@mail.med.upenn.edu (J.D.L.); bushman@mail.med.upenn.edu (F.D.B.)

Diet strongly affects human health, partly by modulating gut microbiome composition. Here we used diet inventories and 16S rDNA sequencing to characterize fecal samples from 98 individuals. Fecal communities clustered into enterotypes distinguished primarily by levels of *Bacteroides* and *Prevotella*. Enterotypes were strongly associated with long-term diets, particularly protein and animal fat (*Bacteroides*) versus carbohydrates (*Prevotella*). A controlled-feeding study of 10 subjects showed that microbiome composition changed detectably within 24 hours of initiating a high fat/low fiber or low fat/high fiber diet, but that enterotype identity remained stable during the 10 day study. Thus, alternative enterotype states are associated with long-term diet.

We coexist with our gut microbiota as mutualists, but this relationship sometimes becomes pathological, as in obesity, diabetes, atherosclerosis, and inflammatory bowel diseases (*1*, *2*). Factors including age, genetics, and diet may influence microbiome composition (*3*). Of these, diet is easiest to modify and presents the simplest route for therapeutic intervention. Recently, an analysis of gut microbial communities proposed three predominant variants, or "enterotypes", dominated by *Bacteroides*, *Prevotella*, and *Ruminococcus* respectively (*4*). The basis for enterotype clustering is unknown but appears independent of nationality, sex, age, or body mass index (BMI).

Here we investigated the association of dietary and environmental variables with the gut microbiota. First, in a cross-sectional analysis of 98 healthy volunteers (abbreviated "COMBO" for "cross-sectional study of diet and stool microbiome composition), we collected diet information using two questionnaires, which queried recent diet ("Recall"), and habitual long-term diet (food frequency questionnaire, or "FFQ"). Second, ten individuals were sequestered in a hospital environment in a controlled-feeding study (abbreviated "CAFE" for "controlled feeding experiment") to compare high-fat/low-fiber and low-fat/high-fiber diets. Stool samples were collected (*5*), and DNA samples analyzed by 454/Roche pyrosequencing (*6*) of 16S rDNA gene segments and, for selected samples, shotgun metagenomics (*7*). In CAFE, rectal biopsy samples were also collected and analyzed on days 1 and 10.

For COMBO, we calculated pairwise UniFrac distances (*8*) among the microbial communities using 16S rDNA sequences. We assessed both relative abundance data (weighted analysis) and presence/absence information (unweighted analysis). Specific nutrients associated with variation in the gut microbiome for the 98 subjects were extracted, along with demographic factors (table S1). For each nutrient, we performed PERMANOVA (*9*) to test for nutrient microbiome association, from which we identified 72 and 97 microbiome-associated nutrients in Recall and FFQ, respectively, at a false discovery rate (FDR) of 25% (the relatively high value was used so as not to miss possible effects of diet on low abundance bacteria). Both weighted and unweighted UniFrac identified similar nutrients, though the discrimination was sharper with unweighted, indicating that change in community membership rather than community composition was the main factor.

For each of these nutrients, we identified the associated bacterial genera using Spearman correlations. We considered only the 78 taxa that had abundance ≥0.2% in at least one sample and appeared in more than 10% of the samples. Figure 1 shows a heatmap summarizing Spearman

Downloaded from www.sciencemag.org on September 1, 2011

correlations between nutrients from the FFQ and bacterial taxa. For a given taxon, individual nutrients account for 3-20% of the between-subject variation in abundance.

Nutrients of the same food groups from Recall and FFQ tended to cluster together (fig. S1A). The nutrients from fat and plant products/fiber showed inverse associations with microbial taxa (Spearman rho = –0.68, p < 0.0001). Inverse associations were also seen with amino acids/protein and carbohydrates (Spearman rho = –0.73, p < 0.0001) and with fat and carbohydrates (Spearman rho = –0.61, p = 0.0001). Phyla positively associated with fat but negatively associated with fiber were predominately *Bacteroidetes* and *Actinobacteria*, while *Firmicutes* and *Proteobacteria* showed the opposite association. However, within each Phylum, not all lower level taxa demonstrated similar correlations with dietary components (fig. S1B). Taxa correlated with body mass index (BMI) also correlated with fat and percent calories from saturated fatty acids (fig. S1B and table S1).

Following the suggestion by Arumugam *et al*. (*4*) that the human gut microbiome can be partitioned into enterotypes, we investigated whether the 98 COMBO samples partitioned into clusters that were detectably associated with dietary or demographic data (Fig. 2). Several methods for data processing and clustering were compared (fig. S2). In one analytical approach (Weighted UniFrac, no Lane masking; fig. S2), PAM analysis favored partitioning into three clusters, though with quite low support (silhouette score 0.2) suggesting that clustering could be due to chance. Comparison to the three genera specified by Arumugam *et al*. (*4*) showed that relatively high levels of the genera *Bacteroides* and *Prevotella* distinguished two of the clusters, while the third showed slightly higher levels of *Ruminococcus*. However, most methods showed two clusters, with stronger support (Fig. 2; Jensen-Shannon Distance, silhouette score 0.66), in which the *Bacteroides* enterotype was fused with the less well distinguished *Ruminococcus* enterotype. As described below, dietary effects primarily discriminate the *Prevotella* enterotype from the *Bacteroides* enterotype.

At a false discovery rate of 5%, 6 genera differed between the *Prevotella* and *Bacteroides* enterotypes (fig. S3). The *Bacteroides* enterotype was distinguished by the additional presence of *Alistipes,* and *Parabacteroides* (Phylum Bacteroidetes). The *Prevotella* enterotype was distinguished by the additional presence of *Paraprevotella* (Phylum Bacteriodetes) and *Catenibacterium* (Phylum Firmicutes) (fig. S3). The enterotype clustering is driven primarily by the ratio of the two dominant genera, *Prevotella* to *Bacteroides*, which defines a gradient across the two enterotypes (fig. S5).

At an FDR of 25%, nutrients from the long-term FFQ but not the short-term Recall questionnaire were associated with enterotype composition, indicating that long-term diet strongly correlates with enterotype (the relatively high FDR was used to avoid excessively strict filtering and to visualize the full pattern). The *Bacteroides* enterotype was highly associated with animal protein, a variety of amino acids, and saturated fats (Fig. 2C), suggesting that meat consumption as in a Western diet characterized this enterotype. The *Prevotella* enterotype, in contrast, was associated with low values for these groups but high values for carbohydrates and simple sugars, indicating association with a carbohydrate-based diet more typical of agrarian societies. Self-reported vegetarians (n = 11) showed enrichment in the *Prevotella* enterotype (27% *Prevotella* enterotype vs. 10% *Bacteriodes* enterotype; p = 0.13). The one self-reported vegan was in the *Prevotella* enterotype. No significant associations were seen with demographic data at this FDR.

A short-term controlled-feeding experiment (CAFE) was carried out to test the stability of the gut microbiome and the observed nutrient-microbiome associations. Ten subjects were sequestered and randomized to diets containing high fat/low fiber or low fat/high fiber, and sampled over 10 days (Fig. 3). Analysis of 16S tag data from stool samples showed that inter-subject variation was by far the predominant source of variance in the data (*10*). Figure 3A shows sharp clustering of the microbiome sequence data by individual in unweighted UniFrac, emphasizing that distinctive lineages are present in each subject. Over 10 days of controlled feeding, there was no reduction in UniFrac distances for stool or biopsy samples between individuals fed the same diet, demonstrating that a short-term identical diet does not overcome intersubject variation.

Remarkably, changes in microbiome composition were detectable within 24 hours of initiating controlled feeding. For each individual sampled, the first sampling day represented an outlier (Fig. 3B; p = 0.0003, 10000 permutations), indicating rapid change. Similar results were seen in the unweighted analysis (p = 0.0002). The taxa affected differed among individuals.

The relationship of changes in microbiome composition to the transit time of material through the gut was also investigated. Subjects swallowed X-ray-opaque markers at the start of the study, allowing quantification of transit time by abdominal X-ray. Transit time was faster with the high fiber diet (2-4 days) compared with the high fat diet (2-7 days; p = 0.02; two-sided Wilcoxon rank sum test), as expected. All patients retained at least one of the 24 markers 48 hours after the start of the experimental diet. Thus, the changes in microbiome composition, which occurred within 24 hours, were faster than clearance of residual material from the gut.

To probe metabolic functionality during the CAFE study, we also analyzed changes in total gene content using shotgun metagenomics. We compared stool samples from day 1 and

Downloaded from www.sciencemag.org on September 1, 2011

day 10 (1.05×10^6 sequence reads total). Sequence reads were annotated for function using the KEGG database (*11*), then interrogated to assess the taxa and classes of genes present. No significant changes in proportions among Archaea, Bacteria, and Eukarya were detected, and bacterial taxa inferred from shotgun metagenomic data paralleled the 16S rDNA data (fig. S4). We investigated gene groups that changed significantly between day 1 and 10 and differed between the high-fat and high-fiber groups. To control for between-subject variability, we used the day 1 samples as within-subject controls, and subtracted each subject's day 1 functional category counts from day 10 samples from that same subject. Functional categories that differentiated diets include bacterial secretion system ($p = 0.01$, *t* test), protein export ($p = 0.022$), and lipoic acid metabolism ($p = 0.045$), thus indicating bacterial functions potentially involved in responding to these dietary changes.

We next assessed the response of enterotypes to the controlled feeding regimen. Each of the samples from the 10 subjects was assigned to an enterotype category based on their microbiome distances to the medoids (*12*) of the enterotype clusters as defined in the COMBO data. All subjects started in the *Bacteroides* enterotype (high protein and fat). None switched stably to the *Prevotella* (carbohydrate) enterotype over the duration of the study. A single specimen scored in the *Prevotella* (carbohydate) enterotype, but reverted by the time of the next sample. Thus, over the 10 days of the dietary intervention, we did not see stable switching between the two enterotype groups characterized by the dietary extremes, despite feeding of a low fat/high fiber diet to half the subjects.

Finally, several factors were significantly correlated with microbiome composition but not with enterotype partitioning. Examples include BMI, red wine, and aspartame consumption (*7*). Thus, not all associations between host and microbiota are captured in the enterotype distinctions.

In summary, comparison of long-term and short-term dietary data showed that only the long-term diet was correlated with enterotype clustering in the cross-sectional study. In the interventional study, changes were significant and rapid, but the magnitude of the changes was modest, and not sufficient to switch individuals between the enterotype clusters associated with protein/fat and carbohydrates. Thus, our data indicate that long-term diet is particularly strongly associated with enterotype partitioning. The dietary associations seen here parallel a recent study comparing European children, who eat a typical Western diet high in animal protein and fat, to children in Burkina Faso, who eat high carbohydrate/low animal protein diets (*13*). The European microbiome was dominated by taxa typical of the *Bacteroides* enterotype, whereas the African microbiome was dominated by the *Prevotella* enterotype, the same pattern seen here. There are of course many differences between Europe and Burkina Faso that might influence the gut microbiome, but dietary differences provide an attractive potential explanation. Having confirmed enterotype partitioning and established the association with dietary patterns, it will be important to determine whether individuals with the *Bacteroides* enterotype have a higher incidence of diseases associated with a Western diet, and whether long-term dietary interventions can stably switch individuals to the *Prevotella* enterotype. If an enterotype is ultimately shown to be causally related to disease, then long-term dietary interventions may allow modulation of an individual's enterotype to improve health.

References and Notes

1. L. V. Hooper, J. I. Gordon, *Science* **292**, 1115 (2001).
2. A. Emminger, E. Kahmann, D. S. Savage, *Cancer Lett.* **2**, 273 (1977).
3. S. R. Gill *et al.*, *Science* **312**, 1355 (2006).
4. M. Arumugam *et al.*, *Nature* **473**, 174 (2011).
5. G. D. Wu *et al.*, *BMC Microbiol.* **10**, 206 (2010).
6. M. Margulies *et al.*, *Nature* **437**, 376 (2005).
7. See supporting material on *Science* Online.
8. C. Lozupone, M. Hamady, R. Knight, *BMC Bioinformatics* **7**, 371 (2006).
9. B. H. McArdle, M. J. Anderson, *Ecology* **82**, 290 (2001).
10. R. E. Ley, P. J. Turnbaugh, S. Klein, J. I. Gordon, *Nature* **444**, 1022 (2006).
11. M. Kanehisa, S. Goto, S. Kawashima, Y. Okuno, M. Hattori, *Nucleic Acids Res.* **32** (database issue), D277 (2004).
12. P. J. Rousseeuw, *J. Comput. Appl. Math.* **20**, 53 (1987).
13. C. De Filippo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 14691 (2010).

Acknowledgments: This work was supported by Project UH2DK083981, the Penn Genome Frontiers Institute; NIH AI39368 (GDW); Penn Digestive Disease Center (P30 DK050306); The Joint Penn-CHOP Center for Digestive, Liver, and Pancreatic Medicine; S10RR024525; UL1RR024134, K24-DK078228; and the Howard Hughes Medical Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center for Research Resources, National Institutes of Health, or Pennsylvania Department of Health. Accession numbers (Sequence Read Archive): CaFE: SRX021237, SRX021236, SRX020587, SRX020379, SRX020378 (metagenomic). COMBO: SRX020773, SRX020770, SRX089367.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1208344/DC1
Materials and Methods
Figs. S1 to S5

Downloaded from www.sciencemag.org on September 1, 2011

Tables S1 to S10
References (*14–21*)

13 May 2011; accepted 17 August 2011
Published online 1 September 2011;
10.1126/science.1208344

Fig. 1. Correlation of diet and gut microbial taxa identified in the cross-sectional COMBO analysis. Columns correspond to bacterial taxa quantified using 16S rDNA tags, rows correspond to nutrients measured by dietary questionnaire. Red color indicates positive association and blue color negative association. The intensity of the colors represents the degree of association between the taxa abundances and nutrients as measured by the Spearman's correlations. Bacterial Phyla are summarized by the color code on the bottom, lower level taxonomic assignments specified are in fig. S1. The * indicates the associations that are significant at the false discovery rate (FDR) of 25%. The FFQ data was used for this comparison (both FFQ and Recall dietary data are shown together in fig. S1). Columns and rows are clustered by Euclidean distance, with rows separated by the predominate nutrient category.

Fig. 2. Clustering of gut microbial taxa into enterotypes is associated with long-term diet. (**A**) Clustering in the COMBO cross-sectional study using Jensen-Shannon Distance. The left panel shows that the data are most naturally separated into two clusters by the PAM method. The x-axis shows cluster number, the y-axis shows silhouette width, a measure of cluster separation (*12*). The right panel shows the clustering on the first two PCs. (**B**) Proportions of bacterial taxa characteristic of each enterotype. .Boxes represent the interquartile range (IQR) and the line inside represents the median. Whiskers denote the lowest and highest values within 1.5 × IQR. (**C**) The association of dietary components with each enterotype. The strength and direction of each association, as measured by the means of the standardized nutrient measurements, is shown by the color key at the lower right. Enterotype is shown at the right. Red indicates greater amounts, blue lesser amounts of each nutrient in each enterotype (complete lists of nutrients are in table S2). Columns were clustered by Euclidean distance.

Fig. 3. Changes in bacterial communities during controlled feeding. Ten subjects were randomized to high fat/low fiber or low fat/high fiber diets, and microbiome composition was monitored longitudinally for 10 days by sequencing 16S rDNA gene tags (CAFE study). (**A**) Cluster diagram based PCoA analysis using unweighted UniFrac. Colors indicate samples from each individual. (**B**) Day 1 samples are outliers compared to all other days, indicating change in the gut microbiome within 24 hours of initiating controlled feeding. In this analysis, weighted Unifrac distances between samples are compared within subjects in two groups. The first collection of distances compares the day 1 samples to days 2-10; the second group compares samples from all days to all others excluding day 1, indicating rapid change ($p = 0.0003$, 10000 permutations). Error bars indicate 1 SD of the distances.

Downloaded from www.sciencemag.org on September 1, 2011

Bacteroides *Prevotella*

Vitamin E, Food Fortification
Added Germ from wheats
Eriodictyol, flavonone

Carbohydrates

Maltose
Glycemic Index
Sucrose
Total Sugars
Carbohydrates
Glucose
Fructose

Fats

Gamma linolenic fatty acid
Gamma Linolenic fatty acid
Eicosenoic fatty acid
Total Trans/Cis Trans Linoleic
Total Trans
Trans Oleic fatty acid
Dihydrophylloquinone Vitamin K1
Sodium
Palmitelaidic trans fatty acid
Stearic fatty acid
Saturated fat
Palmitic fatty acid
c9,t11 conjug diene isomer 18:2 Linoleic
Animal fat
Hydroxyproline
Palmitoleic fatty acid
Taurine
Cholesterol
Choline, Sphingomyelin
Animal Protein

Amino Acids and Choline

Dairy Calcium
Dairy Protein
Calcium w/o vit. pills
Calcium
Phosphorous w/o suppl.
Phosphorous
Choline, Phosphocholine
Choline, Glycerophosphocholine
Vitamin D w/o vit. pills
Proline
Glutamate
Cystine
Sum of Betaine & Choline
Choline w/o suppl.
Total Choline, no betaine
Choline, Phosphatidylcholine w/o suppl.
Choline, Phosphatidylcholine
Asparate
Arginine
Alanine
Glycine
Tryptophan
Serine
Phenylalanine
Protein
Isoleucine
Valine
Tyrosine
Leucine
Lysine
Methionine
Threonine
Histidine

Riboflavin B2 w/o vit. pills
Folate Equivalents, suppl. & fort. foods
Total Folate post 1998
Retinol Equivalents of Vit A
Retinol

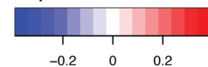
Fiber and Plant Derived Compounds

Caffeine
Aspartamefort. foods
Aspartic Acid, Aspartame
Phenylalanine, Aspartame
Alcohol
Catechin, flavan−3−ol
Proanthocyanidin, dimers
Proanthocyanidin, polymers
Cyanidin, anthocyanidin
Proanthocyanidin, 7−10mers
Proanthocyanidin, trimers
Proanthocyanidin, 4−6mers
Vitamin E w/o vit. suppl.
Naringenin, flavanone
Pantothenic Acid w/o suppl.
AOAC fiber
Natural Food Folate
Free Choline w/o suppl.
Free Choline, choline−contrib. metabolite
Magnesium w/o suppl.
Magnesium
Potassium w/o suppl.
Potassium
Pelargonidin, anthocyanidin
Delphinidin, anthocyanidin
Total anthocyanidins
Petunidin, anthocyanidin
Malvidin, anthocyanidin
Peonidin, anthocyanidin

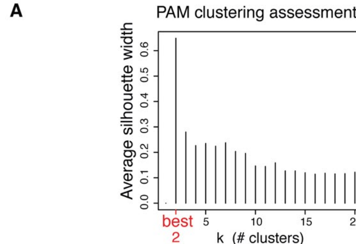
Phylum

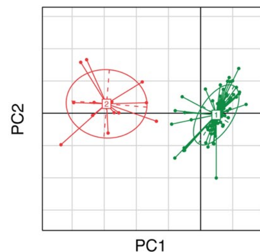
Firmicutes
Bacteroidetes
Actinobacteria
Proteobacteria

Spearman Correlation

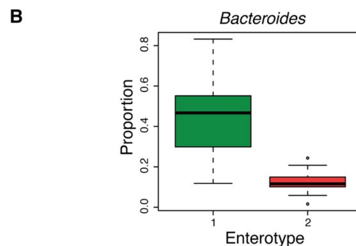


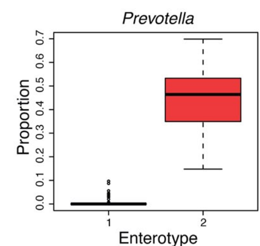
A



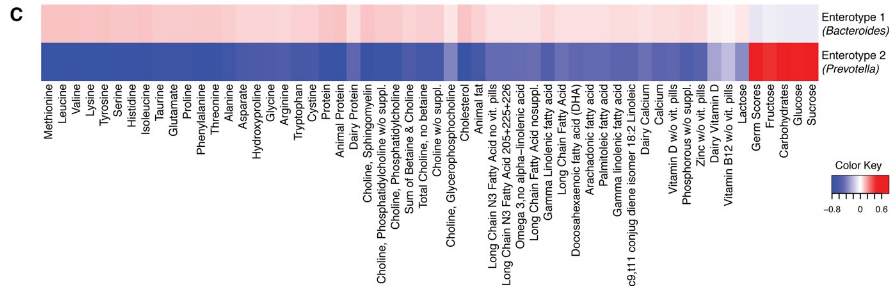


B





C



A

PC3 (3.8%)

PC1 (7.9%)

PC2 (4.5%)

B

Day 1 to others

Between other days

Weighted UniFrac Distance

0.0 0.1 0.2 0.3 0.4 0.5

2004 2006 2009 2016 2019 2005 2008 2011 2012 2020

Low Fat Diet

High Fat Diet

Subject Number