Supporting Information

Goodman et al. 10.1073/pnas.1102938108

SI Materials and Methods

Culturing of Fecal Microbiota. Samples were placed in prereduced PBS with 0.1% cysteine (PBS_C) 15 mL g^{-1} feces. The fecal material was suspended by vortexing for 5 min, and the suspension was allowed to stand at room temperature for 5 min to permit large insoluble particles to settle to the bottom of the tube. 16S ribosomal RNA (rRNA) sequencing of the starting material and of the supernatant obtained after the settling step showed no significant differences in community composition. This settling step dramatically increased the reproducibility of subsequent dilutions. Diluted (10^{-4}) samples were plated on plates 150 mm in diameter containing prereduced, nonselective Gut Microbiota Medium (GMM) (Table S2) so that colonies were dense but distinct (~5,000 colonies per plate) after a 7-d incubation at 37 °C under an atmosphere of 75% N_2, 20% CO_2, and 5% H_2. GMM is modified from supplemented trypone-yeast-glucose (TYG_S) medium (1) and contains only commercially available components. To discourage colony overgrowth, the concentration of glucose, tryptone, and yeast extract is reduced fivefold compared with TYG_S. Colonies were harvested *en masse* from each of six plates by scraping with a cell scraper (BD Falcon) into 10 mL of prereduced PBS_C. Stocks were generated by adding prereduced glycerol containing 0.1% cysteine to the fecal or cultured samples (final concentration of glycerol, 20%). Stocks were stored in anaerobic glass vials in a standard −80 °C freezer.

To determine the optimal number of plates to be surveyed for each fecal sample, a freshly discarded sample from one of the anonymous human donors was processed as above, and the 10^{-4} dilution was plated on 10 prereduced GMM plates. De-noised, chimera-checked variable region 2 (V2)-directed 16S rRNA reads generated from the pooled colonies obtained from each plate separately after a 7-d incubation were assigned to operational taxonomic units (OTUs) at 97% nucleotide sequence identity (ID) using QIIME v1.1 (2). Each additional plate contributed new OTUs, although, as shown by the rarefaction curves plotted in Fig. S8*A*, the contribution of each added plate fell to <20 new OTUs after approximately six plates. For this reason, each subsequent cultured sample reflects pooled scraped material from six plates.

Gnotobiotic Mouse Husbandry. Mice were housed under a strict 12-h light/dark cycle and fed a standard, autoclaved low-fat/plant polysaccharide-rich (LF/PP) chow diet (B&K Universal) ad libitum. Mice were colonized by gavage (0.2 mL of the resuspended fecal material or pooled cultured organisms recovered from GMM after 7-d incubation as above, per germfree recipient). Animals receiving different microbial inoculations were placed in separate gnotobiotic isolators before gavage; once gavaged, all mice were caged individually.

After 4 wk of acclimatization on the LF/PP diet, mice were transitioned to a Western diet (Harlan-Teklad TD96132) provided ad libitum for 2 wk and then were returned to the LF/PP diet for 2 wk. 16S rRNA analysis of fecal samples collected 1, 4, 7, and 14 d after gavage indicated that both complete and cultured communities reached a steady state well before the diet transition. During the initial LF/PP diet phase, fecal samples were collected at postgavage days 4, 7, 14, and either on day 32 when the input community was a complete microbiota or on day 25 in the case of an input cultured community. Mice were sampled on days 1, 3, 7, and 14 after the shift to the Western diet and on days 1, 3, 8, and 15 upon return to LF/PP chow. The animals then were fasted for 24 h and returned to the LF/PP diet for 1 wk before they were killed.

Fecal samples for subsequent culture were collected from each mouse directly into BBL thioglycollate medium (BD), transported to an anaerobic chamber within 30 min, then diluted and plated on GMM as above. Fecal samples from fasted mice were not cultured because 16S rRNA analysis did not show significant changes in community composition at either the 12-h or 24-h fasting time points. After mice were sacrificed, the intestine was subdivided into 16 segments of equivalent length numbered from 1 (proximal) to 16 (distal). Contents from small intestine segments 2, 5, and 13, plus cecum and colon contents, were snap-frozen in liquid nitrogen and stored at −80 °C.

DNA Extraction and Purification. Fecal samples (0.1 g) were resuspended into 710 μL of 200 mM NaCl, 200 mM Tris, 20 mM EDTA plus 6% SDS (Buffer A). After the addition of 0.5 mL of 0.1-mm zirconia/silica beads (BioSpec Products) and 0.5 mL of phenol/chloroform/isoamyl alcohol, pH 7.9 (Ambion), cells were lysed by mechanical disruption with a bead-beater (BioSpec Products) for 3 min. Samples were centrifuged for 3 min at 6,800 × *g*, and the aqueous phase was collected and subjected to a second phenol-chloroform-isoamyl alcohol extraction using Phase Lock Gel tubes (5 Prime). DNA in the aqueous phase was precipitated by the addition of an equal volume of isopropanol and 0.1 volumes of 3 M sodium acetate, pH 5.5 (Ambion). After overnight incubation at −20 °C, samples were centrifuged for 20 min at 4 °C at 18,000 × *g*, and the supernatant was removed. Pelleted DNA was washed once with 0.5 mL of 100% ethanol, and dried using a vacuum evaporator. DNA pellets then were resuspended in 0.2 mL of Tris-EDTA containing 4 μg RNAseA (Qiagen). Crude DNA extracts were column-purified using the Rapid PCR Purification Kit (Marligen Biosciences). DNA concentrations were adjusted to 50 ng/μL for subsequent 16S rRNA or shotgun pyrosequencing.

16S rRNA Sequencing. PCR reactions were carried out in triplicate using 2.5× Master Mix (5 Prime), forward primer (FLX-8F; 5′-GCCTTGCCAGCCCGCTCAGTC*AGAGTTTGATCCTGGC-TCAG*-3′; the 454 FLX Amplicon primer B sequence is underlined, and the 16S rRNA primer sequence 8F is italicized), barcoded (3) reverse primer (FLX-BC-338R; 5′-GCCTCCCT-CGCGCCATCAGNNNNNNNNNNNCA*TGCTGCCTCCCG-TAGGAGT*-3′; the 454 FLX Amplicon primer A sequence is underlined, "N" indicates the barcode sequence, and the 16S rRNA primer sequence 338R is shown in italics), and 50 ng input DNA purified as described above. Reactions were incubated for 2 min at 95 °C, followed by 30 cycles of 20 s at 95 °C, 20 s at 52 °C, and 1 min at 65 °C. Triplicate reactions were pooled, inspected by gel electrophoresis, and purified on AMPure beads (Agencourt Biosciences). For each barcoded primer, negative control reactions lacking input DNA were conducted in parallel. Purified samples were combined at equimolar concentrations and sequenced with FLX chemistry on a 454 pyrosequencer (Roche).

16S rRNA Sequence Analysis. For 16S rRNA sequence analysis, sequences were preprocessed to remove reads with low-quality scores (sliding window set to 50 bp), ambiguous characters, and incorrect lengths (<200 or >300 bp). Reads passing these criteria were assigned to specific samples based on their error-corrected barcode sequence, de-noised using default parameters, grouped into OTUs at 97%ID, and a representative sequence was se-

lected from each OTU using default parameters in QIIME v.1.1. These representative sequences then were filtered for possible chimeric sequences using ChimeraSlayer (http://microbiomeutil.sourceforge.net) with default parameters (sequences designated "unknown" were not discarded). Filtered datasets were subsampled to 1,000 sequences per sample (the only exceptions were the datasets from the human-derived complete and cultured samples collected at the day 148 time point shown in Fig. 1*A*, which were subsampled to 5,000 sequences). Beta-diversity calculations were conducted using Jaccard (nonphylogenetic) and UniFrac (phylogenetic) metrics in QIIME v1.1.

Weighted Taxonomic Analysis. Analysis of full-length and V2-region delimited sequences from the 16S rRNA genes of taxonomically defined bacteria indicated that discrete %ID cutoffs do not correspond closely to established taxonomic levels. Histograms of the distributions of %ID values of 16S rRNA sequences between representatives of two species in the same genus, two genera in the same family, and so forth, overlap to a large extent (Fig. S8*B* shows comparisons of 16S V2 regions, which are used commonly in multiplex pyrosequencing studies, between 4,041 bacterial species selected from the SILVA database (4) v102). For this reason, we primarily used a taxonomic assignment method to compare uncultured and cultured communities across taxonomic levels rather than approximating taxonomic groups by selecting arbitrary %ID cutoffs to represent each taxonomic level. For example, Fig. S8*B* illustrates that there is no clear %ID cutoff that distinguishes species-level from genus-level groups or family-level from order-level groups. As an alternate, taxonomy-independent approach, we grouped de-noised 16S rRNA sequences into OTUs at a range of %ID cutoffs (80%, 90%, 95%, 97%) using uclust (5) in QIIME v1.1. At each %ID cutoff, OTUs were filtered for chimeras as above. The proportion of reads in the uncultured sample that belonged to OTUs also identified in the corresponding cultured sample were determined as in Fig. 1*A* (see Fig. S1 *A–F* for a comparison of results obtained from different assignment methods).

SILVA-VOTE: A Computational Pipeline for Improved Accuracy in Taxonomic Assignments of V2 16S rRNA Sequences. We observed that commonly used tools for taxonomy assignment often failed to correctly assign V2 16S rRNA sequences derived from known human gut microbes. To generate a nonredundant, curated 16S rRNA database for taxonomy assignment, we downloaded the v102 SILVA database (4) prefiltered for redundancy at a 99%ID (SSURef_102_SILVA_NR_99.fasta; http://www.arb-silva.de). This database is composed of 262,092 full-length sequences from the small subunit rRNAs of Eukaryotes, Bacteria, and Archaea. A total of 297 sequences whose accession numbers had been removed from or modified by GenBank or were not associated with a complete National Center for Biotechnology Information (NCBI) taxonomy (i.e., phylum, class, order, family, genus, and species designations) were excluded. The remaining sequences were aligned using PyNast as implemented in QIIME v1.1: 224,899 sequences were aligned successfully and contained more than 90% of the V2 region. These V2 sequences were filtered for redundancy by clustering and selection of a representative sequence from each cluster, using uclust at a 99% identity (2). To assign consensus taxonomies to the representative sequences, we applied a 75% majority voting scheme: For each taxonomic level, the representative sequence was assigned a taxonomic designation if more than 75% of the sequences within the cluster shared the same assignment; otherwise, the cluster was labeled "unknown" at that taxonomic level. Taxonomic designations of sequences within a cluster that included the nonunique identifiers "unknown," "uncultured," "candidatus," or "bacterium" were not considered in the 75% majority vote for taxonomy assignment of the representative sequence. Species-level annotations with numbers or decimal points (which in almost all cases refer to strains rather than species) also were excluded. After removal of sequence clusters with little or no consensus taxonomy (i.e., with 50% or more of the taxonomic levels labeled "unknown" after the voting analysis), 34,181 nonredundant, annotated, bacterial 16S rRNA V2 sequences remained and were designated as our reference database.

To assign taxonomy to the 16S rRNA V2-amplicon pyrosequencing reads, significant matches to the 34,181-sequence reference database were identified by BLAST (the top 100 hits with an e-value cutoff of $\leq 10^{-30}$ were retained). All the BLAST hits with a score within 10% of the score of the best BLAST hit were considered for the taxonomy assignment. Taxonomy was assigned for each phylogenetic level independently by using a majority voting scheme: A read was assigned a taxonomic designation if 50% or more of the selected reference sequences (whose BLAST scores were within 10% of the top score for that query sequence) shared the same taxonomic assignment. As above, sequences designated "unknown" were not taken into account for the voting. When no assignment was conserved in >50% of the selected BLAST hits, the query sequence was labeled as "nonidentified" at that taxonomic level.

To test this method, 16S rRNA V2 sequences were extracted from the genomes of 66 human gut microbes (Table S11). Taxonomy was assigned to these test sequences in QIIME v.1.1 using three methods: (*i*) the Ribosomal Database Project Bayesian classifier (v.2.0) (6); (*ii*) a BLAST top hit-based query of the Greengenes core sequence set [composed of 4,938 sequences; downloaded September 15, 2009; (http://greengenes.lbl.gov/Download/Sequence_Data/Fasta_data_files/core_set_aligned.fasta.imputed) (7)]; and (*iii*) with SILVA-VOTE. Comparison of these results suggests that SILVA-VOTE yields a significantly increased number of correct taxonomic assignments, particularly at the genus and species levels (Fig. S8*C*).

Control Experiments to Address the Influence of Lysed or Nongrowing Cells on 16S rRNA Datasets from Colonies Collected from Agar Plates. Although each average-sized colony among the ~30,000 colonies obtained from each cultured sample likely contributed ~10^9 cells to the pooled population (8), in theory genetic material from lysed or nongrowing cells could also contribute to the sequences obtained from the plated samples. To test this possibility directly, fecal samples were diluted 10,000-fold and plated onto GMM and also onto plates containing ingredients that should not support growth of bacteria and thus represent the background expected if 100% of the plated material was nongrowing or lysed [control PARC plates contained Phosphate buffer, noble Agar, Resazurin (oxygen indicator), and Cysteine (reducing agent); Table S9]. After a 7-d anaerobic incubation, no colonies were observed on the PARC plates. Twenty randomly selected single colonies from the GMM plates were picked, an aliquot was reserved for 16S rRNA gene sequencing, and the remainder was pooled with the scraped surfaces of the PARC plates. Sequencing this pool revealed that >98% of the 16S rRNA reads could be attributed to the 20 colonies from the GMM plates; among the remainder, none belonged to OTUs represented by more than two reads per 1,000. Together, these findings suggest that at least 98% of the reads generated from 30,000 pooled colonies are not derived from nongrowing or lysed bacteria.

Testing the Possible Contribution of Lysed or Nongrowing Cells to Microbial Communities in Gnotobiotic Mice Gavaged with a Readily Cultured Human Gut Microbiota. As noted in the main text, the initial cultured inoculum was prepared by scraping GMM plates *en masse*. In theory, this input could include cells that did not actually grow in these conditions but instead remained dormant, below the limit of detection by 16S rRNA sequencing, over the 7-d in vitro incubation period. To determine whether such non-

growing taxa contributed to the distal gut communities of mice that received a readily cultured microbiota, a control sample containing material harvested from PARC plates and pooled with 20 visible colonies picked directly from GMM plates was introduced into five age-matched, individually caged, germfree mice fed a LF/PP diet. Fecal samples were collected at 3, 7, and 14 d postgavage and were subjected to V2-targeted 16S rRNA pyrosequencing. Community composition, as determined both by alpha-diversity and beta-diversity metrics, was stable after the 7-d time point (average UniFrac distance within 7-d or 14-d samples = 0.319; average distance between 7-d and 14-d samples = 0.321; $P > 0.89$ based on unpaired, two-tailed Student's t test; Fig. S8 *D–E*). When 16S rRNA sequences obtained from the 14-d fecal samples were compared with sequences obtained from the 20 picked colonies, we found that the mice harbored only two OTUs, both mapping to *Akkermansia muciniphila*, that could not be attributed to the 20 colonies and that were not observed on GMM plates in any other experiments. *Akkermansia muciniphila* type strain ATCC BAA-835 contains three 16S rRNA genes and grows readily on GMM. This species was a minor component in the fecal microbiota of the two donors (one or fewer reads per 1,000 from eight samples collected over time); in fecal communities sampled from mice that received the readily cultured component of either donor's microbiota, abundance averaged 1.8% across all time points. These findings indicate that just 20 actively growing colonies are able to exclude virtually all non-growing species that may be present on GMM plates from colonizing germfree mice.

Shotgun Pyrosequencing. Five hundred nanogram aliquots of DNA prepared from selected complete and cultured microbiota were sheared and ligated to the default 454 Titanium multiplex identifiers (MIDs; Roche Rapid Library Preparation Method Manual, GS FLX Titanium Series, October 2009). 16S rRNA sequencing of samples from individually caged mice colonized with the same community indicated a high degree of similarity between individual animals. For this reason, fecal DNAs from replicate mice were pooled (n = 3–5 mice per pool), and the 12 pooled samples, each labeled with a unique MID, were sequenced in a single 454 Titanium run.

Shotgun pyrosequencing reads were parsed by MID and filtered to remove short sequences (<60 bp), low-quality sequences (three or more N bases in the sequence or two continuous N bases), and replicate sequences (>97%ID over the length of the read, with identical sequences over the first 20 bases) (9). Reads reflecting host DNA contamination were identified by BLAST (against the mouse genome for samples isolated from mice and against the human genome for all other samples) and were removed *in silico* (≥75% identify, E-value ≤ 10^{-5}, bitscore ≥ 50). Remaining sequences were queried against the KEGG Orthology (KO) database (v52) with a Blastx e-value cutoff of 10^{-5}. KO assignments were mapped further to enzyme commission (EC) and KEGG pathway annotations.

Bio-Prospecting for Antibiotic-Resistance Genes in Uncultured and Readily Cultured Microbiota. DNA fragments from complete and cultured communities were cloned into an expression vector, electroporated into *Escherichia coli*, and screened for their ability to confer resistance to 15 different antibiotics. To this end, 10 μg of DNA purified from the two human donors' fecal microbiota, from the derived cultured communities (samples with 16S rRNA samples 2, 3, 5, and 6 in Tables S10 and S14), plus pooled contents of the arrayed strain collection (Table S8) were sheared to 1.5- to 4-kB fragments (Diagenode Bioruptor XL), followed by size selection (1% agarose gel electrophoresis). Sheared DNA then was end-repaired (Epicentre EndIt Kit), column-purified (Qiagen QiaQuick PCR Purification Kit), and concentrated in a vacuum evaporator (Thermo Fisher SpeedVac). The expression vector pZE21-MCS1 (10) was subjected to PCR amplification using primers flanking the HincII site (pZE21_126_146FOR, 5′-GACGGTATCGATAAGCTTGAT- 3′; pZE21_111_123rcREV, 5′-GACCTCGAGGGGGGG-3′). The resulting linear product was gel-purified, dephosphorylated with calf intestinal phosphatase, and column purified. Approximately 500 ng of the DNA fragments were ligated to 100 ng of the linearized vector in an overnight reaction (Epicentre FastLink Ligation Kit). The ligation reaction was desalted by dialysis in double-distilled H_2O and electroporated into *E. coli* MegaX DH10B T1R cells (Invitrogen). After 1 h of recovery, a small (1 μL) aliquot of the library was titered with serial dilutions onto LB agar plates containing 50 μg/mL kanamycin (to select for pZE21 transformants) and incubated at 37 °C for 16 h. The insert size distribution for each library was characterized by gel electrophoresis of amplicons obtained using primers flanking the HincII site in the multiple cloning site of pZE21 MCS1. The total size of each library was estimated by multiplying average insert size by the number of cfu in a given library. The remainder of the recovered cells was inoculated into 10 mL LB containing 50 μg/mL kanamycin and grown, with shaking, at 27 °C for 16 h. The culture subsequently was diluted with an equal volume of LB medium containing 30% glycerol and stored at −80 °C before screening.

For functional selections, 100 μL of each library freezer stock (corresponding to 0.5–1 × 10^8 cfu) was plated on an LB agar plate containing kanamycin (50 μg/mL) plus one of 15 different antibiotics (Table S6). The total number of cells plated on each antibiotic represented ∼10 copies of each original unique transformant. Antibiotic-resistant colonies were scored after plates had been incubated at 37 °C for 16 h. Inserts contained in colonies with amikacin-, piperacillin- and piperacillin/tazobactam-resistant phenotypes were subjected to bidirectional Sanger sequencing (Beckman Coulter Genomics) using primers pZE21_81_104_57C (5′-GAATTCATTAAAGAGGAGAAA GGT-3′) and pZE21_151_174rc_58C (5′-TTTCGTTTTATTTGATGCCTCTAG-3′). Resulting reads were trimmed to remove low-quality and vector sequences and subjected to within-library contig assembly (≥200 bp of 97%ID sequence required). Contigs and unassembled reads were mapped by BLAST to the National Center for Biotechnology Information (NCBI) nonredundant database and to a custom database of 122 human gut microbial genomes. All sequence datasets have been deposited in the NCBI Sequence Read Archive (SRA) under accession no. SRA026271.

Amikacin-resistant strains were quantified from each donor, in triplicate, by plating diluted fecal samples on GMM with and without amikacin (4,100 μg/mL; lower concentrations produced high background). Amikacin-resistant colonies were quantified after 5 d incubation under anaerobic conditions, and colony counts were normalized to the total number of colonies obtained in the absence of the antibiotic. A total of 48 fecal isolates (12 from the GMM+amikacin selection and 12 from the non-selective plates, from each of two donors) were chosen for a PCR-based survey for the amikacin-resistance genes captured in the *E. coli* libraries described above and for 16S rRNA sequencing.

Preparation of an Arrayed Species Collection. A single vial of the −80 °C anaerobic glycerol stock containing an aliquot of a fecal sample from Donor 2 was diluted into prereduced TYG_s medium lacking resazurin (1) in an anaerobic chamber and was dispensed into prereduced 384-well flat-bottomed trays (170 μL per well). To determine the dilution at which a high percentage of wells received a single viable cell in the initial inoculation (Fig. S8*F*), two- and fourfold serial dilutions were performed (from 10^{-6} to 10^{-10}) in a trial inoculation (48 wells per dilution; 170 μL per well). Trays were sealed with sterile foil lids and incubated anaerobically at 37 °C for 5 d. The dilution at which ∼30% of wells were turbid (OD_{630} > 0.2) was chosen for the

subsequent large-scale culturing (Fig. S8*G*). To this end, a second vial of the frozen anaerobic glycerol stock from the same donor was added to 500 mL of prereduced TYG_S medium lacking resazurin at the calculated dilution and dispensed into ten 384-well culture trays (170 μL per well). Trays were sealed and incubated as above. Cells were then resuspended in each well of each tray by pipetting, and 25 μL aliquots were transferred to each of two archive trays containing 25 μL prereduced TYG_S (resazurin included) plus 40% glycerol per well. The arrayed archive trays were sealed with aluminum foil, frozen on dry ice inside the anaerobic chamber, and transported on dry ice to a conventional −80 °C freezer for storage. Cultures stored in this fashion remain anaerobic, as judged colorimetrically using resazurin in the medium and by recovery of strict anaerobes (as long as they are transported frozen, on dry ice, into an anaerobic chamber for strain recovery). Another 50 μL aliquot from the culture trays (not from the archive trays) was measured by OD_{630} and stored at −80 °C for PCR amplification.

Taxonomies were assigned to each strain in the 3,840-well collection by two-step barcoded 454 FLX pyrosequencing. A 1 μL aliquot from each well was transferred to a new tray, and cells were lysed in 10 μL of lysis buffer (25 mM NaOH, 0.2 mM EDTA; incubation for 30 min at 95 °C) followed by the addition of 10 μL of neutralization buffer (40 mM Tris-HCl). To reduce the amplification of background DNA present from dead or lysed cells, the neutralized lysate was diluted 1:10 into EB buffer (Qiagen). To barcode each bacterial strain uniquely before amplicon sequencing, the V2 region of their 16S rRNA gene was targeted for PCR using 2x Phusion HF Master Mix (Finnzymes), 2 μL of input DNA, primer 454_16S_8F and 1 of 96 barcoded (Roche Multiplex Identifiers) reverse primers (454_16S_338R_barcode1) that include a 12-bp tail sequence (Table S12) in a 10-μL reaction (384-well format). Duplicate reactions were incubated for 30 s at 98 °C, followed by 30 cycles of 10 s at 98 °C, 30 s at 61 °C, and 30 s at 72 °C.

Reactions were combined so that each pool contained one representative associated with each barcode (four pools per 384-well tray), passed over a PCR cleanup column (Qiagen), and diluted to 0.5 ng/μL. Pools then were subjected to a second round of PCR amplification with 0.5 ng of pool DNA in a 25-μL reaction. Reactions were incubated for 30 s min at 98 °C, followed by five cycles of 10 s at 98 °C, 30 s at 54 °C, and 30 s at 72 °C, followed by an additional 25 cycles of 10 s at 98 °C and 30 s at 70 °C. In this second PCR, the reverse primers were replaced with a second barcoded (3) linker primer (454_linker_barcode2; Table S13) specific to the 12-bp tail sequence added in the first PCR. In this way, the 16S rRNA V2 regions of the bacterial genomes in each initial well were associated with a unique two-barcode 'pointer' sequence (Fig. 4*A* and legend). After the second-round PCR, reactions were pooled and run over a PCR cleanup column (Qiagen), and DNA in the expected size range (200–300 bp) was gel purified (Qiagen). The final product was quantified and subjected to multiplex 454 FLX pyrosequencing at a depth expected to yield 250 sequences per well (25% of one sequencing run).

The resulting reads were assigned to wells in the archive trays based on their associated barcodes (Table S8). Of the 3,840 wells, 1,181 (30.8%) were turbid as defined by $OD_{630} \geq 0.2$; of the turbid wells, 1,172 (99.2%) had at least a single sequence with the correct barcode combination. Barcode combinations mapping to wells with culture $OD_{630} < 0.2$ were also identified in the sequencing dataset. However, the total number of reads that mapped to these wells was much lower than to turbid wells (51,925 turbid versus 14,427 nonturbid), and the percentage of mapped wells was much lower for the nonturbid subset (62.2% vs. 99.2%).

The taxonomy of the most abundant sequence associated with each barcode combination was assigned using SILVA-VOTE. These most abundant sequences were also clustered into 97%ID OTUs using uclust in QIIME 1.1. To evaluate the diversity captured at varying taxonomic levels, representative 97%ID OTU sequences were assigned taxonomy using SILVA-VOTE. Reads designated "nonidentified" by SILVA-VOTE were not considered to represent an additional taxonomic group unless they were associated with a distinct higher-order taxonomic classification (e.g., sequences annotated as "Family Clostridiaceae; Genus nonidentified" were scored as representing a different genus-level group than sequences annotated as "Family Ruminococcaceae; Genus nonidentified"). Rarefaction analysis of the number of additional taxa added with each additional 384-well culture tray is shown in Fig. S8*H*. Mapping was verified by recovery of strains from the archive trays, colony purification, and full-length Sanger sequencing of their 16S rRNA gene.

Identification of Human Gut Isolates in the German Resource Centre for Biological Material Culture Collection. The German Resource Centre for Biological Material (DSMZ) bacterial culture collection (http://www.dsmz.de/microorganisms/bacteria_catalogue.php; October 14, 2010) was searched under the terms "gut," "faeces," "feces," "fecal," and "stool." Search results were filtered to exclude strains from nonhuman sources. Strains that matched the search terms without host species information were included, as were noncommensals (i.e., pathogens).

1. Goodman AL, et al. (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* 6:279–289.
2. Caporaso JG, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336.
3. Fierer N, Hamady M, Lauber CL, Knight R (2008) The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proc Natl Acad Sci USA* 105: 17994–17999.
4. Pruesse E, et al. (2007) SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 35: 7188–7196.
5. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461.
6. Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73:5261–5267.
7. DeSantis TZ, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72:5069–5072.
8. Mashimo K, Nagata Y, Kawata M, Iwasaki H, Yamamoto K (2004) Role of the RuvAB protein in avoiding spontaneous formation of deletion mutations in the Escherichia coli K-12 endogenous tonB gene. *Biochem Biophys Res Commun* 323:197–203.
9. Gomez-Alvarez V, Teal TK, Schmidt TM (2009) Systematic artifacts in metagenomes from complex microbial communities. *ISME J* 3:1314–1317.
10. Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* 25:1203–1210.

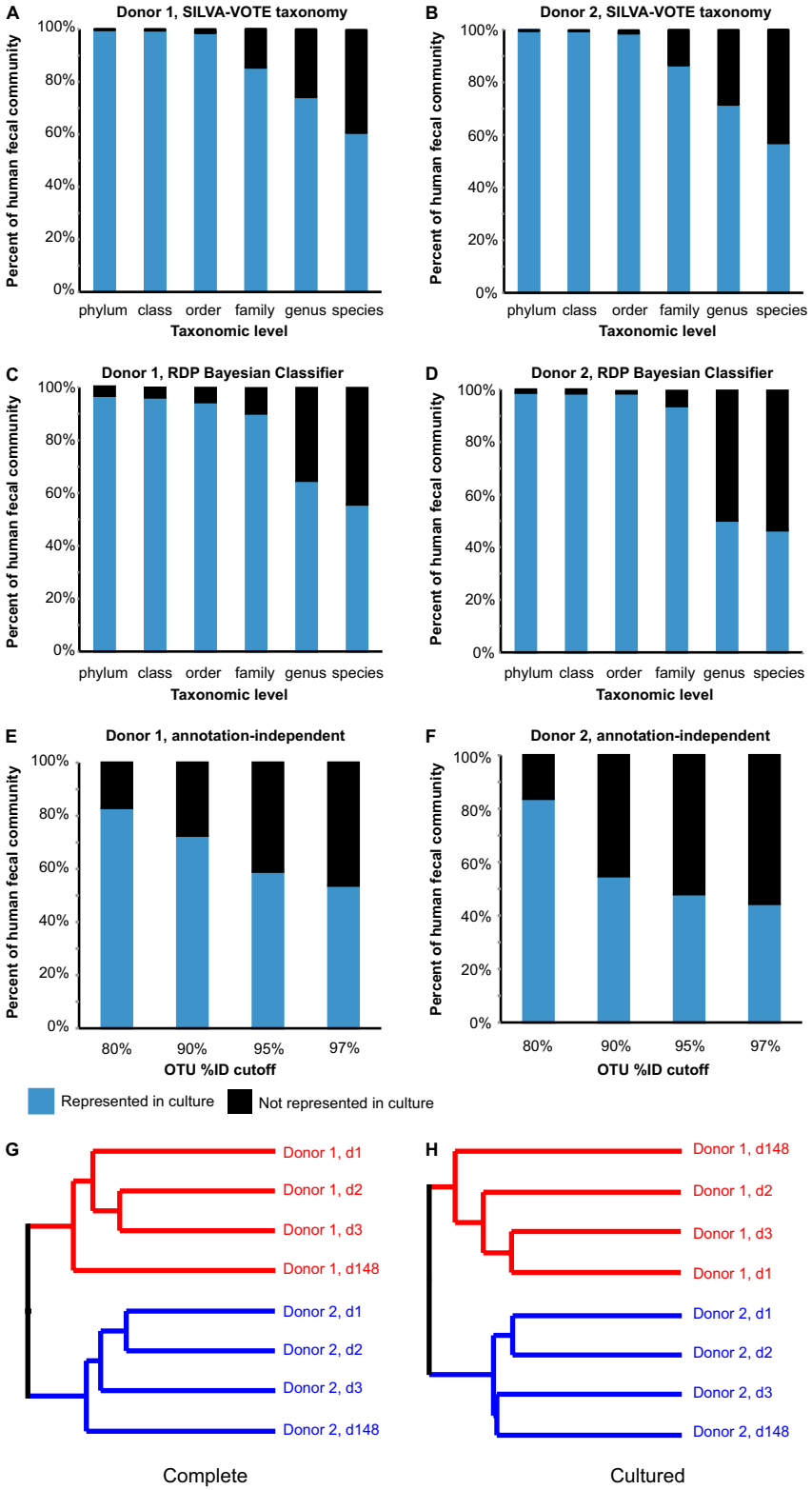


Fig. S1. Abundance of readily cultured taxa in fecal samples from two unrelated human donors, as determined by SILVA-VOTE, Ribosomal Database Project (RDP)-based 16S rRNA annotation, and annotation-independent (OTU %ID cutoff) methods. (*A–F*) Analyses were performed as described in Fig. 1. Unsupervised hierarchical clustering of 16S rRNA datasets generated from either complete uncultured (*G*) or readily cultured (*H*) human gut microbial communities separates all samples from Donor 1 (red) from Donor 2 (blue). Unweighted pair group method with arithmetic mean (UPGMA) clustering of unweighted UniFrac distances between samples (rarefied to 1,000 de-noised, chimera-checked sequences each) is shown. Nonphylogenetic distance metrics (Jaccard, Hellinger, Bray-Curtis) produce similar results (data not shown). In both *G* and *H*, bootstrap support separating Donor 1 samples from Donor 2 samples is 100% [100 iterations; 500 sequences subsampled from each complete (nonrarefied) dataset].

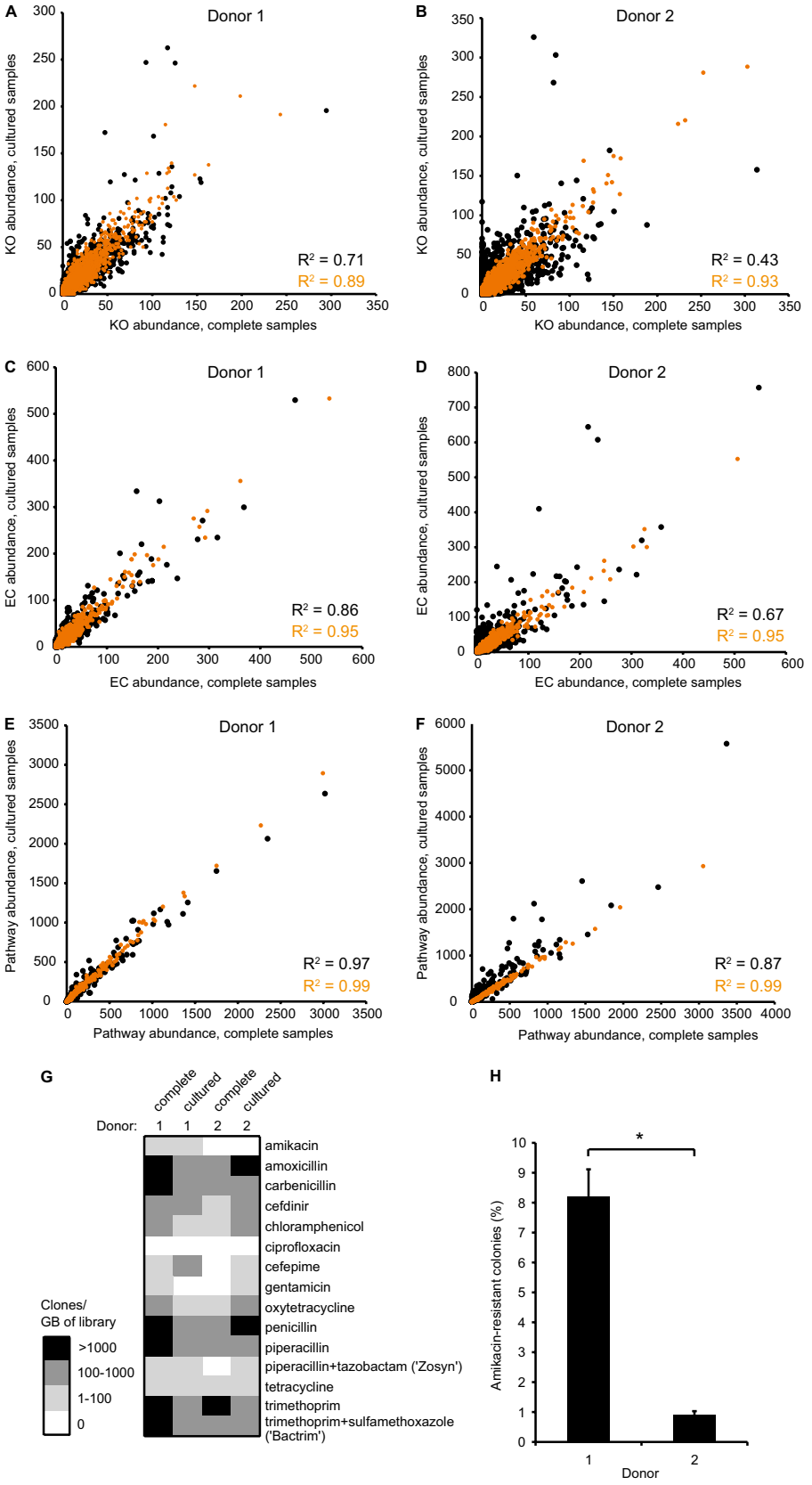


Fig. S2. Relative abundance of functional annotations in the uncultured (complete) and readily cultured fecal communities of two unrelated donors. From each donor, complete and cultured fecal samples also were introduced into germfree mice. After a 4-wk acclimatization period on a standard LF/PP diet, fecal microbiomes were characterized by shotgun pyrosequencing. Reads were mapped to KO (*A* and *B*), EC (*C* and *D*), and level 2 KEGG pathways (*E* and *F*). In these graphs, each point represents a functional annotation, and the axes represent the relative abundance (per 100,000 shotgun pyrosequencer reads) of these

Legend continued on following page

predicted functions in comparisons of complete versus cultured microbiomes (black points) and in comparisons of complete versus cultured microbiomes after each had been introduced into germfree mice (orange points). In each comparison, the goodness of fit (R^2) values increase in communities that share the same environment (mouse gut) regardless of the donor. (*G*) Annotation-independent comparison of functions encoded in the microbiomes of uncultured complete or cultured fecal communities: capture of antibiotic resistance genes. Shades represent number of *E. coli* clones (per GB of subcloned DNA from complete or readily cultured microbial communities from each of two human donors), resulting from each of 15 antibiotic selections. (*H*) Proof-of-principle study connecting captured genes with their associated bacterial sources. Fecal DNA fragments cloned into *E. coli* from Donor 1, but not from Donor 2, conferred resistance to the aminoglycoside amikacin. Replating these fecal samples directly on high levels of amikacin (4,100 μg/mL) reveals that this functional difference is mirrored in the source communities. Mean values ± SEM of triplicate samples (separate frozen aliquots of the original fecal material) are plotted. *$P < 0.005$ based on Student's *t* test (unpaired, two-tailed, assuming equal variance; $P < 0.02$ by heteroscedastic test).

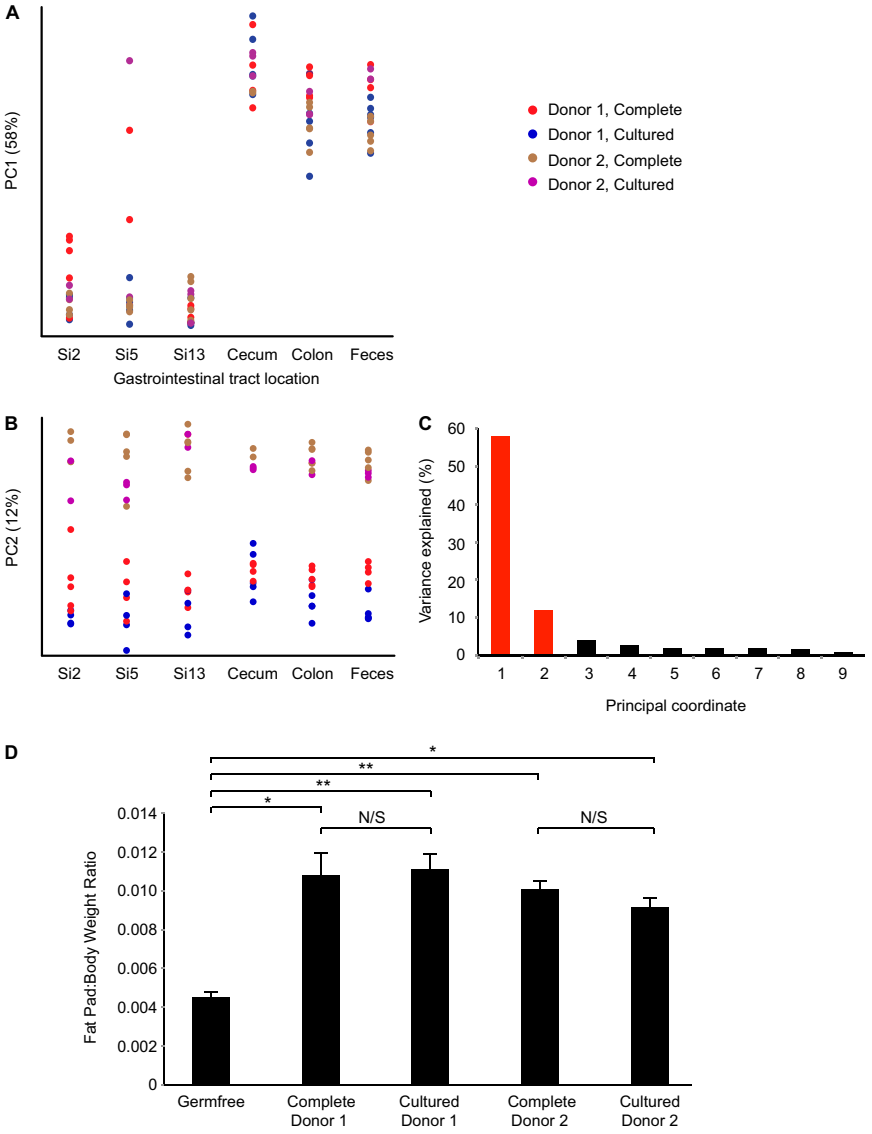


Fig. S3. Biogeography of complete and readily cultured human gut microbial communities in gnotobiotic mice and the impact of colonization on host adiposity. (*A–C*) Principal coordinate analysis (PCoA) of weighted UniFrac distances between samples collected along the length of the gut indicates that mice colonized with readily cultured microbial communities have microbiota similar to those colonized from an uncultured source. (*A*) Principal coordinate 1 (PC1) separates samples by location. (*B*) Principal coordinate 2 (PC2) separates samples by donor. (*C*) No other coordinate explains ≥5% of the total variance between samples. (*D*) Epididymal fat pad:body weight ratios in germfree mice and those colonized with complete or readily cultured microbial communities from the two human donors. Ratios represent the average ± SEM from $n = 5$ mice per group (except the Donor 2 cultured community; $n = 3$). Asterisks indicate statistically significant differences based on an unpaired, two-tailed Student's *t* test. *$P < 0.005$; **$P < 0.001$; N/S, not significant.

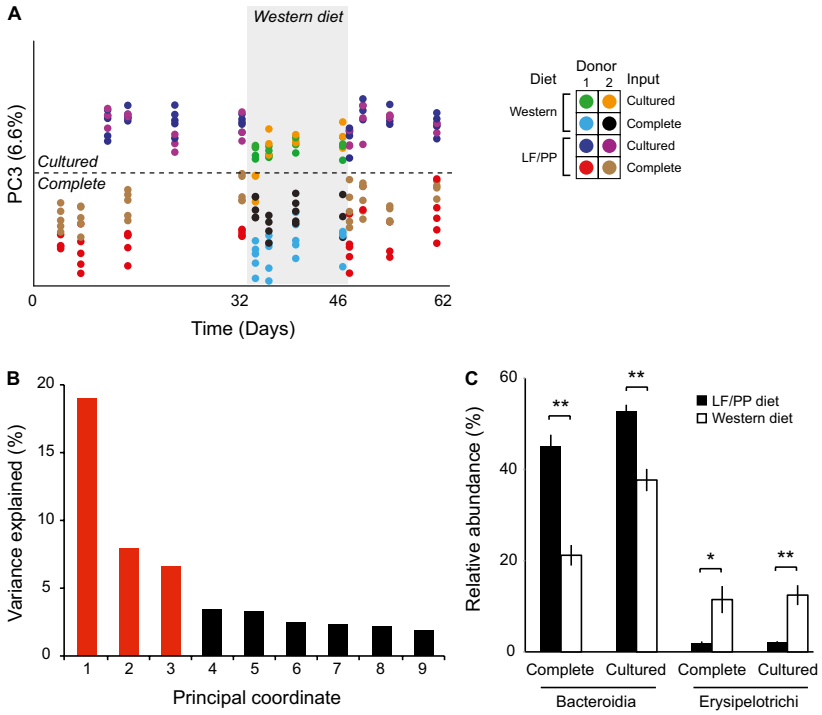


Fig. S4. (Continued)

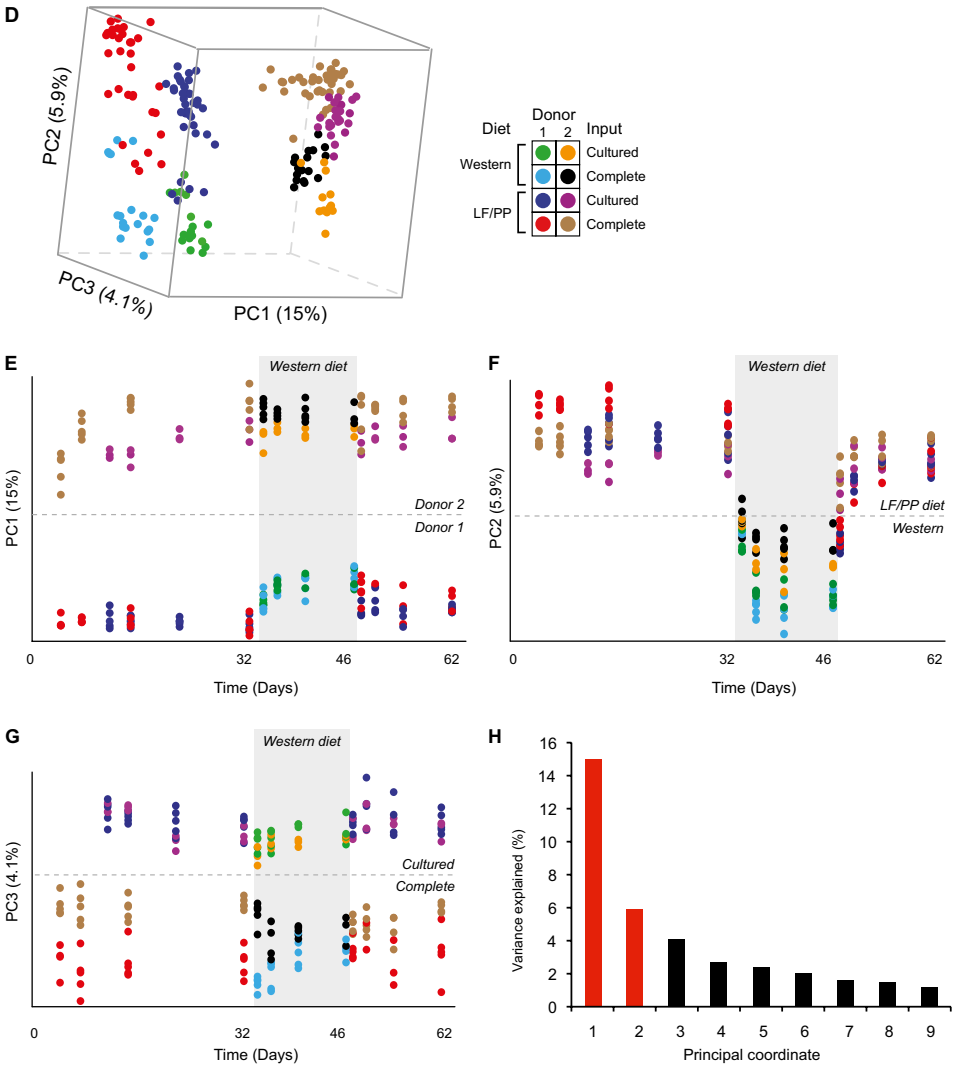


Fig. S4. Diet shapes complete and readily cultured human gut microbial communities in a similar manner. (*A*) PCoA of unweighted UniFrac distances between fecal samples obtained from mice colonized with complete or cultured microbial communities from two unrelated human donors. On day 33 after gavage, mice were switched from a LF/PP chow to a high-fat, high-sugar Western diet (gray rectangle). On day 47 they were returned to the original LF/PP diet. Variance along principal coordinate 3 (PC3) is plotted against time. (*B*) Scree plot from PCoA analysis. Only PC1-PC3 (red) account for ≥5% of the variance between samples. (*C*) In gnotobiotic mice, communities composed of readily cultured human gut microbes and communities containing a complete human gut microbiota undergo similar diet-dependent changes in abundance of Bacteroidia and Erysipelotrichi upon changes in host diet. Each mouse in each treatment group was caged individually, and each group that received a given community was housed in a separate gnotobiotic isolator. Mean values ± SEM and *P* values (**P* < 0.05; ***P* < 0.01 based on a paired, two-tailed Student's *t* test) are indicated when mice were consuming a LF/PP diet (black bars) and then switched to the Western diet (white bars). (*D–H*) Diet-dependent community-wide shifts in bacterial species representation as measured by PCoA analysis based on a non-phylogenetic (binary Jaccard) distance measurement. (*D*) PCoA plot of binary Jaccard distances between all samples. (*E–G*) Separate PCoA values plotted against time. (*H*) Scree plot of variance explained by PCoA axes.

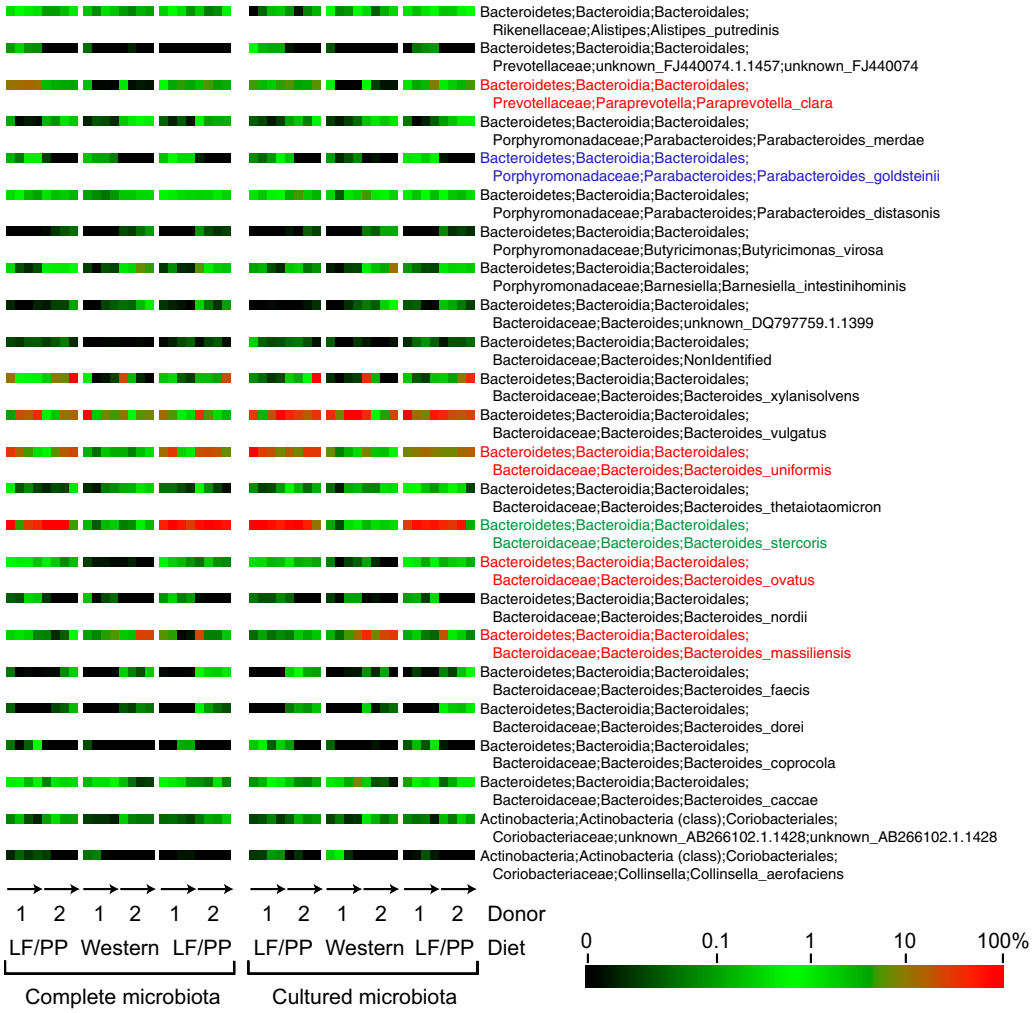


Fig. S5. (Continued)

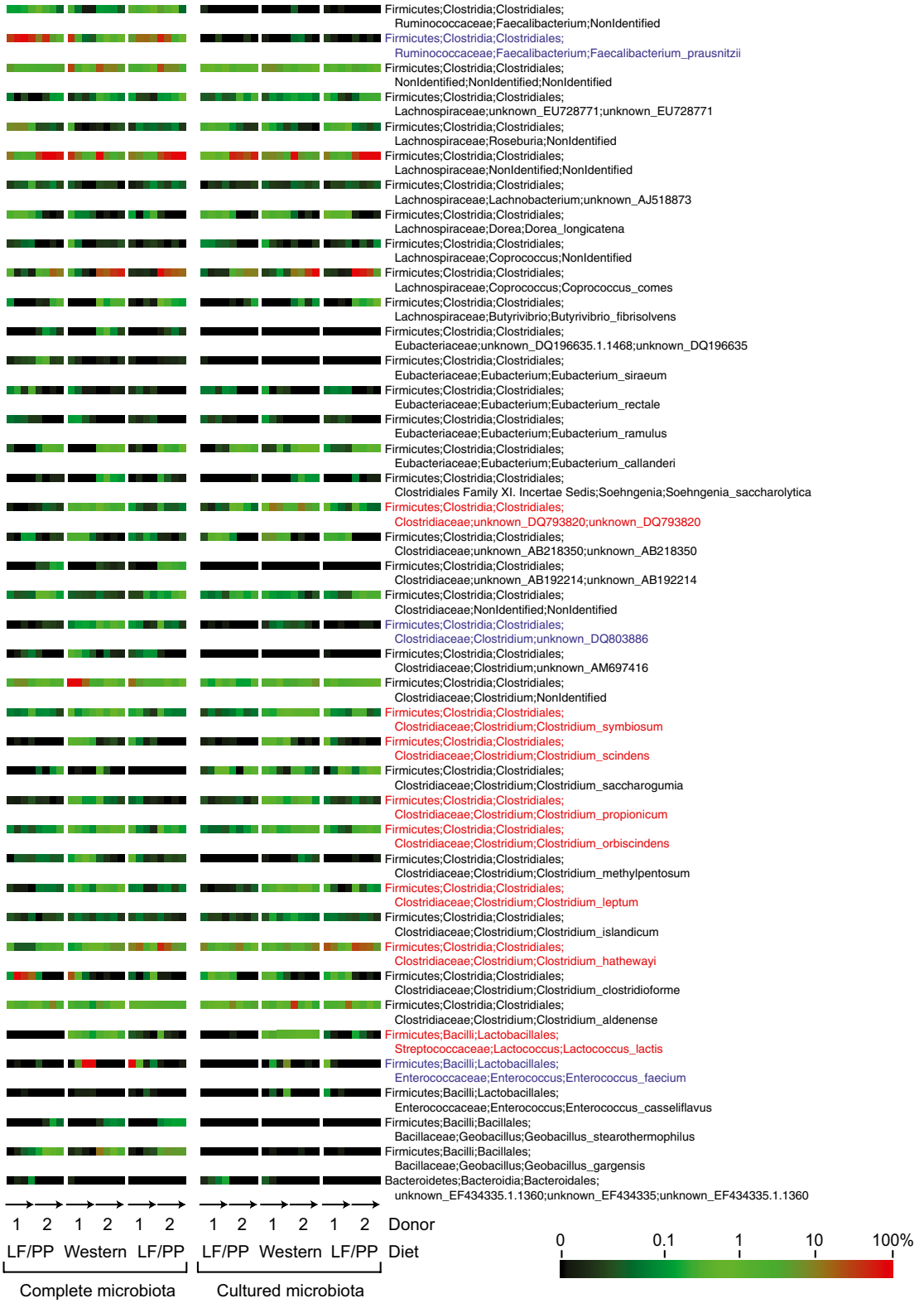


Fig. S5. (Continued)

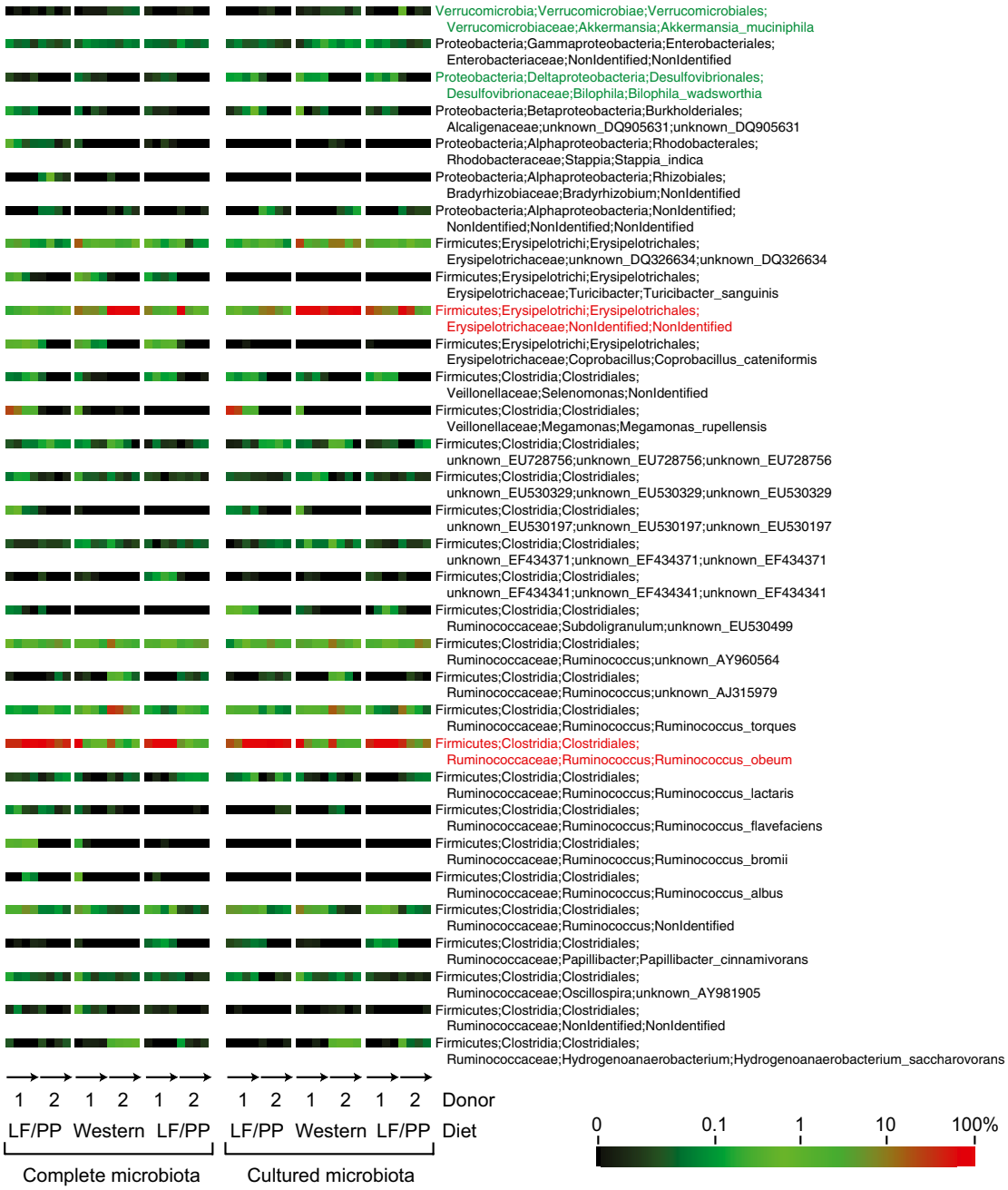


Fig. S5. Relative abundances of species-level taxa in fecal samples obtained from gnotobiotic mice over time. All identified taxa present at an abundance of ≥0.1% in at least a single time point are shown. Species significantly influenced by diet in either the complete community (blue names), the readily cultured community (green names), or both (red names) are plotted over time (arrows) during the initial LF/PP, subsequent Western, and final LF/PP phases of the diet-oscillation experiment. Significance ($P \leq 0.01$ after Bonferroni correction) was determined by unpaired, two-tailed Student's t test, assuming equal variances; $n = 97$ taxa tested. The assumption of equal variances was tested by F test ($P < 0.02$).

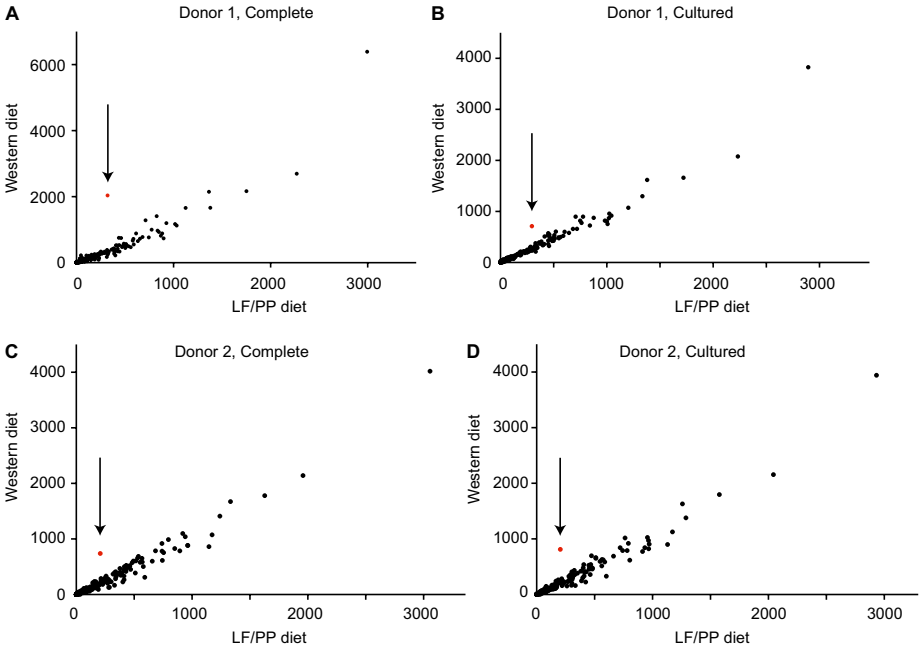
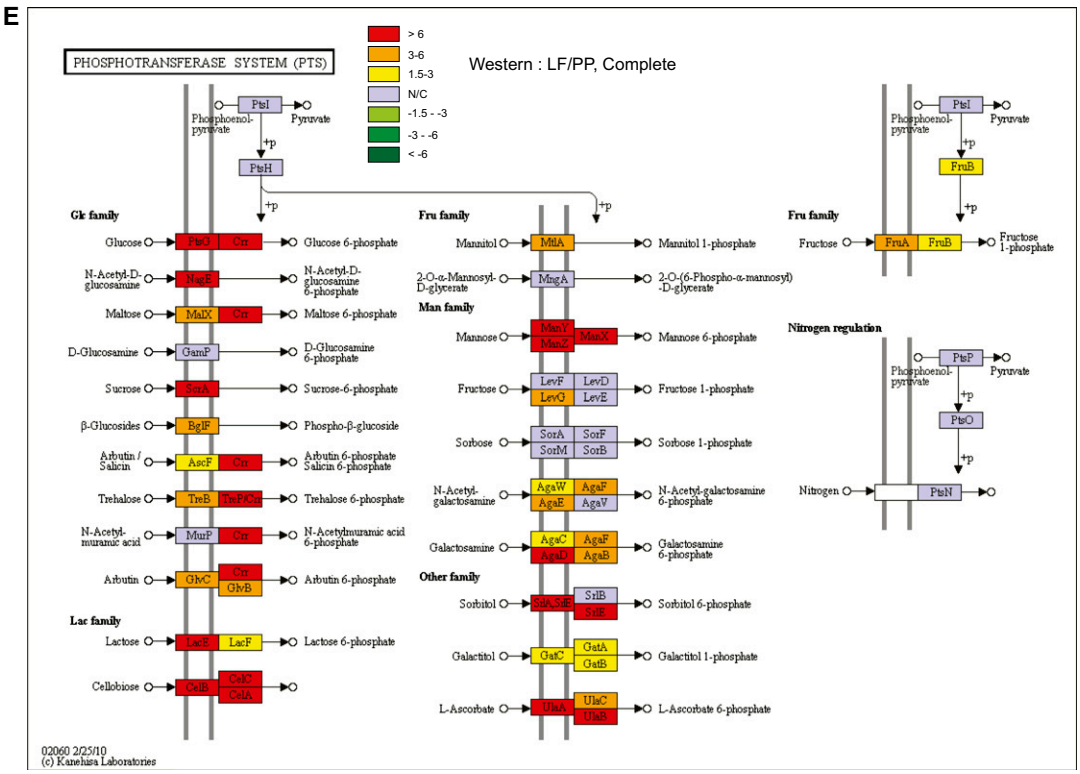


Fig. S6. (Continued)



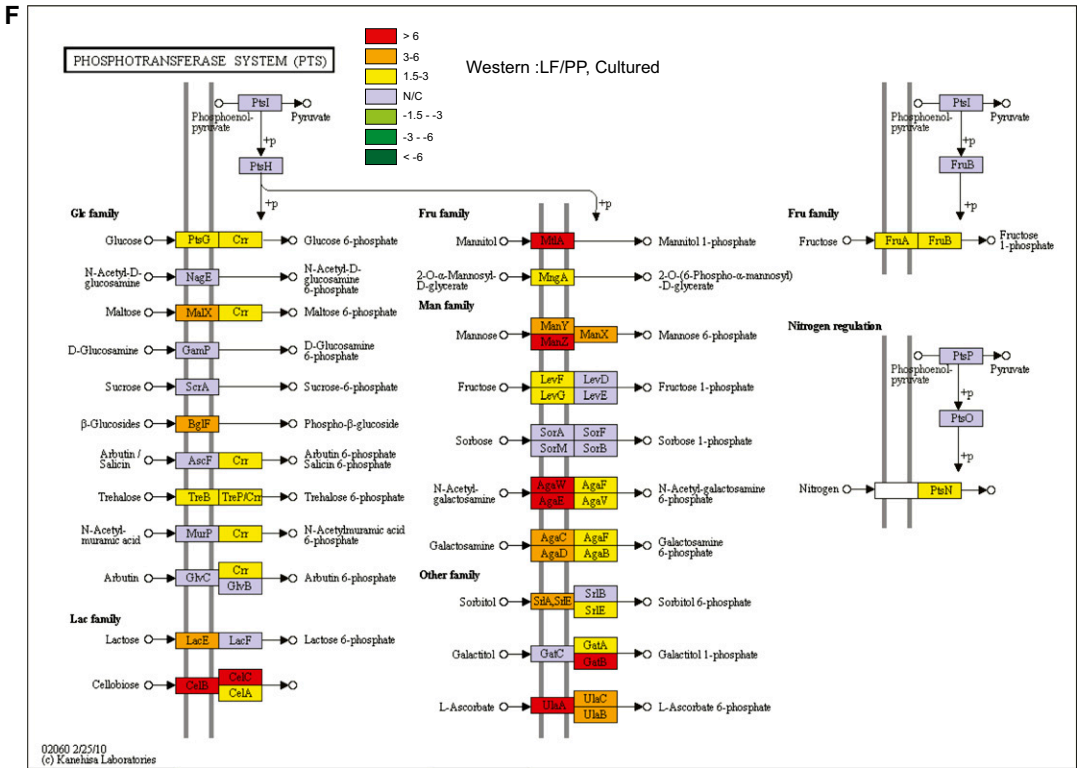


Fig. S6. KEGG level 2 pathway-based analysis of fecal microbiomes obtained from LF/PP- and Western diet-fed mice colonized with a complete or cultured human gut microbiota from two human donors. (*A*–*D*) Phosphotransferase system (PTS) pathways are marked in red and highlighted with arrows. (*E* and *F*) Multiple predicted KEGG PTS pathway components are enriched in the fecal microbiomes of mice colonized with complete (*E*) or cultured (*F*) human gut microbial communities and maintained on a high-fat, high-sugar Western diet. KO level-predicted functional annotations are colored by average fold-difference in their representation in microbiomes obtained from mice on the different diets (Western versus LF/PP). Data represent averages from mice colonized with the complete or cultured fecal communities from two unrelated human donors.

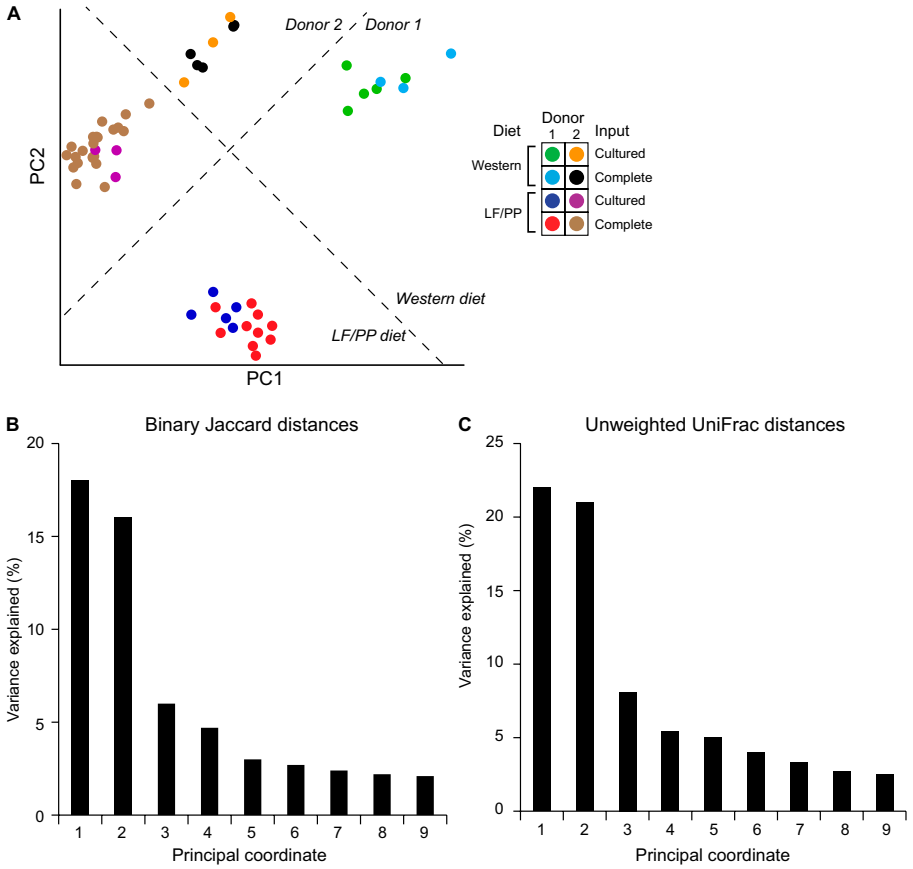


Fig. S7. Plated communities of human gut microbes can be reshaped through diet selection in gnotobiotic mice. (*A*) PCoA analysis of unweighted UniFrac distances between communities collected from mice before and after a diet switch and plated on GMM. Scree plots display variance explained by PCoA analysis of binary Jaccard (*B*) or unweighted UniFrac (*C*) distances between samples.

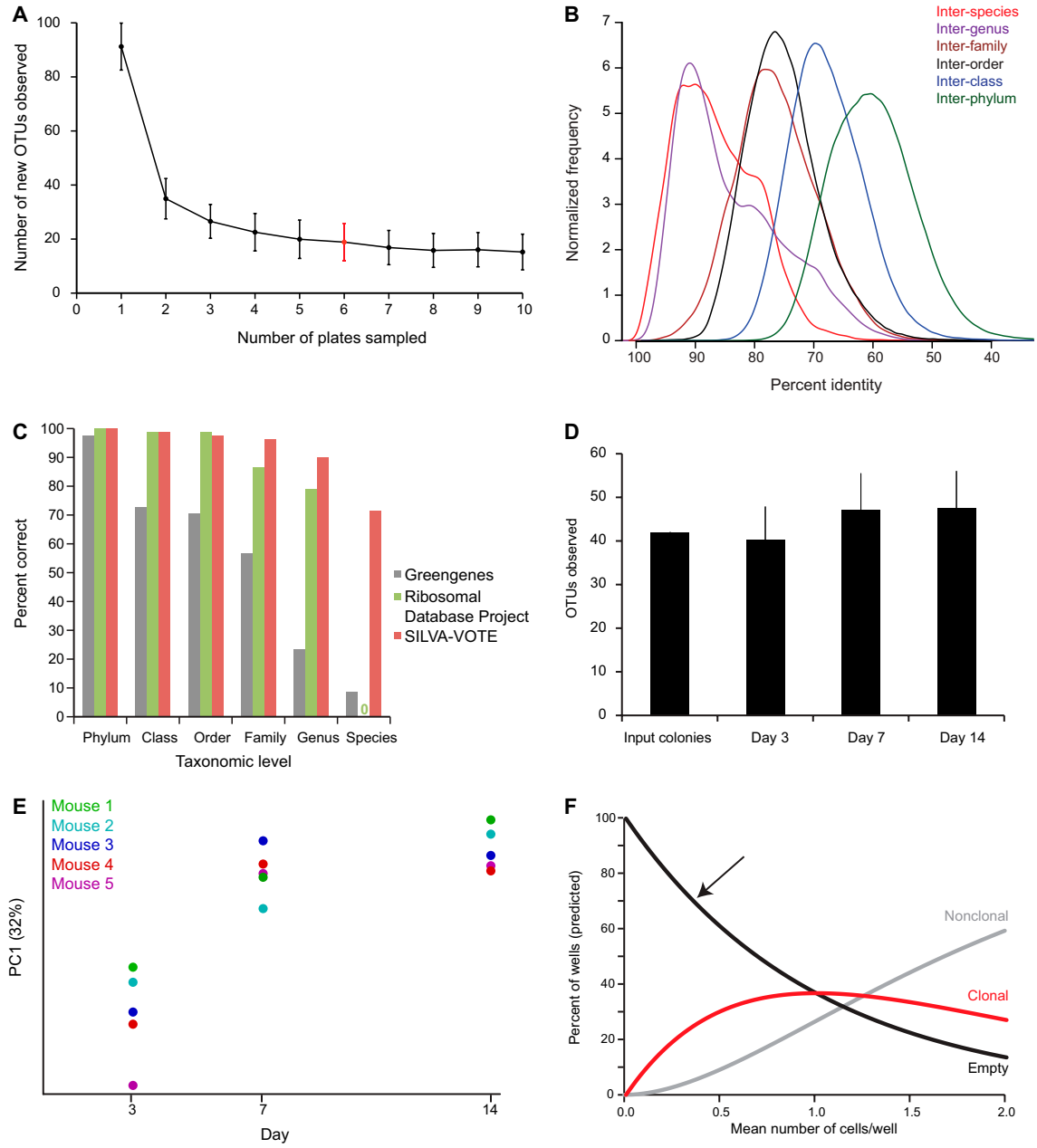


Fig. S8. (Continued)

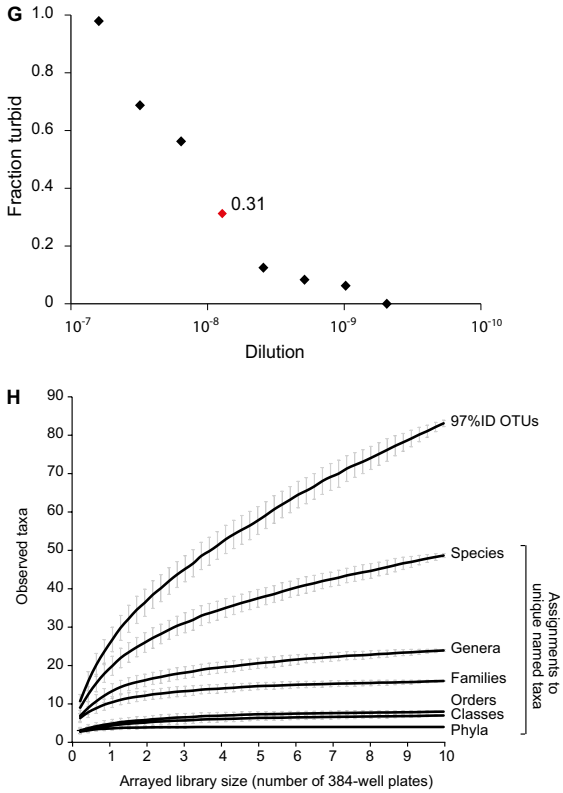


Fig. S8. Experimental parameters for *en masse* culturing, taxonomic assignment, inoculation of germfree mice, and arrayed strain collections. (*A*) Most readily cultured OTUs in a human fecal sample are observed in six GMM plates. Rarefaction analysis describes the number of new OTUs observed with each additional agar plate added to the dataset (10 plates were prepared independently). Points reflect the mean value after 100 iterations (plates sampled in random order without replacement); error bars represent one SD. The mean number of new OTUs identified per plate drops below 20 with six or more plates (red). (*B*) Distributions of percent identities between pairwise comparisons of V2 16S rRNA gene sequences from representatives of bacterial taxa with varying degrees of shared phylogeny. We selected 4,041 16S rRNA sequences from the SILVA database (v102) that contain complete V2 regions and full (species-level) SILVA-VOTE annotations. Sequences were aligned using PyNast. After removal of gap-only columns, the %ID between V2 regions was calculated for each pairwise comparison. The resulting %ID distributions are plotted for members of two different species within the same genus (interspecies), two different genera of the same family (intergenus), and so forth. (*C*) Comparison of three methods for assigning taxonomy to V2 16S rRNA sequences. Reference genomes are listed in Table S11. (*D* and *E*) Sequences identified in gnotobiotic mice colonized with a readily cultured human gut microbiota do not reflect nongrowing or dormant cells. (*D*) Alpha-diversity analysis of fecal microbial communities of mice that had been inoculated with the control sample described in *SI Materials and Methods*. Diversity is similar at the 7-d and 14-d time points. (*E*) Time-course beta-diversity analysis of gnotobiotic mice inoculated with the control sample. UniFrac distances were calculated between all samples and represented spatially by PCoA. In this figure, variance along the major coordinate of variance (PC1) is plotted against time; colors represent individual mice. (*F*) Communities of cultured human gut microbes can be clonally archived in 384-well format by limiting dilution. At a dilution of a human fecal sample that produces 70% empty wells, a Poisson distribution predicts that 25% of wells will contain a clonal population of cells and that 5% of wells will be nonclonal (arrow). (*G*) Optimization of dilutions for arrayed strain collections. A 1.28×10^8-fold dilution of the stored fecal aliquot and subsequent inoculation into 384-well trays (170 μL per well) produces 70% empty wells. (*H*) Rarefaction analysis indicates that ten 384-well trays are sufficient to capture nearly all the genus-level diversity that can be retrieved from a single fecal sample using this arrayed culturing method.

Other Supporting Information Files

Table S1 (XLS)
Table S2 (XLS)
Table S3 (XLS)
Table S4 (XLS)
Table S5 (XLS)
Table S6 (XLS)
Table S7 (XLS)
Table S8 (XLS)
Table S9 (XLS)
Table S10 (XLS)
Table S11 (XLS)
Table S12 (XLS)
Table S13 (XLS)
Table S14 (XLS)