www.sciencemag.org/content/357/6353/802/suppl/DC1

Supplementary Materials for

Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania

Samuel A. Smits,* Jeff Leach,* Erica D. Sonnenburg, Carlos G. Gonzalez, Joshua S. Lichtman, Gregor Reid, Rob Knight, Alphaxard Manjurano, John Changalucha, Joshua E. Elias, Maria Gloria Dominguez-Bello, Justin L. Sonnenburg†

*These authors contributed equally to this work.
†Corresponding author. Email: jsonnenburg@stanford.edu

Published 25 August 2017, *Science* **357**, 802 (2017)
DOI: 10.1126/science.aan4834

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:
(available at www.sciencemag.org/content/357/6353/802/suppl/DC1)

Tables S1, S3, and S4 (CSV format)

Materials and Methods

Sample Collection

Samples for this study were collected in 2013 and 2014 spanning 5 sub-seasons (n=64, 2013-LD; n=62, 2014-EW; n=58, 2014-LW; n=33, 2014-ED; n=133, 2014-LD**; Fig. S1**), were limited to those from individuals more than 3 years of age and to two adjacent (less than 7km) camps, Sengeli and Hukamako. The number of samples collected was not pre-specified for ensuring adequate statistical power. Permission for the study was obtained from the National Institute of Medical Research (MR/53i 100/83, NIMR/HQ/R.8a/Vol.IX/1542) and the Tanzania Commission for Science and Technology. We obtained verbal consent from the Hadza after having described the study's intent and scope. At the time of sample collection when available we recorded the age (either provided by the individual or estimated based on information provided in their recall), gender, weight and related information. After samples were retrieved from the individuals they were immediately placed into liquid nitrogen containers in the field, transported to Arusha frozen and mailed on dry ice by air to the United States and maintained at -80˚C until further processing.

16S rRNA gene sequencing and analysis

DNA was extracted from the samples according to the Earth Microbiome Project (EMP) standard protocols (*32*) using Powersoil-htp extraction kits (MO BIO Laboratories) and EMP 515F/806R primers as previously described (*33*). The libraries were sequenced on an Illumina MiSeq platform yielding 250nt per read. The QIIME 1.9.1 package (*34*) was used for performing quality filtering and demultiplexing, and the reads were assigned to open-reference OTUs using UCLUST (*35*) with a threshold of 97% identity and seeded with Greengenes 13.8 database (*36*) sequences. The resulting OTU counts per sample were rarefied to 11,000. We analyzed diversity with QIIME using UniFrac (*37*) that depended on PyNAST alignment (*34*) and FastTree for tree-building, along with Shannon and Chao1 metrics (*38*). We used R *vegan* and *ggplot2* libraries for analyses and visualization, and Python's *numpy*, *scipy*, *scikit-learn* (*39*) and *scikit-bio* (scikit-bio.org) for OTU-tracking, machine learning and statistics. We generated custom code for producing visualizations including the geographic map, Sankey plots, steamgraph and heatmaps using D3.js (d3js.org).

Volatility analysis was performed by assembling a list of OTU abundances across all individuals (n=8) that were tracked over all three seasons. The OTU list was then filtered to a total abundance threshold that represented at least 0.1% of an individual's rarefied OTU count (at least 11 OTUs). OTUs were then aggregated to taxonomic families (taxonomic level 5) and further filtered to only those families that consisted of at least 10 different OTUs, resulting in a total of 13 families. We then calculated a Volatility Statistic between the Dry and Wet season per taxonomic family for each individual:

$$Volatility\ Statistic = \log_2\left(\frac{Number\ of\ OTUs\ observed\ in\ the\ Dry\ season\ but\ not\ in\ the\ Wet}{Number\ of\ OTUs\ observed\ in\ the\ Wet\ season\ but\ not\ in\ the\ Dry}\right)$$

These Volatility Statistics are plotted in **Fig. S3C**. Families defined to possess the most abundant representation of volatile OTUs are those whose final absolute mean statistic exceeded 2: Prevotellaceae, Spirochaetaceae, Paraprevotellaceae and Succinivibrionaceae.

16S rRNA Sequence Meta-Analyses

Raw sequencing data and mapping data were downloaded for all studies included in the analyses (**Table S1, S5**). All sequences were trimmed to 100nt and closed-reference OTU picking was performed using UCLUST (*35*) at 97% identity with Greengenes 13.8 (*36*)

sequences using QIIME 1.9.1 (*34*). Each sample was rarefied to 5,000 OTUs and those that did not have at least 5,000 OTUs were excluded from the analysis. In order to overcome challenges presented by 16S rRNA amplicon sequencing datasets targeting varying gene regions, we evaluated paired HMP samples (**Table S6**) that were sequenced using primers targeting different 16S gene regions (V1-V3 and V3-V5) for robustness of taxonomic assignments according to taxonomic levels. Taxonomic levels 1-5 are capable of describing at least 93% of the OTUs identified (**Table S7**), performed significantly better than a random sampling set (**Fig. S8A**), and Bray-Curtis dissimilarity PCoAs performed significantly better than random sampling sets (**Fig. S8B, S8C**), while taxonomic levels 1-5 retained most of the information (**Fig. S8D**). We therefore performed all subsequent analysis at taxonomic level 5 (Family), and applied Bray-Curtis ordination to the taxonomic profiles, a method which treats each family equally and does not penalize taxonomies that may be underrepresented in reference databases.

Metagenomic Sequencing and Metabolomics Processing

We selected 40 samples across the seasons by attempting to maximize the same subjects across seasons for comparative analyses. For this study, analyses of these samples were limited to 35 samples by eliminating any repeated sampling within a single season (n=3) and those under 3 yrs of age (n=2). We processed the samples using ~100mg of fecal material from each and extracted the DNA by first transferring the frozen material into an Eppendorf tube with ~500ul acid-washed and autoclaved 0.1mm Zirconia beads (BioSpec Products), 500ul of extraction buffer (200mmol/l each of Tris (pH 8.0) and NaCl, 20mmol/l EDTA), 210ul of 20% SDS and 500ul of phenol:chloroform:isoamyl alcohol (25:24:1). The sample was then subjected to bead beating on high for 4 minutes at 4C, after which we extracted the DNA after multiple rounds of centrifugation, aspiration and resuspension in phenol:chloroform:isoamyl alcohol, and finally precipitated with isopropanol. The extracted DNA was fragmented using a Covaris M220 and libraries prepared using the TruSeq DNA PCR-Free Library Preparation Kit (Illumina) and sequenced on an Illumina HiSeq 4000.

The same samples were also separately processed for metabolomics and kept on ice or at 4C during processing. The samples were weighed and transferred to a 2ml Eppendorf tubes with autoclaved and acid-washed 0.1mm Zirconia beads (BioSpec Products), to which 100ul of acidified ddH_2O (0.01% Formic Acid (FA)) and 100ul acetonitrile (ACN) were added and then subjected to bead-beating for 4 minutes on high at 4C. They were then centrifuged at 9x1000rcf for 10 minutes at 4C, aspirated, diluted with ddH_2O and FA to a final 5:5:90 ACN:FA:ddH_2O concentration and filtered with 13mm Fisherbrand 0.22um filters that were prepared by pre-eluting 2ml of MeOH. The extraction was maintained at -80C until run on the instrument, at which time they were further diluted (1:5-1:100). Injections were 1-2ul using reversed-phase chromatography on a Dionex Ultimate 3000 HPLC with an in-house laser-pulled 100um ID nanospray column packed to ~150mm with 3um 2A C18 beads (Reprosil). Buffer A of the mobile phase contained 0.2% FA and 5% DMSO in HPLC-grade water, while buffer B contained 0.2% FA and 5% DMSO in ACN. An initial 3 minute isocratic gradient flowing 3% B was followed by a linear increase up to 95% B for 18 minutes whereupon B was held for 3 minutes and returned back to baseline (1 min) and held for 14 minutes, for a total of 35 minutes. The HPLC flow rate was 0.400 uL/minute. The HPLC was coupled to a Thermo Orbitrap Velos mass spectrometer that collected MS data in positive ion mode within the 95-1000 m/z range. A top-five MS/MS method was employed with an initial Orbitrap scan resolution of 60,000. This was followed by collision-induced dissociation and analysis of top 5 analytes in the ion trap with

dynamic exclusion enabled (repeat count of 1, exclusion duration of 15 s). The automatic gain control for FT full MS was set to 5e5 and for ITMSn was set to 1.5e5. The data were aligned using XCMS (*40*) and analyses were performed using Python libraries *numpy*, *scipy* and *scikit-learn* (*39*).

Metagenomic Sequencing Analyses

The metagenomic reads were filtered for human DNA with DeconSeq (*41*) using Genome Reference Consortium GRCh38 sequences (assembly date December 2013). We used 32 HMP metagenomic samples to serve as controls and processed them together with our Hadza metagenomic reads similarly in the following steps (**Table S8**) (*23*). The reads were quality filtered and trimmed using DynamicTrim (*42*) and coding sequences identified with FragGeneScan (*43*). Reads were assigned to KEGG pathways using HUMAnN2 (*44*) with default parameters. We used HMMER 3.0 (*45*) to identify CAZymes using dbCAN (*46*) and antibiotic resistance genes with ResFams (*47*), with E-value<1e-5 cutoff. CAZymes relevant to the utilization of various carbohydrate sources were gleaned from literature and are presented in **Table S9** (*48*). The reads were normalized to the number of coding sequences identified per sample. Diversity analyses were performed with Python libraries *numpy*, *scipy*, *scikit-learn* (*39*) and *scikit-bio* (scikit-bio.org).

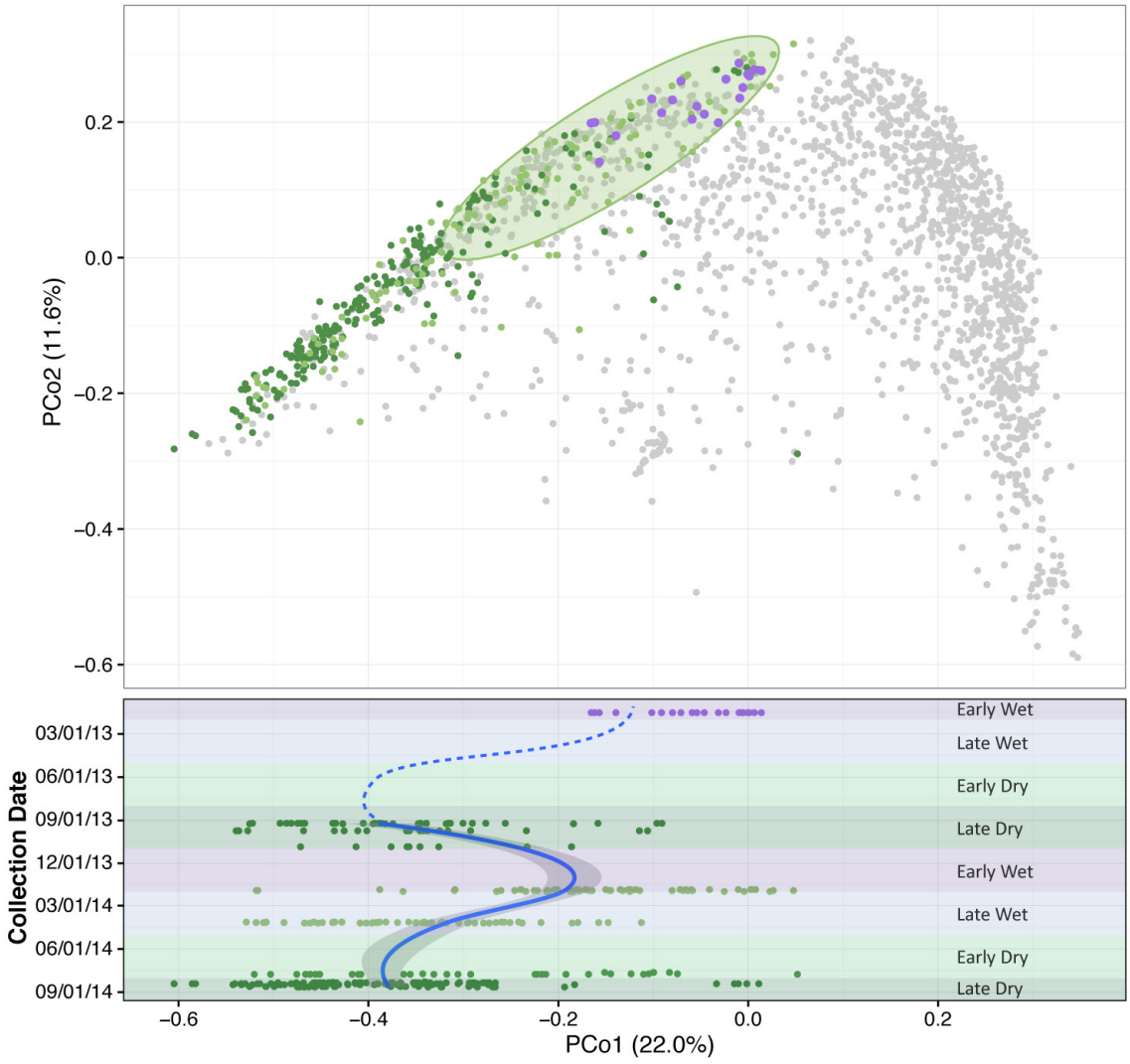

Fig. S1.

Seasonal fluctuations of the Hadza gut microbiome over two years.
Top panel. The Bray-Curtis PCoA from Figure 3 is adapted to highlight the Hadza samples from this study and a previous study by Schnorr et. al. This study's samples are colored in green (n=350), Schnorr's in purple (n=20) and an ellipse captures those samples from this study that were collected in the Early-Wet season of 2014 (Schnorr's samples were collected in the Early-Wet season of 2013). Both studies' samples co-segregate with each other and with traditional population samples. **Bottom panel.** The Hadza samples are plotted according to their collection date on the y-axis and their position on the first principal coordinate of the Bray-Curtis PCoA in the top panel on the x-axis. The sub-seasons are labeled and indicated by shading. Loess regression was applied to these points using the collection date and PCo1 coordinates, and the curve was plotted in blue with a 95% pointwise confidence interval band in gray on the plot using the data within this study. The dashed blue line is a continuation of the regression curve yet is an implied regression curve assuming the appropriate inflection points are captured with data from our study.

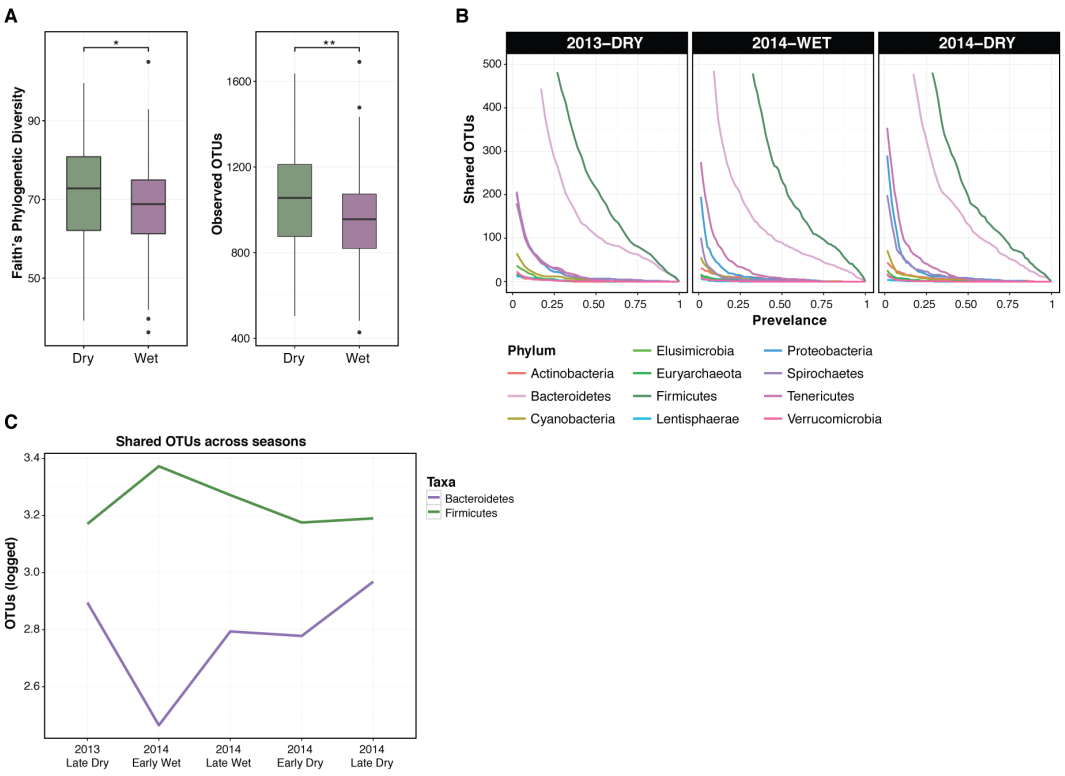

Fig. S2

The incidence of individual microbes varies across the seasons in the Hadza microbiome. (**A)** Alpha-diversity metrics reveal lower phylogenetic diversity and observed OTUs in the Wet seasons (p=0.02 and p=9e-4, respectively; Wilcoxon). Boxplot distributions are tested using the non-parametric two-sided Wilcoxon rank sum test with Holm correction for multiple hypothesis testing, center values indicate the median and error bars the s.d. *, p-values < 0.05, ** < 0.01. (**B)** The number of shared OTUs of phyla are plotted by their prevalence (percent incidence) in the population and aggregated at the phylum level for indicated seasons (LD: Late-Dry; ED: Early-Dry; LW: Late-Wet; EW: Early-Wet) (left panel); colors correspond to phylum as indicated **(C)** The logged abundances of shared Bacteroidetes and Firmicutes OTUs by at least 10 percent of the population at each timepoint is plotted.

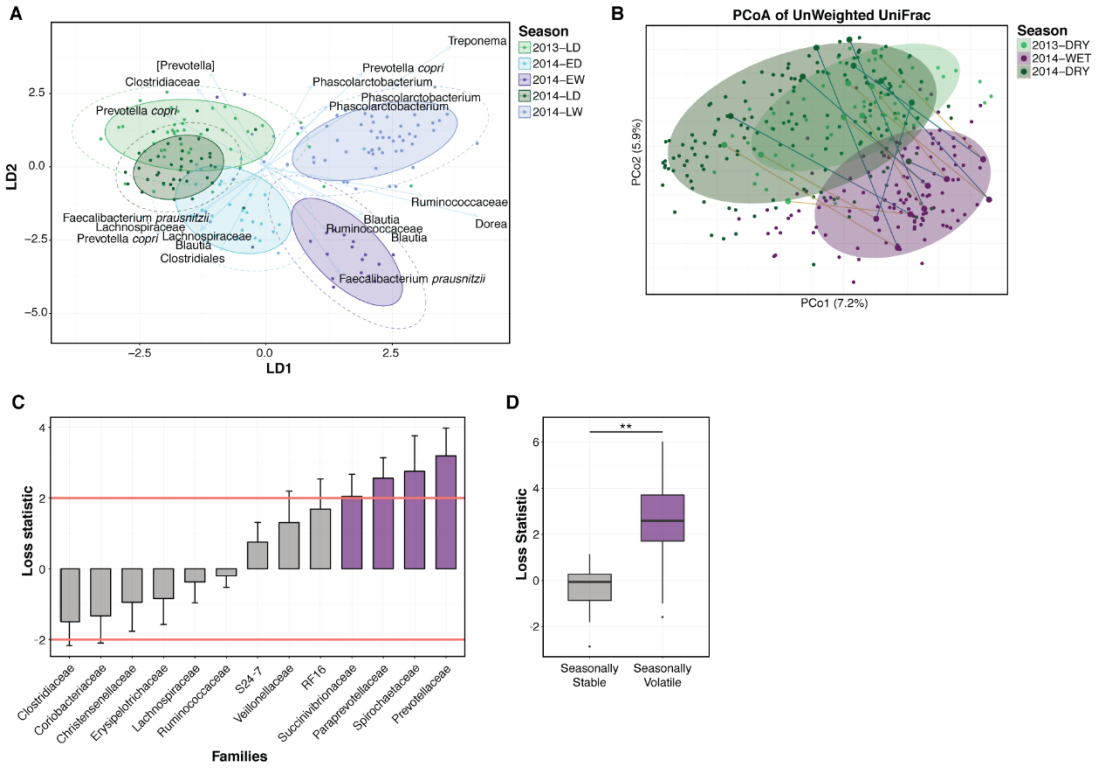

Fig. S3

Bacterial taxa show differential volatility and representation across seasons within the Hadza microbiome.

(**A)** Linear discriminant analysis (LDA) was used to differentiate between microbiota composition in sub-seasons across years. The two Late-Dry seasons of sequential years are too similar for LDA to differentiate. The length and direction of the arrows indicate the normalized scalings for each of the features (OTUs). **(B)** The eight individuals with repeated sampling across the three seasons is plotted on an unweighted UniFrac PCoA. Colored lines connect the samples from the same individual across seasons (yellow: 2013-Dry to 2014-Wet; blue: 2014-Wet to 2014-Dry). (**C)** Summary statistics of volatile OTUs in individuals tracked across seasons. The barplots indicate the mean Volatility Statistic, and the error bars indicate the standard deviation. Those with more than a 2-fold loss statistic change are colored purple. (**D)** Volatility Statistics aggregated from the four most volatile families (Succinivibrionaceae, Paraprevotellaceae, Spirochaetaceae and Prevotellaceae) are significantly more volatile than those families exhibiting seasonal stability (Ruminococcaceae, Lachnospiraceae; **Fig. 1F**) (p=7e-13, Wilcoxon). Boxplot distributions are tested using the non-parametric two-sided Wilcoxon rank sum test with Holm correction for multiple hypothesis testing, center values indicate the median and error bars the s.d. *, p-values < 0.05, ** < 0.01. The ellipses in all plots represent the 80% confidence interval, dotted elipses represent the 95% confidence interval.

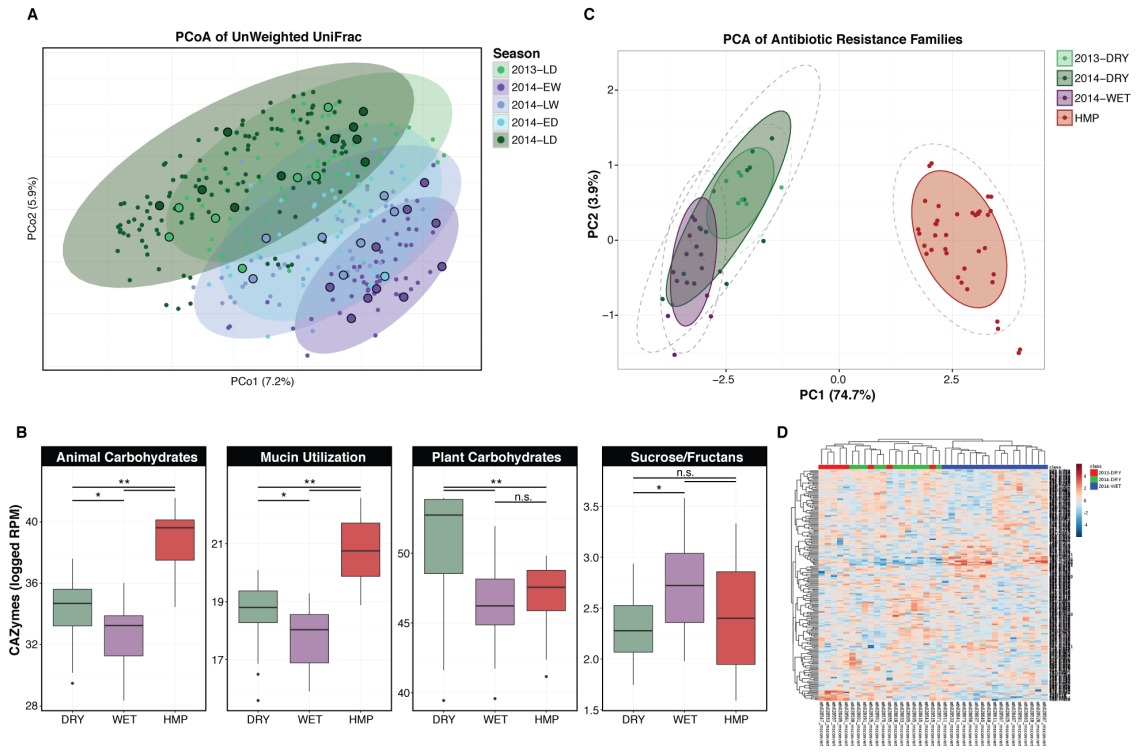

Fig. S4

Functional differences between the American and Hadza gut microbiome, and between seasons within the Hadza. (A) The 35 samples used for metagenomics are highlighted on an unweighted UniFrac PCoA plot. (**B)** Hadza Wet-season microbiotas are enriched for CAZYmes involved in the utilization of sucrose and fructans; the enrichment over the Dry season (p=0.03, Wilcoxon) and American population (p=0.03, Wilcoxon) coincides with a season increased in berry-foraging activities. The Hadza Dry-season capacity for sucrose and fructan utilization does not differ from the American reference population. All boxplot distributions are tested using the non-parametric two-sided Wilcoxon rank sum test with Holm correction for multiple hypothesis testing, center values indicate the median and error bars the s.d. p-values, * < 0.05, ** < 0.01. **(C)** The repertoire of Antibiotic Resistance Genes (ARGs) are distinct between the Hadza and US residents (HMP). Despite the remoteness of many hunter-gatherer populations, antibodies against many crowd diseases have been identified in remote Tanzanian and South American populations (*49, 50*). Similarly, geographic divide is not a significant barrier for the passage of antibiotic resistant genes to remote populations, and we wondered if the repertoire of these elements might differ in Hadza (*15, 51*). Unsupervised learning could differentiate the repertoire of ARGs, which were found to be completely differentiable between populations (pseudo-F=497.1, p=0.001; PERMANOVA). (**D)** Untargeted gut metabolome features differentiate between the samples from Wet and Dry seasons, but not the Dry seasons of two different years. The top 200 features with the greatest effect size were used to differentiate the groups. Unsupervised clustering methods resolved differences in between the Dry and Wet seasons perfectly. The ellipses in all plots represent the 80% confidence interval, dotted elipses represent the 95% confidence interval.

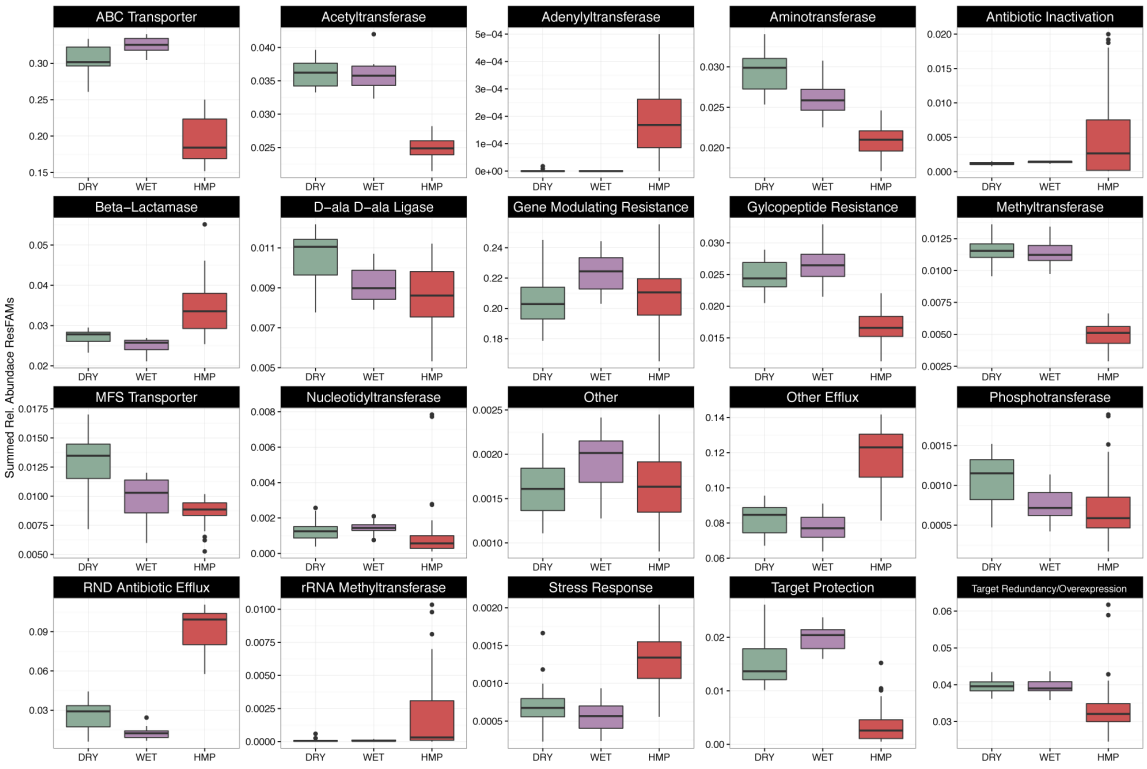

Fig. S5

The relative abundance of Antibiotic Resistance Genes (ARGs) encoded in microbial metagenomes varies between the American and Hadza gut microbiome, and between seasons within the Hadza. The relative abundance of reads assigned to ARG categories are shown as distributions across the Hadza metagenomes obtained in Dry (green) and Wet (purple) seasons, and samples from the Human Microbiome Project (HMP). Center values of the boxplot distributions indicate the median and error bars the s.d.

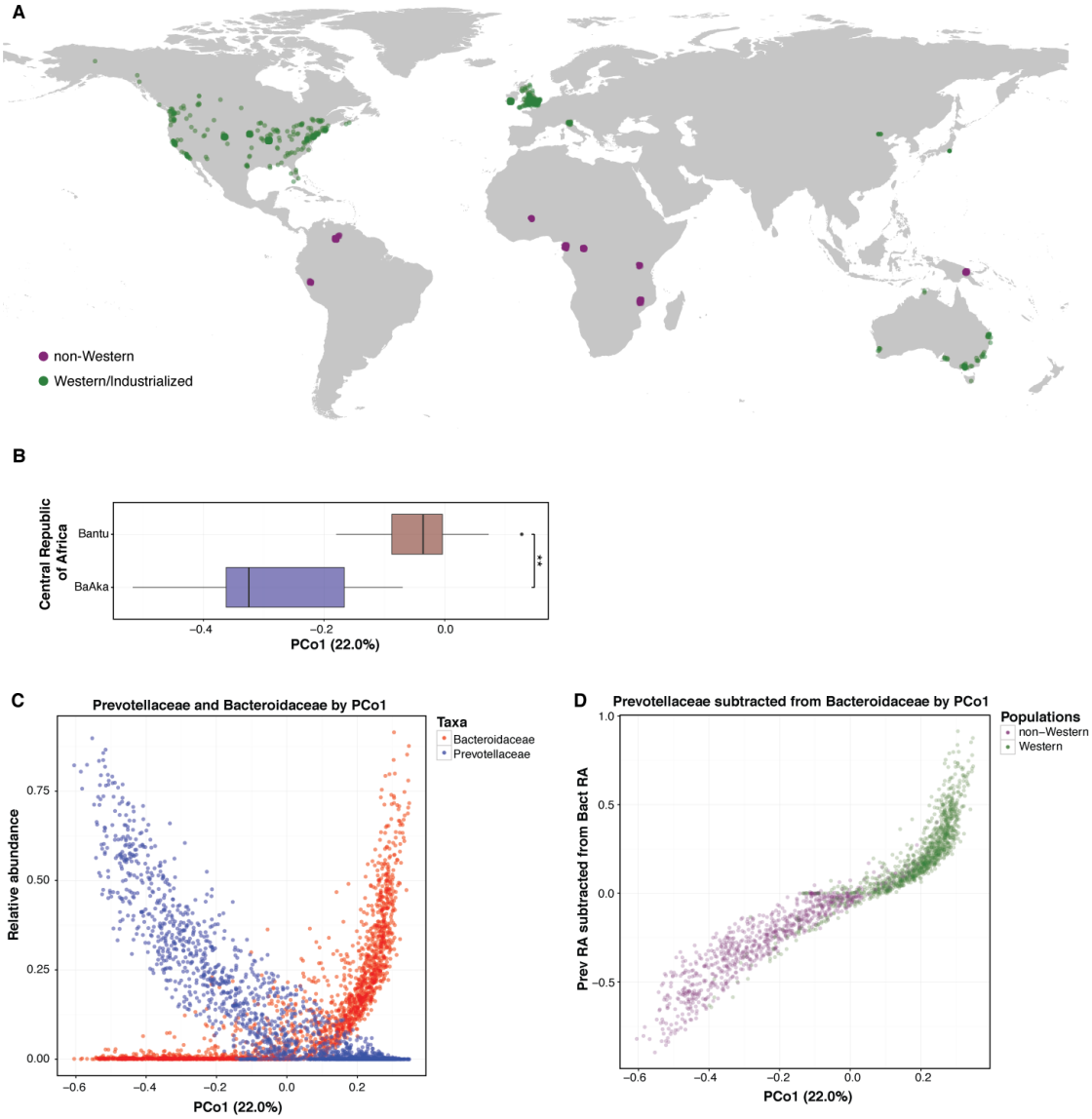

Fig. S6

Common patterns are found around the world in microbiome compositional features. (A) Geographic map indicating where microbiota samples were collected. The points were plotted according to annotated data where latitude and longitude coordinates were provided. In cases where the geographic coordinates were not provided, we inputted coordinates according to a description of where the samples were collected in the relevant papers. Points are colored according to a Western or non-Western lifestyle described for the populations in the originating papers. **(B)** Individuals living in the Central Republic of Africa are differentiated according to their lifestyle (Bantu agriculturalists and BaAka hunter-gatherers) on a PCoA plot (**Figure 3**) on PCo1 ($p<9e-13$, Wilcoxon). Boxplot distributions are tested using the non-parametric two-sided Wilcoxon rank sum test with Holm correction for multiple hypothesis testing, center values indicate the median and error bars the s.d. **(C)** Prevotellaceae and Bacteroidaceae have an inverse relationship. Samples plotted in **Figure 3** are re-plotted here on the y-axis according to their

relative abundance of Bacteroidaceae (red), and Prevotellaceae (blue) versus the first principal coordinate of the original PCoA. (**D)** Prevotellaceae and Bacteroidaceae relative abundances are correlated across traditional and Western populations. The relative abundances of Prevotellaceae were subtracted from the relative abundances of Bacteroidaceae and plotted on the y-axis, while their position on PCo1 (x-axis) are maintained. A latent continuous variable arising from such data has been previously described (*27, 28*).

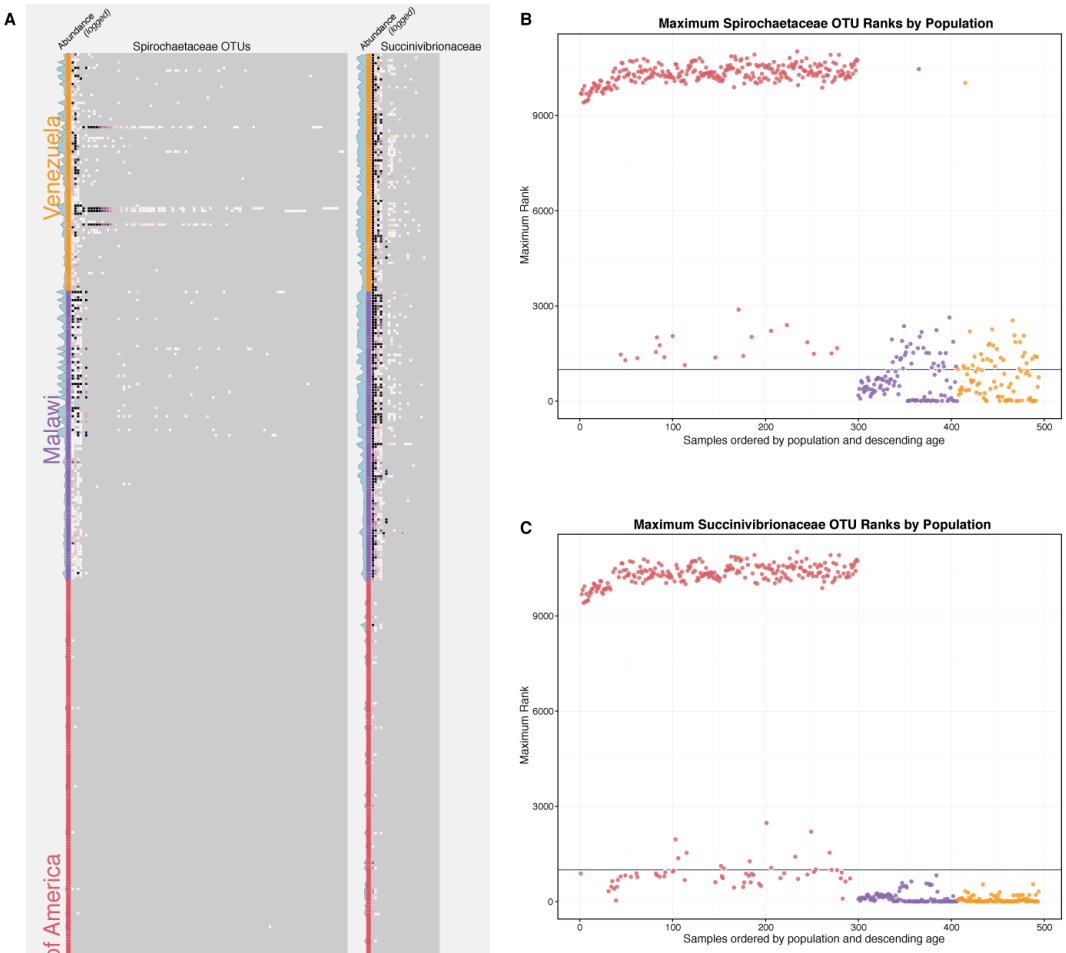

Fig. S7

Prevalent families within traditional populations, Succinivibrionaceae and Spirochaetaceae, are rare or absent in Western microbiotas. (A) All unique OTUs (x-axis) identified within these two families across samples (y-axis) with at least 1 million assigned reads are represented in the heatmap. The summed logged relative abundance of all identified OTUs

within the sample are represented in the line graphs. **(B)** All OTUs identified as Spirochaetaceae were ranked by abundance per sample, with the maximum ranked Spirochaetaceae OTU plotted. The blue line represents the 1000th ranked OTU threshold. (C) All OTUs identified as Succinivibrionaceae were ranked by abundance per sample, with the maximum ranked Succinivibrionaceae OTU plotted. The blue line represents the 1000th ranked OTU threshold.

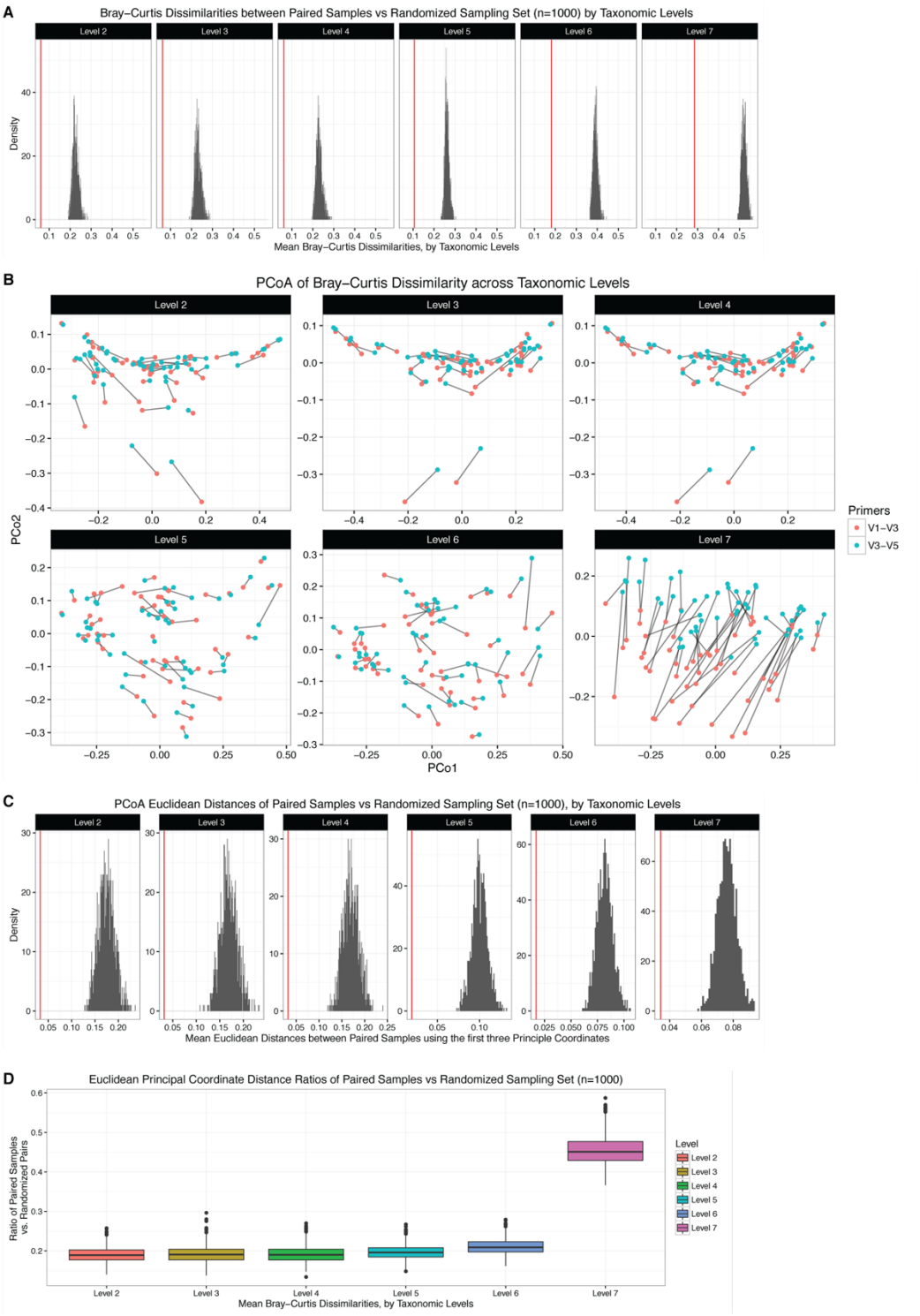

Fig. S8

Comparison of microbiome profiles at different taxonomic levels for V1-V3 versus V3-V5 16S rRNA regions amplified and sequenced from the same stool sample. (A) The mean

dissimilarity between paired samples was calculated (vertical red line in plots), and a randomized sampling set (n=1000) of mean dissimilarities was calculated by shuffling the labels at random. The data show that at every taxonomic level paired samples retain more similarity than likely to be explained by random shuffling. (**B)** Principal coordinates analyses of Bray-Curtis dissimilarities**.** Red and salmon colored circles are the V1-V3 and V3-V5 datasets, respectively, and lines join identical samples across the sequencing runs. Levels 5 and 6 show data that are well dispersed suggesting retention of differentiating information, whereas level 7 reveals significant systematic dissimilarity within paired samples (longer connecting lines). (**C)** Euclidean distances between paired samples versus distances between randomized sampling set. Level 7 histograms show that the variance between randomized sampling paired sets and the variance between paired samples are significantly more similar than all other levels. Histograms were calculated by randomly shuffling labels (n=1000) and calculating the mean Euclidean distances. The red vertical line indicates the true mean of the Euclidean distance between paired samples. (**D)** Ratios of Euclidean distances between paired samples and randomized paired sampling. The Level 7 distribution is significantly higher than all other levels, indicating that less differentiating information between samples are conserved across the paired samples. Boxplot center values indicate the median and error bars the s.d.

Table S1.

Samples incorporated in meta-analyses. A list of all samples for which publically-available sequencing datasets were downloaded and analyzed (americangut.org) (*13-17, 19, 23, 52-54*). Additionally, columns "Metagenomics", "SEASONALITY_UNIQUE_ONLY_1" and "8_individuals" indicate the Hadza samples in this study used in these particular analyses.

Table S2.

Top 20 differentiating species between sub-seasons using Linear Discriminant Analysis. A list of OTUs that have the greatest scalings and therefore influence on the separation of the sub-season microbial communities.

Phylum	Taxon	LD1	LD2
Spirochaetes	Treponema	-2.831378164	4.789367564
Firmicutes	Faecalibacterium prausnitzii	-1.432512183	-4.2833568
Bacteroidetes	Prevotella copri	-1.300532615	3.619762979
Firmicutes	Blautia	-2.431455116	-1.856358489
Bacteroidetes	Prevotella copri	1.728255305	-2.575572826
Firmicutes	Phascolarctobacterium	-1.536529914	2.682235025
Firmicutes	Lachnospiraceae	1.145963462	-3.56332021
Bacteroidetes	Prevotella copri	-2.489647716	1.606536621
Firmicutes	Clostridiaceae	1.128514312	3.298200781
Firmicutes	Coprococcus	-1.437660688	2.535283573
Firmicutes	Ruminococcaceae	-2.542096184	-1.298789137
Bacteroidetes	[Prevotella]	1.126406407	2.816404125
Bacteroidetes	Prevotella copri	1.260384862	-2.488762085
Firmicutes	Blautia	1.156103097	-2.659617047
Bacteroidetes	Prevotella copri	1.789338057	1.708711109
Firmicutes	Ruminococcaceae	-1.694313079	1.780940526
Spirochaetes	Treponema	0.688454241	-4.099192496
Firmicutes	[Ruminococcus]	-0.9326603	-3.014316791
Firmicutes	Ruminococcaceae	-1.036890955	-2.585859848
Firmicutes	Ruminococcaceae	-1.308683411	2.030038687

Table S3.

Relative abundance of genes assigned to KEGG Category pathways in Hadza and Human Microbiome Project metagenomes. The relative abundance of reads in each sample that are assigned to KEGG Category pathways are listed by sample, project and season (where applicable).

Table S4.

Relative abundance of genes assigned to KEGG Carbohydrate Metabolism pathways in Hadza and Human Microbiome Project metagenomes. The relative abundance of reads in each sample that are assigned to KEGG Carbohydrate Metabolism pathways are listed by sample, project and season (where applicable).

Table S5.

Studies incorporated in meta-analyses. A list of all studies for which publically-available sequencing datasets were downloaded and analyzed (americangut.org) (*13-17, 19, 23, 52-55*). Some studies are further sub-categorized according to country or surveyed population. The calculated mean and standard deviation of the relative abundances of taxa by study are shown.

Country	Population	Cohort	Study	Bacteroidaceae Mean	Bacteroidaceae Std. Dev.	Prevotellaceae Mean	Prevotellaceae Std. Dev.	Spirochaetes Mean
Venezuela	Venezuela-Yanomami	Yanomami	Clemente, JC	0.042640985	0.096945013	0.566918365	0.216993844	0.000529624
Venezuela	Venezuela-Guahibo	Guahibo	Yatsunenko, T	0.021146189	0.052449591	0.169920802	0.11273392	0.004123553
United Kingdom	United Kingdom	United Kingdom	AGP_United Kingdom	0.206059905	0.137053297	0.080954908	0.136074445	2.18E-05
USA	USA	USA	Yatsunenko, T	0.148667691	0.107215025	0.017865216	0.055266915	4.39E-07
USA	USA	USA	HMP v35	0.41670006	0.217115087	0.020702242	0.068295319	2.73E-06
USA	USA	USA	HMP v13	0.433945396	0.195921446	0.035196605	0.095134713	0
USA	USA	USA	AGP_USA	0.28916648	0.169706882	0.040143475	0.101842817	0.000526985
USA	USA	Norman	Obregon-Tito, AJ	0.051164044	0.080754249	0.010736089	0.035310373	5.86E-05
Tanzania	Tanzaznia-Hadza	Hadza	Smits, SA	0.004602549	0.013716683	0.375284453	0.202854996	0.051918068
Tanzania	Tanania-Hadza	Hadza	Schnorr, SL	0.001795487	0.001299744	0.080391722	0.062969305	0.03095364
Peru	Peru	Tunapuco	Obregon-Tito, AJ	0.000518567	0.000865524	0.388462282	0.248761458	0.072686616
Papua New Guinea	Papua New Guinea	Nauna Village (Sausi)	Martinez, I	0.001178316	0.002678695	0.027113883	0.017055402	0.000289237
Papua New Guinea	Papua New Guinea	Asaro Valley	Martinez, I	0.004309885	0.012228371	0.027069973	0.010021909	0.00025744
Malawi	Malawi	Malawi	Yatsunenko, T	0.008467165	0.033232491	0.204799068	0.126572155	0.0061134
Japan	Japan	Japan	Nakayama, J	0.092936499	0.044913195	0.045468592	0.043523376	0
Italy	Italy	Italy	Schnorr, SL	0.077009709	0.081625891	0.003940417	0.010416325	3.06E-05
Italy	Italy	Italy	De Filippo, C	0.148730837	0.107658257	0.003119597	0.00651162	1.62E-05
Ireland	Ireland	Ireland	Claesson, MJ	0.246030567	0.156641187	0.009291253	0.024554066	2.23E-07
China	China	China	Nakayama, J	0.124111633	0.082633126	0.082802804	0.130969503	0
Canada	Canada	Canada	AGP_Canada	0.29584755	0.162792663	0.032805014	0.085858374	0
Cameroon	Cameroon	Pygmy	Morton, ER	0.007317948	0.02810863	0.191760033	0.145741488	0.007741059
Cameroon	Cameroon	Bantu	Morton, ER	0.002988122	0.005223492	0.274630524	0.20151479	0.019728938
Central African Republic	CAR-Bantu	Bantu	Gomez, A	0.010665166	0.033102467	0.062924203	0.047873366	0.004962307
Central African Republic	CAR-BaAka	BaAka	Gomez, A	0.001687046	0.004277162	0.351862143	0.170259441	0.009287682
Burkina Faso	Burkina Faso	Burkina Faso	De Filippo, C	0.003489556	0.007626952	0.528215247	0.262084296	0.004967505
Australia	Australia	Australia	AGP_Australia	0.213348128	0.14539617	0.019475407	0.050900743	0

Spirochaetes Std. Dev.	Succinivibrionaceae Mean	Succinivibrionaceae Std. Dev.	Verrucomicrobia Mean	Verrucomicrobia Std. Dev.
0.000942161	0.034100437	0.046064673	4.85E-06	1.45E-05
0.010410716	0.037668837	0.05578374	0.001115096	0.003498309
0.000264051	0.000626963	0.005040135	0.018968266	0.049216296
5.50E-06	7.42E-06	7.56E-05	0.012051736	0.025509717
1.61E-05	0	0	0.005609641	0.020343195
0	0	0	0.007040841	0.024305977
0.007360901	9.63E-06	7.30E-05	0.010157917	0.027296431
2.35E-05	3.73E-05	2.17E-05	0.048219429	0.107959432
0.074450956	0.031468408	0.05879793	0.000253662	0.003029955
0.053470273	0.033414014	0.048891535	0	0
0.099685178	0.044050334	0.083605191	0.000215359	0.001040939
0.000792542	0.000373593	0.001418094	0.00519832	0.022314687
0.000833933	0.000102887	0.000277503	0.000426972	0.00110096
0.018757111	0.03200643	0.055409294	0.000491453	0.002612723
0	0	0	0.001676449	0.005702328
6.59E-05	0.000195553	0.000678568	0.001313722	0.002990558
3.54E-05	2.52E-05	7.96E-05	0	0
2.97E-06	0.00089181	0.010580761	0.001856127	0.003820279
0	6.29E-05	7.76E-05	0.001119084	0.00346707
0	1.28E-07	8.23E-07	0.031459064	0.070301736
0.010608748	0.120447505	0.13840355	3.61E-05	9.05E-05
0.038326618	0.040255529	0.060642559	0.000315224	0.001566441
0.012054773	0.005019365	0.007523674	0.000303682	0.000903868
0.016148471	0.007477908	0.010698817	1.98E-05	5.66E-05
0.005614893	0.017895956	0.023505528	0	0
0	0.000151832	0.001199919	0.02037787	0.044280349

Table S6.
HMP samples for which 16S rRNA amplicons were generated using both V1-V3 primers and V3-V5 primers. A list of all 53 HMP samples that were used in analyses for assessing potential primer bias.

SRS_SampleID	RunID_v13	RunID_v35	SampleID_v13	SampleID_v35
SRS011410	SRR045027	SRR044946	SRS011410.SRX020659.V13	SRS011410.SRX020666.V35
SRS014287	SRR045664	SRR045506	SRS014287.SRX020511.V13	SRS014287.SRX020551.V35
SRS014442	SRR045656	SRR045494	SRS014442.SRX020511.V13	SRS014442.SRX020551.V35
SRS014659	SRR045649	SRR045497	SRS014659.SRX020511.V13	SRS014659.SRX020551.V35
SRS014823	SRR043699	SRR045736	SRS014823.SRX020546.V13	SRS014823.SRX020555.V35
SRS014885	SRR043822	SRR049220	SRS014885.SRX020554.V13	SRS014885.SRX020558.V35
SRS014999	SRR044050	SRR044015	SRS014999.SRX020577.V13	SRS014999.SRX020535.V35
SRS015190	SRR043731	SRR047400	SRS015190.SRX020554.V13	SRS015190.SRX020558.V35
SRS015247	SRR044079	SRR044113	SRS015247.SRX020577.V13	SRS015247.SRX020535.V35
SRS015281	SRR043805	SRR047403	SRS015281.SRX020554.V13	SRS015281.SRX020558.V35
SRS015389	SRR043661	SRR045688	SRS015389.SRX020546.V13	SRS015389.SRX020555.V35
SRS015518	SRR044146	SRR044243	SRS015518.SRX020539.V13	SRS015518.SRX020573.V35
SRS015578	SRR044021	SRR043983	SRS015578.SRX020577.V13	SRS015578.SRX020535.V35
SRS015599	SRR044148	SRR047411	SRS015599.SRX020539.V13	SRS015599.SRX020573.V35
SRS015663	SRR044161	SRR044257	SRS015663.SRX020539.V13	SRS015663.SRX020573.V35
SRS015724	SRR044132	SRR044209	SRS015724.SRX020539.V13	SRS015724.SRX020573.V35
SRS015782	SRR044139	SRR044216	SRS015782.SRX020539.V13	SRS015782.SRX020573.V35
SRS015911	SRR044585	SRR044585	SRS015911.SRX020510.V13	SRS015911.SRX020510.V35
SRS015960	SRR044087	SRR044119	SRS015960.SRX020577.V13	SRS015960.SRX020535.V35
SRS016018	SRR044156	SRR044251	SRS016018.SRX020539.V13	SRS016018.SRX020573.V35
SRS016056	SRR049357	SRR044252	SRS016056.SRX020536.V13	SRS016056.SRX020573.V35
SRS016095	SRR044317	SRR044395	SRS016095.SRX020516.V13	SRS016095.SRX020522.V35
SRS016203	SRR044343	SRR044423	SRS016203.SRX020516.V13	SRS016203.SRX020522.V35
SRS016335	SRR044344	SRR044424	SRS016335.SRX020516.V13	SRS016335.SRX020522.V35
SRS018559	SRR044352	SRR044432	SRS018559.SRX020516.V13	SRS018559.SRX020522.V35
SRS019267	SRR044360	SRR044439	SRS019267.SRX020516.V13	SRS019267.SRX020522.V35
SRS043804	SRR046050	SRR046050	SRS043804.SRX020531.V13	SRS043804.SRX020531.V35
SRS053335	SRR046056	SRR046056	SRS053335.SRX020531.V13	SRS053335.SRX020531.V35
SRS046349	SRR046051	SRR046051	SRS046349.SRX020531.V13	SRS046349.SRX020531.V35
SRS045414	SRR047697	SRR049611	SRS045414.SRX020563.V13	SRS045414.SRX020518.V35
SRS046382	SRR047262	SRR046402	SRS046382.SRX020540.V13	SRS046382.SRX020563.V35
SRS065466	SRR047056	SRR049589	SRS065466.SRX020532.V13	SRS065466.SRX020579.V35
SRS064321	SRR046501	SRR047340	SRS064321.SRX020519.V13	SRS064321.SRX020548.V35
SRS011011	F51YIRY01	F6J9Z3U02	700014390.V13.241827	700014390.V35.241827

SRS011033	715342091112.1.TCAG	715342091112.2.TCAG	700014489.V13.241766	700014489.V35.241766
SRS011034	F51YIRY01	F5672XE02	700014490.V13.241766	700014490.V35.241766
SRS011035	F51YIRY01	F6JVTJB02	700014503.V13.241762	700014503.V35.241762
SRS011036	715342091112.1.TCAG	715342091112.2.TCAG	700014504.V13.241762	700014504.V35.241762
SRS011060	F51YIRY01	F6J9Z3U02	700014561.V13.241810	700014561.V35.241810
SRS011061	715342091112.1.TCAG	715342091112.2.TCAG	700014562.V13.241810	700014562.V35.241810
SRS011084	715342091112.1.TCAG	715342091112.2.TCAG	700014724.V13.241722	700014724.V35.241722
SRS011134	720589091112.1.TCAG	720589091112.2.TCAG	700014837.V13.241820	700014837.V35.241820
SRS011136	720589091112.1.TCAG	720589091112.2.TCAG	700014845.V13.241773	700014845.V35.241773
SRS011205	F6J9Z3U01	F6RMMXF02	700015117.V13.241755	700015117.V35.241755
SRS011239	715342091112.1.TCAG	715342091112.2.TCAG	700015181.V13.241719	700015181.V35.241719
SRS011298	720589091112.1.TCAG	720589091112.2.TCAG	700015399.V13.241665	700015399.V35.241665
SRS011302	720589091112.1.TCAG	720589091112.2.TCAG	700015415.V13.241712	700015415.V35.241712
SRS011382	720589091112.1.TCAG	720589091112.2.TCAG	700015704.V13.241804	700015704.V35.241804
SRS011411	F5672XE01	F6J9Z3U02	700016013.V13.241709	700016013.V35.241709
SRS011451	F5672XE01	F6J9Z3U02	700016141.V13.241760	700016141.V35.241760
SRS011528	F57CATM01	F6JVTJB02	700016609.V13.241837	700016609.V35.241837
SRS011585	F57CATM01	F6JVTJB02	700033152.V13.241783	700033152.V35.241783
SRS011652	F57CATM01	F6JVTJB02	700033748.V13.241764	700033748.V35.241764

Table S7.

Percent of taxonomic information preserved at various taxonomic levels across primer sets. Taxonomic levels 1-5 preserve 93% of taxonomic information between HMP datasets using primers targeting the V1-V3 and V3-V5 rRNA regions.

Taxonomic Level	V1-V3	V3-V5
1	1	1
2	1	1
3	0.999988732	0.999994203
4	0.99996338	0.999994203
5	0.932405634	0.929756522
6	0.768960563	0.724304348
7	0.229014085	0.15522029

Table S8.

HMP metagenomic datasets used as a reference set. A list of all 32 samples and runs used in the analyses involving HMP metagenomic datasets.

Run	BioSample
SRR528006	SAMN00084837
SRR528372	SAMN00088016
SRR528373	SAMN00097954
SRR528374	SAMN00076282
SRR528377	SAMN00070035
SRR528379	SAMN00088832
SRR528381	SAMN00097897
SRR528384	SAMN00085508
SRR528385	SAMN00037656
SRR528386	SAMN00088409
SRR528392	SAMN00085245
SRR528393	SAMN00085131
SRR528396	SAMN00073658
SRR528399	SAMN00069411
SRR528402	SAMN00139730
SRR528404	SAMN00139862
SRR528405	SAMN00078079
SRR528407	SAMN00074361
SRR528410	SAMN00041840
SRR528417	SAMN00042371
SRR528418	SAMN00045881
SRR528420	SAMN00082513
SRR528422	SAMN00083997
SRR528425	SAMN00065748
SRR528431	SAMN00144714
SRR528433	SAMN00088586
SRR528473	SAMN00073582
SRR528481	SAMN00075542
SRR528487	SAMN00034187
SRR528504	SAMN00042437
SRR528525	SAMN00096542
SRR528369	SAMN00034067

Table S9.
CAZyme families related to the utilization of broad substrate categories.

Broad Substrate	CAZy Families
Plant Cell Wall Carbohydrates	GH1; GH2; GH3; GH4; GH5; GH8; GH9; GH11; GH12; GH15; GH16; GH17; GH26; GH27; GH28; GH29; GH36; GH39; GH43; GH44; GH48; GH51; GH53; GH55; GH67; GH74; GH78; GH93; GH94; GH95; GH115; GH117; GH121; PL1; PL2; PL6; PL7; PL9; PL11; PL15; PL22
Animal Carbohydrates	GH1; GH2; GH3; GH4; GH18; GH19; GH20; GH29; GH33; GH38; GH58; GH79; GH84; GH85; GH88; GH89; GH92; GH95; GH98; GH99; GH101; GH105; GH109; GH110; GH113; PL6; PL8; PL12; PL13; PL21
Sucrose/Fructans	GH32; GH68; GH70; GH91
Mucin	GH2; GH20; GH27; GH29; GH33; GH35; GH36; GH95; GH89; GH110; GH129

References and Notes

1. L. V. Hooper, D. R. Littman, A. J. Macpherson, Interactions between the microbiota and the immune system. *Science* **336**, 1268–1273 (2012). doi:10.1126/science.1223490 Medline

2. J. L. Sonnenburg, F. Bäckhed, Diet-microbiota interactions as moderators of human metabolism. *Nature* **535**, 56–64 (2016). doi:10.1038/nature18846 Medline

3. E. A. Mayer, R. Knight, S. K. Mazmanian, J. F. Cryan, K. Tillisch, Gut microbes and the brain: Paradigm shift in neuroscience. *J. Neurosci.* **34**, 15490–15496 (2014). doi:10.1523/JNEUROSCI.3299-14.2014 Medline

4. N. T. Mueller, E. Bakacs, J. Combellick, Z. Grigoryan, M. G. Dominguez-Bello, The infant microbiome development: Mom matters. *Trends Mol. Med.* **21**, 109–117 (2015). doi:10.1016/j.molmed.2014.12.002 Medline

5. J. K. Goodrich, E. R. Davenport, J. L. Waters, A. G. Clark, R. E. Ley, Cross-species comparisons of host genetic associations with the microbiome. *Science* **352**, 532–535 (2016). doi:10.1126/science.aad9379 Medline

6. A. H. Moeller, A. Caro-Quintero, D. Mjungu, A. V. Georgiev, E. V. Lonsdorf, M. N. Muller, A. E. Pusey, M. Peeters, B. H. Hahn, H. Ochman, Cospeciation of gut microbiota with hominids. *Science* **353**, 380–382 (2016). doi:10.1126/science.aaf3951 Medline

7. L. A. David, C. F. Maurice, R. N. Carmody, D. B. Gootenberg, J. E. Button, B. E. Wolfe, A. V. Ling, A. S. Devlin, Y. Varma, M. A. Fischbach, S. B. Biddinger, R. J. Dutton, P. J. Turnbaugh, Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014). doi:10.1038/nature12820 Medline

8. C. Lupp, M. L. Robertson, M. E. Wickham, I. Sekirov, O. L. Champion, E. C. Gaynor, B. B. Finlay, Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae. *Cell Host Microbe* **2**, 204 (2007). doi:10.1016/j.chom.2007.08.002 Medline

9. J. S. Lichtman, J. A. Ferreyra, K. M. Ng, S. A. Smits, J. L. Sonnenburg, J. E. Elias, Host-Microbiota Interactions in the Pathogenesis of Antibiotic-Associated Diseases. *Cell Reports* **14**, 1049–1061 (2016). doi:10.1016/j.celrep.2016.01.009 Medline

10. E. K. Costello, K. Stagaman, L. Dethlefsen, B. J. Bohannan, D. A. Relman, The application of ecological theory toward an understanding of the human microbiome. *Science* **336**, 1255–1262 (2012). doi:10.1126/science.1224203 Medline

11. L. Dethlefsen, D. A. Relman, Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl. Acad. Sci. U.S.A.* **108** (suppl. 1), 4554–4561 (2011). doi:10.1073/pnas.1000087107 Medline

12. E. D. Sonnenburg, S. A. Smits, M. Tikhonov, S. K. Higginbottom, N. S. Wingreen, J. L. Sonnenburg, Diet-induced extinctions in the gut microbiota compound over generations. *Nature* **529**, 212–215 (2016). doi:10.1038/nature16504 Medline

13. T. Yatsunenko, F. E. Rey, M. J. Manary, I. Trehan, M. G. Dominguez-Bello, M. Contreras, M. Magris, G. Hidalgo, R. N. Baldassano, A. P. Anokhin, A. C. Heath, B. Warner, J. Reeder, J. Kuczynski, J. G. Caporaso, C. A. Lozupone, C. Lauber, J. C. Clemente, D.

Knights, R. Knight, J. I. Gordon, Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012). Medline

14. C. De Filippo, D. Cavalieri, M. Di Paola, M. Ramazzotti, J. B. Poullet, S. Massart, S. Collini, G. Pieraccini, P. Lionetti, 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 (2010). doi:10.1073/pnas.1005963107 Medline

15. J. C. Clemente, E. C. Pehrsson, M. J. Blaser, K. Sandhu, Z. Gao, B. Wang, M. Magris, G. Hidalgo, M. Contreras, Ó. Noya-Alarcón, O. Lander, J. McDonald, M. Cox, J. Walter, P. L. Oh, J. F. Ruiz, S. Rodriguez, N. Shen, S. J. Song, J. Metcalf, R. Knight, G. Dantas, M. G. Dominguez-Bello, The microbiome of uncontacted Amerindians. *Sci. Adv.* **1**, e1500183 (2015). doi:10.1126/sciadv.1500183 Medline

16. A. J. Obregon-Tito, R. Y. Tito, J. Metcalf, K. Sankaranarayanan, J. C. Clemente, L. K. Ursell, Z. Zech Xu, W. Van Treuren, R. Knight, P. M. Gaffney, P. Spicer, P. Lawson, L. Marin-Reyes, O. Trujillo-Villarroel, M. Foster, E. Guija-Poma, L. Troncoso-Corzo, C. Warinner, A. T. Ozga, C. M. Lewis, Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat. Commun.* **6**, 6505 (2015). doi:10.1038/ncomms7505 Medline

17. I. Martínez, J. C. Stegen, M. X. Maldonado-Gómez, A. M. Eren, P. M. Siba, A. R. Greenhill, J. Walter, The gut microbiota of rural papua new guineans: Composition, diversity patterns, and ecological processes. *Cell Reports* **11**, 527–538 (2015). doi:10.1016/j.celrep.2015.03.049 Medline

18. T. A. Suzuki, M. Worobey, Geographical variation of human gut microbial composition. *Biol. Lett.* **10**, 20131037 (2014). doi:10.1098/rsbl.2013.1037 Medline

19. S. L. Schnorr, M. Candela, S. Rampelli, M. Centanni, C. Consolandi, G. Basaglia, S. Turroni, E. Biagi, C. Peano, M. Severgnini, J. Fiori, R. Gotti, G. De Bellis, D. Luiselli, P. Brigidi, A. Mabulla, F. Marlowe, A. G. Henry, A. N. Crittenden, Gut microbiome of the Hadza hunter-gatherers. *Nat. Commun.* **5**, 3654 (2014). doi:10.1038/ncomms4654 Medline

20. F. Marlowe, *The Hadza: Hunter-Gatherers of Tanzania*. Origins of human behavior and culture, 3 (Univ. of California Press, Berkeley, 2010).

21. N. G. Blurton Jones, L. C. Smith, J. F. O'Connell, K. Hawkes, C. L. Kamuzora, Demography of the Hadza, an increasing and high density population of Savanna foragers. *Am. J. Phys. Anthropol.* **89**, 159–181 (1992). doi:10.1002/ajpa.1330890204 Medline

22. E. R. Davenport, O. Mizrahi-Man, K. Michelini, L. B. Barreiro, C. Ober, Y. Gilad, Seasonal variation in human gut microbiome composition. *PLOS ONE* **9**, e90731 (2014). doi:10.1371/journal.pone.0090731 Medline

23. C. Huttenhower, D. Gevers, R. Knight, S. Abubucker, J. H. Badger, A. T. Chinwalla, H. H. Creasy, A. M. Earl, M. G. FitzGerald, R. S. Fulton, M. G. Giglio, K. Hallsworth-Pepin, E. A. Lobos, R. Madupu, V. Magrini, J. C. Martin, M. Mitreva, D. M. Muzny, E. J. Sodergren, J. Versalovic, A. M. Wollam, K. C. Worley, J. R. Wortman, S. K. Young, Q. Zeng, K. M. Aagaard, O. O. Abolude, E. Allen-Vercoe, E. J. Alm, L. Alvarado, G. L. Andersen, S. Anderson, E. Appelbaum, H. M. Arachchi, G. Armitage, C. A. Arze, T. Ayvaz, C. C. Baker, L. Begg, T. Belachew, V. Bhonagiri, M. Bihan, M. J. Blaser, T.

Bloom, V. Bonazzi, J. Paul Brooks, G. A. Buck, C. J. Buhay, D. A. Busam, J. L. Campbell, S. R. Canon, B. L. Cantarel, P. S. G. Chain, I.-M. A. Chen, L. Chen, S. Chhibba, K. Chu, D. M. Ciulla, J. C. Clemente, S. W. Clifton, S. Conlan, J. Crabtree, M. A. Cutting, N. J. Davidovics, C. C. Davis, T. Z. DeSantis, C. Deal, K. D. Delehaunty, F. E. Dewhirst, E. Deych, Y. Ding, D. J. Dooling, S. P. Dugan, W. Michael Dunne, A. Scott Durkin, R. C. Edgar, R. L. Erlich, C. N. Farmer, R. M. Farrell, K. Faust, M. Feldgarden, V. M. Felix, S. Fisher, A. A. Fodor, L. J. Forney, L. Foster, V. Di Francesco, J. Friedman, D. C. Friedrich, C. C. Fronick, L. L. Fulton, H. Gao, N. Garcia, G. Giannoukos, C. Giblin, M. Y. Giovanni, J. M. Goldberg, J. Goll, A. Gonzalez, A. Griggs, S. Gujja, S. Kinder Haake, B. J. Haas, H. A. Hamilton, E. L. Harris, T. A. Hepburn, B. Herter, D. E. Hoffmann, M. E. Holder, C. Howarth, K. H. Huang, S. M. Huse, J. Izard, J. K. Jansson, H. Jiang, C. Jordan, V. Joshi, J. A. Katancik, W. A. Keitel, S. T. Kelley, C. Kells, N. B. King, D. Knights, H. H. Kong, O. Koren, S. Koren, K. C. Kota, C. L. Kovar, N. C. Kyrpides, P. S. La Rosa, S. L. Lee, K. P. Lemon, N. Lennon, C. M. Lewis, L. Lewis, R. E. Ley, K. Li, K. Liolios, B. Liu, Y. Liu, C.-C. Lo, C. A. Lozupone, R. Dwayne Lunsford, T. Madden, A. A. Mahurkar, P. J. Mannon, E. R. Mardis, V. M. Markowitz, K. Mavromatis, J. M. McCorrison, D. McDonald, J. McEwen, A. L. McGuire, P. McInnes, T. Mehta, K. A. Mihindukulasuriya, J. R. Miller, P. J. Minx, I. Newsham, C. Nusbaum, M. O'Laughlin, J. Orvis, I. Pagani, K. Palaniappan, S. M. Patel, M. Pearson, J. Peterson, M. Podar, C. Pohl, K. S. Pollard, M. Pop, M. E. Priest, L. M. Proctor, X. Qin, J. Raes, J. Ravel, J. G. Reid, M. Rho, R. Rhodes, K. P. Riehle, M. C. Rivera, B. Rodriguez-Mueller, Y.-H. Rogers, M. C. Ross, C. Russ, R. K. Sanka, P. Sankar, J. Fah Sathirapongsasuti, J. A. Schloss, P. D. Schloss, T. M. Schmidt, M. Scholz, L. Schriml, A. M. Schubert, N. Segata, J. A. Segre, W. D. Shannon, R. R. Sharp, T. J. Sharpton, N. Shenoy, N. U. Sheth, G. A. Simone, I. Singh, C. S. Smillie, J. D. Sobel, D. D. Sommer, P. Spicer, G. G. Sutton, S. M. Sykes, D. G. Tabbaa, M. Thiagarajan, C. M. Tomlinson, M. Torralba, T. J. Treangen, R. M. Truty, T. A. Vishnivetskaya, J. Walker, L. Wang, Z. Wang, D. V. Ward, W. Warren, M. A. Watson, C. Wellington, K. A. Wetterstrand, J. R. White, K. Wilczek-Boney, Y. Q. Wu, K. M. Wylie, T. Wylie, C. Yandava, L. Ye, Y. Ye, S. Yooseph, B. P. Youmans, L. Zhang, Y. Zhou, Y. Zhu, L. Zoloth, J. D. Zucker, B. W. Birren, R. A. Gibbs, S. K. Highlander, B. A. Methé, K. E. Nelson, J. F. Petrosino, G. M. Weinstock, R. K. Wilson, O. White; Human Microbiome Project Consortium, Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012). doi:10.1038/nature11234 Medline

24. S. Rampelli, S. L. Schnorr, C. Consolandi, S. Turroni, M. Severgnini, C. Peano, P. Brigidi, A. N. Crittenden, A. G. Henry, M. Candela, Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota. *Curr. Biol.* **25**, 1682–1693 (2015). doi:10.1016/j.cub.2015.04.055 Medline

25. B. L. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, The Carbohydrate-Active EnZymes database (CAZy): An expert resource for Glycogenomics. *Nucleic Acids Res.* **37**, D233–D238 (2009). doi:10.1093/nar/gkn663 Medline

26. A. Marcobal, P. C. Kashyap, T. A. Nelson, P. A. Aronov, M. S. Donia, A. Spormann, M. A. Fischbach, J. L. Sonnenburg, A metabolomic view of how the human gut microbiota impacts the host metabolome using humanized and gnotobiotic mice. *ISME J.* **7**, 1933–1943 (2013). doi:10.1038/ismej.2013.89 Medline

27. P. Diaconis, S. Goel, S. Holmes, Horseshoes in multidimensional scaling and local kernel methods. *Ann. Appl. Stat.* **2**, 777–807 (2008). doi:10.1214/08-AOAS165

28. A. Gorvitovskaia, S. P. Holmes, S. M. Huse, Interpreting Prevotella and Bacteroides as biomarkers of diet and lifestyle. *Microbiome* **4**, 15 (2016). doi:10.1186/s40168-016-0160-7 Medline

29. J. L. Sonnenburg, J. Xu, D. D. Leip, C. H. Chen, B. P. Westover, J. Weatherford, J. D. Buhler, J. I. Gordon, Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* **307**, 1955–1959 (2005). doi:10.1126/science.1109051 Medline

30. E. C. Martens, H. C. Chiang, J. I. Gordon, Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**, 447–457 (2008). doi:10.1016/j.chom.2008.09.007 Medline

31. M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W. T. Liu, M. Crüsemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C. C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrewe, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C. C. Liaw, Y. L. Yang, H. U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. Boya, P, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P. M. Allard, P. Phapale, L. F. Nothias, T. Alexandrov, M. Litaudon, J. L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D. T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. O. Palsson, K. Pogliano, R. G. Linington, M. Gutiérrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein, N. Bandeira, Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016). doi:10.1038/nbt.3597 Medline

32. J. A. Gilbert, J. K. Jansson, R. Knight, The Earth Microbiome project: Successes and aspirations. *BMC Biol.* **12**, 69 (2014). doi:10.1186/s12915-014-0069-1 Medline

33. J. G. Caporaso, C. L. Lauber, W. A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S. M. Owens, J. Betley, L. Fraser, M. Bauer, N. Gormley, J. A. Gilbert, G. Smith, R. Knight, Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624 (2012). doi:10.1038/ismej.2012.8 Medline

34. J. G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Peña, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J.

Zaneveld, R. Knight, QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010). doi:10.1038/nmeth.f.303 Medline

35. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010). doi:10.1093/bioinformatics/btq461 Medline

36. T. Z. DeSantis, P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, G. L. Andersen, Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006). doi:10.1128/AEM.03006-05 Medline

37. C. Lozupone, R. Knight, UniFrac: A new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**, 8228–8235 (2005). doi:10.1128/AEM.71.12.8228-8235.2005 Medline

38. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009). doi:10.1093/molbev/msp077 Medline

39. F. Pedregosa *et al*., Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

40. C. A. Smith, E. J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **78**, 779–787 (2006). doi:10.1021/ac051437y Medline

41. R. Schmieder, R. Edwards, Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLOS ONE* **6**, e17288 (2011). doi:10.1371/journal.pone.0017288 Medline

42. M. P. Cox, D. A. Peterson, P. J. Biggs, SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* **11**, 485 (2010). doi:10.1186/1471-2105-11-485 Medline

43. M. Rho, H. Tang, Y. Ye, FragGeneScan: Predicting genes in short and error-prone reads. *Nucleic Acids Res.* **38**, e191 (2010). doi:10.1093/nar/gkq747 Medline

44. S. Abubucker, N. Segata, J. Goll, A. M. Schubert, J. Izard, B. L. Cantarel, B. Rodriguez-Mueller, J. Zucker, M. Thiagarajan, B. Henrissat, O. White, S. T. Kelley, B. Methé, P. D. Schloss, D. Gevers, M. Mitreva, C. Huttenhower, Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLOS Comput. Biol.* **8**, e1002358 (2012). doi:10.1371/journal.pcbi.1002358 Medline

45. S. R. Eddy, A new generation of homology search tools based on probabilistic inference. *Genome Inform.* **23**, 205–211 (2009). Medline

46. Y. Yin, X. Mao, J. Yang, X. Chen, F. Mao, Y. Xu, dbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**, W445–451 (2012). doi:10.1093/nar/gks479 Medline

47. M. K. Gibson, K. J. Forsberg, G. Dantas, Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* **9**, 207–216 (2015). doi:10.1038/ismej.2014.106 Medline

48. B. L. Cantarel, V. Lombard, B. Henrissat, Complex carbohydrate utilization by the healthy human microbiome. *PLOS ONE* **7**, e28742 (2012). doi:10.1371/journal.pone.0028742 Medline

49. F. J. Bennett, N. A. Barnicot, J. C. Woodburn, M. S. Pereira, B. E. Henderson, Studies on viral, bacterial, rickettsial and treponemal diseases in the Hadza of Tanzania and a note on injuries. *Hum. Biol.* **45**, 243–272 (1973). Medline

50. F. L. Black, W. J. Hierholzer, F. Pinheiro, A. S. Evans, J. P. Woodall, E. M. Opton, J. E. Emmons, B. S. West, G. Edsall, W. G. Downs, G. D. Wallace, Evidence for persistence of infectious agents in isolated human populations. *Am. J. Epidemiol.* **100**, 230–250 (1974). doi:10.1093/oxfordjournals.aje.a112032 Medline

51. I. L. Brito, S. Yilmaz, K. Huang, L. Xu, S. D. Jupiter, A. P. Jenkins, W. Naisilisili, M. Tamminen, C. S. Smillie, J. R. Wortman, B. W. Birren, R. J. Xavier, P. C. Blainey, A. K. Singh, D. Gevers, E. J. Alm, Mobile genes in the human microbiome are structured from global to individual scales. *Nature* **535**, 435–439 (2016). doi:10.1038/nature18927 Medline

52. M. J. Claesson, I. B. Jeffery, S. Conde, S. E. Power, E. M. O'Connor, S. Cusack, H. M. Harris, M. Coakley, B. Lakshminarayanan, O. O'Sullivan, G. F. Fitzgerald, J. Deane, M. O'Connor, N. Harnedy, K. O'Connor, D. O'Mahony, D. van Sinderen, M. Wallace, L. Brennan, C. Stanton, J. R. Marchesi, A. P. Fitzgerald, F. Shanahan, C. Hill, R. P. Ross, P. W. O'Toole, Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, 178–184 (2012). Medline

53. E. R. Morton, J. Lynch, A. Froment, S. Lafosse, E. Heyer, M. Przeworski, R. Blekhman, L. Ségurel, Variation in Rural African Gut Microbiota Is Strongly Correlated with Colonization by Entamoeba and Subsistence. *PLOS Genet.* **11**, e1005658 (2015). doi:10.1371/journal.pgen.1005658 Medline

54. J. Zhang, Z. Guo, Z. Xue, Z. Sun, M. Zhang, L. Wang, G. Wang, F. Wang, J. Xu, H. Cao, H. Xu, Q. Lv, Z. Zhong, Y. Chen, S. Qimuge, B. Menghe, Y. Zheng, L. Zhao, W. Chen, H. Zhang, A phylo-functional core of gut microbiota in healthy young Chinese cohorts across lifestyles, geography and ethnicities. *ISME J.* **9**, 1979–1990 (2015). doi:10.1038/ismej.2015.11 Medline

55. J. Nakayama, K. Watanabe, J. Jiang, K. Matsuda, S. H. Chao, P. Haryono, O. La-Ongkham, M. A. Sarwoko, I. N. Sujaya, L. Zhao, K. T. Chen, Y. P. Chen, H. H. Chiu, T. Hidaka, N. X. Huang, C. Kiyohara, T. Kurakawa, N. Sakamoto, K. Sonomoto, K. Tashiro, H. Tsuji, M. J. Chen, V. Leelavatcharamas, C. C. Liao, S. Nitisinprasert, E. S. Rahayu, F. Z. Ren, Y. C. Tsai, Y. K. Lee, Diversity in gut bacterial community of school-age children in Asia. *Sci. Rep.* **5**, 8397 (2015). doi:10.1038/srep08397 Medline