Supporting Information

Mahowald et al. 10.1073/pnas.0901529106

SI Methods

Bacterial Culture. Bacterial strains were stored frozen at −80 °C in a prereduced mixture of 2 parts TYG medium (1) to 1 part glycerol. Bacteria were routinely cultured in TYG medium in an anaerobic chamber (Coy Lab Products, Grass Lake, MI) under an atmosphere of 40% CO_2, 58% nitrogen, and 2% H_2. To assay growth of *E. rectale* on specific carbon sources, the organism was cultured on medium containing 1% tryptone, 100 mM potassium phosphate buffer (pH 7.2), 15 mM NaCl, 180 μM $CaCl_2$, 100 μM $MgCl_2$, 50 μM $MnCl_2$, 42 μM $CoCl_2$, 15 μM $FeSO_4$, 1% trace element mix (ATCC), 2 μg/mL folinic acid (calcium salt), 1.2 μg/mL hematin, and 1 mg/mL menadione. Growth curves for different carbon sources were acquired at 37 °C in the Coy anaerobic chamber using a 96-well plate spectrophotometer (Tecan). Growths were scored as positive if the OD_{600} measurement rose by ≥0.2 over a 72-h incubation at 37 °C.

Genome Sequencing. *E. rectale* and *E. eligens* were grown to late log phase under anaerobic conditions in TYG medium. Cells harvested from a 50 mL culture were lysed by incubation for 30 min at 37 °C in 11 mL of Buffer B1 (Qiagen Genomic DNA buffer set; Qiagen) supplemented with 2.2 mg of RNase A, 50 units lysozyme, 50 units mutanolysin, and 600 units achromopeptidase (all from Sigma) followed by addition of 4 mL of Buffer B2 (Qiagen) together with 10 mg (300 units) proteinase K (Sigma) and incubation at 50 °C for 2 h. DNA was precipitated by adding 1.5 mL of 3M sodium acetate and 30 mL of isopropanol, removed with a sterile glass hook, and washed several times with ethanol.

Unlike *E. eligens*, genomic DNA from *E. rectale* was very resistant to standard cloning techniques. This cloning bias made efforts to produce fosmids ineffective, and left vast regions of the genome uncloned in our primary sequencing vector, pOT. Only half (1.7 Mb) of its genome was represented in our initial assembly containing 228 contigs from >9× plasmid shotgun reads with an ABI 3730xl capillary instrument. Therefore, we generated >40× coverage of the *E. rectale* genome through pyrosequencing with a 454 GS20 instrument, and used an additional vector (pJAZZ) for capillary sequencing to obtain a finished genome sequence.

Protein-coding genes were identified with Glimmer 2.13 (2) and GeneMarkS (3), using the start site predicted by GeneMarkS where the two overlapped. "Missed" genes were then added by using a translated BLAST of intergenic regions against the nonredundant (NR) protein database and finding conserved ORFs. Additional missed genes were added to the *E. rectale* genome using YACOP (trained by Glimmer 2.13) (4). tRNA, rRNA and other non-coding RNAs were identified and annotated using tRNAscan-SE (5), RNAMMER (6), and RFAM (7), respectively. Protein-coding genes were annotated with the KEGG Orthology group definition using a National Center for Biotechnology Information BLASTP search (8) of the KEGG genes database (9), with a minimum bit score of 60.

Animal Husbandry. All experiments using mice were performed using protocols approved by the animal studies committee of Washington University. NMRI-KI mice (10) were maintained in flexible plastic film isolators under a strict 12 h light cycle, and fed an irradiated standard low-fat, high plant polysaccharide chow (LF/PP, diet 2018 from Harlan Tecklad, www.tekladcustomdiets.com) or a high fat, high-sugar (HF/HS) Western-style diet (Harlan Teklad 96132) or a corresponding control low fat, high-sugar (LF/HS; Harlan Teklad 03317).

Animals were colonized via gavage with 10^8 CFU from an overnight culture of *B. thetaiotaomicron* or a log-phase culture of *E. rectale*. Gavage with *E. rectale* was repeated on 3 successive days using cells from separate log-phase cultures begun from separate colonies. Cecal contents were flash frozen in liquid nitrogen immediately after animals were killed.

Quantitative (q) PCR Measurements of Colonization. A total of 100–300 mg of frozen cecal contents from each gnotobiotic mouse was added to 2 mL tubes containing 250 μL of 0.1 mm-diameter zirconia/silica beads (Biospec Products), 0.5 mL of Buffer A (200 mM NaCl, 20 mM EDTA), 210 μL of 20% SDS, and 0.5 mL of a mixture of phenol:chloroform:isoamyl alcohol (25:24:1; pH 7.9; Ambion). Samples were lysed with a bead beater (BioSpec; "high" setting for 4 min at room temperature). The aqueous phase was extracted after centrifugation (8,000 × *g* at 4 °C for 3 min), and the extraction repeated with another 0.5 mL of phenol:chloroform:isoamyl alcohol and 1 min of vortexing. DNA was precipitated with 0.1 volume of 3M sodium acetate (pH 5) and 1 volume of isopropanol (on ice for 20 min), pelleted (14,000 × *g*, 20 min at 4 °C) and washed with ethanol. The resulting pellet was resuspended in water and 1/2 (for *E. rectale* mono-associations) or 1/10 of the DNA (for *B. thetaiotaomicron*-colonized samples) cleaned up further using a DNAEasy column (Qiagen).

qPCR was performed using (*i*) primers specific to the 16S rRNA gene of *B. thetaiotaomicron* (11) or to the *Clostridium coccoides*/*E. rectale* group (forward: 5′-CGGTACCTGACTAAGAAGC-3′; reverse: 5′-AGTTT(C/T)ATTCTTGCGAACG-3′) (12), and (*ii*) conditions described previously for *B. thetaiotaomicron* (11). The amount of DNA from each genome in each PCR was computed by comparison to a standard curve of genomic DNA prepared in the same manner from pure cultures of each bacterial species. Data were converted to genome equivalents by calculating the mass of each finished genome (2.8×10^5 genome equivalents (GEq) per ng *E. rectale* DNA, and 1.5×10^5 GEq per ng *B. thetaiotaomicron* DNA).

GeneChip Design, Hybridization, and Data Analysis. A custom, 6-species human gut microbiome Affymetrix GeneChip was designed using the finished genome sequences of *B. thetaiotaomicron, B. vulgatus, P. distasonis* and *M. smithii* (13–15), plus draft versions of the *E. rectale* and *E. eligens* genomes. Gene predictions for the Firmicute assemblies were made using Glimmer3 (2). The design included 14 probe pairs (perfect match plus mismatch) per CDS (protein coding sequence) in each draft assembly, and 11 probe pairs for each CDS in a finished genome.

Non-specific cross-hybridization was controlled in 2 ways. *First*, probe masks for each genome were developed as follows. For analysis of *E. rectale-B. thetaiotaomicron* cocolonizations, probes resulting from misassembly and missing sequences later identified (from the finished genome) in the *E. rectale* draft assembly were removed to avoid cross-hybridization. National Center for Biotechnology Information BLASTN (8) was used, with parameters adjusted for small query size (word size 7, no filtering or gaps), to identify probesets that either failed to find a perfect match in the finished genomes, or that registered a hit to >1 sequence feature with a bit score ≥38 (using the default scoring parameters for BLASTN). This mask reduced the proportion of probesets exhibiting a spurious "Present" call (by

Affymetrix software) by 36%. The resulting CDF file was imported into BioConductor using the altcdfenvs package, and all expression analyses were performed using the MAS5 algorithm implemented in BioConductor's "Affy package" (16), after masking of GeneChip imperfections with Harshlight (17)—in both cases using the default parameters. *Second,* for all analyses, we also identified all *B. thetaiotaomicron* and *E. rectale* probesets that registered a "Present" call due to cross-hybridization with targets generated from RNA isolated from the cecal contents of mice that had been mono-associated with either *E. rectale*, or *B. thetaiotaomicron.* These probesets were also excluded from further analyses Expression values were computed using Bioconductor.

Expression of selected genes was confirmed by qRT-PCR (11). Primers used for these reactions are available from the authors on request.

Proteomic Analyses of Cecal Contents. Cecal contents were processed via a single tube cell lysis and protein digestion method as follows. Briefly, the cell pellet was resuspended in 6 M Guanidine/10 mM DTT, heated at 60 °C for 1 h, followed by an overnight incubation at 37 °C to lyse cells and denature proteins. The guanidine concentration was diluted to 1 M with 50 mM Tris/10 mM $CaCl_2$ (pH 7.8), and sequencing grade trypsin (Promega) was added (1:100; wt/wt). Digestions were run overnight at 37 °C. Fresh trypsin was then added followed by additional 4 h incubation at 37 °C. The complex peptide solution was subsequently de-salted (Sep-Pak C_{18} solid phase extraction; Waters), concentrated, filtered, aliquoted and frozen at −80 °C. All 8 samples were coded and mass spectrometry measurements conducted in a blinded fashion.

Cecal samples were analyzed in technical triplicates using a 2-dimensional (2D) nano-LC MS/MS system with a split-phase column (SCX-RP) (18) on a linear ion trap (Thermo Fisher Scientific) with each sample consuming a 22-h run as detailed elsewhere (19, 20). The linear ion trap (LTQ) settings were as follows: dynamic exclusion set at 1; and 5 data-dependent MS/MS. Two microscans were averaged for both full and MS/MS scans and centroid data were collected for all scans. All MS/MS spectra were searched with the SEQUEST algorithm (21) against a database containing the entire mouse genome, plus the *B. thetaiotaomicron*, and *E. rectale*, genomes (common contaminants such as keratin and trypsin were also included). To find potential food proteins, yeast and rice databases were included. The break down of each database can be found and downloaded from http://compbio.ornl.gov/gnotobiotic_mouse_cecal_metaproteome/databases/. The SEQUEST settings were as follows: enzyme type, trypsin; Parent Mass Tolerance, 3.0; Fragment Ion Tolerance, 0.5; up to 4 missed cleavages allowed (internal lysine and arginine residues), and fully tryptic peptides only (i.e., both ends of the peptide must have arisen from a trypsin-specific cut, except the N and C termini of proteins). All datasets were filtered at the individual run level with DTASelect (22) [Xcorrs of at least 1.8 (+1 ions), 2.5 (+2 ions) 3.5 (+3 ions)]. Only proteins identified with 2 fully tryptic peptides were considered. All resulting DTASelect/Contrast files used in this study are available from http://compbio.ornl.gov/gnotobiotic_mouse_cecal_metaproteome. Also accessible from this site are MS/MS spectra for all identified peptides.

For this study, false-positive rates (FPR) were used to estimate the error associated with peptide identifications. The overall FPR was estimated using the formula: $FPR = 2[n_{rev}/(n_{rev} + n_{real})] \times 100$ where n_{rev} is the number of peptides identified from the reverse database and n_{real} is the number of peptides identified from the real database (23). Reverse and shuffled databases were created to calculate FPRs (23, 24). A reverse database was created by precisely reversing each protein entry (i.e., N terminus became C terminus in each case) and then appending these reversed sequences onto the original database. Two runs—samples 705, Run 1 and 710, Run 2—were randomly selected for estimating a FPR. The observed FPR rates were 0.55% and 0.31% respectively for these 2 runs. An additional database was created by randomly shuffling the amino acids of each protein rather than simply reversing the N terminus and C terminus. A FPR was estimated using a similar formula as that described above except that the number of identified reverse peptides was replaced with the number of shuffled peptides. A FPR was estimated for both samples, 705, Run 1 (0.45%) and 710, Run2 (0.31%) and was similar to the rate determined by the reverse database method. Datasets for calculating FPR rates are available on the web site mentioned above.

In addition to differentiating between true and false peptide identifications with FPRs, label-free quantitation methods were used to estimate relative protein abundance. Several protein quantitation methods are currently available and routinely performed for shotgun proteomics analyses. To estimate relative protein abundance in complex protein mixtures and communities, spectral counts and normalized spectral abundance factors (NSAF) (25) are commonly used. Spectral counting is based on the theory that the more abundant peptides are typically sampled more frequently, resulting in higher spectral counts. Lui et al. (26) has shown that spectral copy number provides a more accurate correlation to protein abundance than peptide count and % coverage. NSAF, however, is based on spectral counts, but takes into account protein size and the total number of spectra from a run, thus normalizing the relative protein abundance between samples (25). Both methods were performed and the results for all samples and runs can be found on the web site. The list of all identified proteins from all runs and sample types with spectral counting approach can be found in Table S6A. The same list with % coverage, peptides and NSAF can be found on the web site.

Biochemical Analyses. Measurements of acetate, propionate, butyrate, NAD+, and NADH in cecal contents were performed as described in ref. 11, with the exception that acetic acid-1-^{13}C,d_4 (Sigma) was used as a standard to control for acetate recovery.

SI Results

Comparative Genomic Studies of Human Gut-Associated Firmicutes and Bacteroidetes. Although the sequenced gut Bacteroidetes all harbor large sets of polysaccharide sensing, acquisition and degradation genes, the gut Firmicutes, including *E. rectale* and *E. eligens,* have smaller genomes and a significantly smaller proportion of genes involved in glycan degradation (Fig. S2). The gut-associated Bacteroidetes possess large families of SusC and SusD paralogs involved in binding and import of glycans, while the genomes of *E. rectale* and other gut Firmicutes are enriched for phosphotransferase systems and ABC transporters (Fig. S2). Lacking adhesive organelles, the ability of gut Bacteroidetes to attach to nutrient platforms consisting of small food particles and host mucus via glycan-specific SusC/SusD outer membrane binding proteins likely increases the efficiency of oligo- and monosaccharide harvest by adaptively expressed bacterial GHs, and preventing washout from the perfused gut bioreactor. Unlike the surveyed Bacteroidetes, several Firmicutes, notably *E. rectale, E. eligens*, *E. siraeum,* and *Anaerotruncus colihominis* (the later belongs to the *Clostridium leptum* cluster) possess genes specifying components of flagellae (Fig. S2): These organelles may contribute to persistence within the continuously perfused gut ecosystem and/or enable these species to move to different microhabitats to access their preferred nutrient substrates.

Table S2 lists predicted GHs and PLs present in the Firmicutes and Bacteroidetes surveyed, sorted into families according to the scheme incorporated into the Carbohydrate Enzymes (CAZy)

database (www.cazy.org). The Firmicutes have fewer total polysaccharide-degrading enzymes than the Bacteroidetes. Nonetheless, most of the sampled Firmicutes have sets of carbohydrate active enzyme families whose proportional representation in their genomes is significantly greater than in the sampled human gut Bacteroidetes. For example, while *E. rectale* and *E. eligens* lack a variety of enzymes to degrade host-derived glycans present in mucus and/or the apical surfaces of gut epithelial cells (e.g., fucosidases and hexosaminidases), *E. rectale* has a disproportionately large number of predicted α-amylases (GH family 13; Table S2 and Fig. S3). *E. eligens* has fewer of the latter, but possesses a number of enzymes for degrading pectins (e.g., GH family 28, PL families 1 and 9) (Table S2). Among the Bacteroidetes "glycobiomes", there is also evidence of niche specialization: Although *B. vulgatus* has fewer GHs and PLs overall than *B. thetaiotaomicron*, it has a larger assortment of enzymes for degrading pectins (GH family 28 and PL families 1, 10 and 11) and possesses enzymes that *B. thetaiotaomicron* lacks that should enable it to degrade certain xylans [GH family 10 and Carbohydrate esterase (CE) family 15] (Fig. S3 and Table S2). The results of in vitro assays of the growth of *B. thetaiotaomicron* and *E. rectale* in defined medium containing mono- di- and polysaccharides are summarized in Table S3.

Proteomic Analysis. For a complete list of the total number of identified spectra, peptides and proteins per sample and run, see Table S6B. Interestingly, the total number of identified spectra were, for the most part, distinct and unique to each bacterial species. Unlike *B. thetaiotaomicron* and *E. rectale*, the number of identified spectra belonging to mouse were redundant: Thus, a higher number of spectra were non-unique spectra. The difference is evident when the total spectra counts are compared with unique spectra counts only. The total average spectra count identified in the control (germ-free) mouse was 10,767 for sample 700 and 11, 221 for sample 799. The total average unique spectra count, however, decreased to 4,394 and 4,168. Therefore, the majority of mouse peptides are not unique within the database.

The total number of unique spectra counts per species and run can be found in Table S6C. The 2 cocolonized mice (710 and 810) had a total of $\approx$77% unique spectra belonging to *B. thetaiotaomicron*, 20% unique spectra belong to *E. rectale*, and only 3% of the 2 species' combined spectra counts were non-unique. This suggests that the majority of identified proteins belonging to *B. thetaiotaomicron* and *E. rectale* are true unique identifications and these species can be easily differentiated by proteomics.

1. Holdeman LV, Cato EP, Moore WEC (1977) *Anerobe Laboratory Manual* (Virginia Polytechnic Institute and State University, Blacksburg, VA).
2. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 27:4636–4641.
3. Besemer J, Lomsadze A, Borodovsky M (2001) GeneMarkS: A self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 29:2607–2618.
4. McHardy AC, Goesmann A, Puhler A, Meyer F (2004) Development of joint application strategies for two microbial gene finders. *Bioinformatics* 20:1622–1631.
5. Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964.
6. Lagesen K, et al. (2007) RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108.
7. Griffiths-Jones S, et al. (2005) Rfam: Annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* 33:D121–124.
8. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
9. Kanehisa M, et al. (2006) From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Res* 34:D354–357.
10. Bry L, Falk PG, Midtvedt T, Gordon JI (1996) A model of host-microbial interactions in an open mammalian ecosystem. *Science* 273:1380–1383.
11. Samuel BS, Gordon JI (2006) A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism. *Proc Natl Acad Sci USA* 103:10011–10016.
12. Rinttila T, Kassinen A, Malinen E, Krogius L, Palva A (2004) Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. *J Appl Microbiol* 97:1166–1177.
13. Samuel BS, et al. (2007) Genomic and metabolic adaptations of Methanobrevibacter smithii to the human gut. *Proc Natl Acad Sci USA* 104:10643–8.
14. Xu J, et al. (2007) Evolution of symbiotic bacteria in the distal human intestine. *PLoS Biol* 5:e156.
15. Xu J, et al. (2003) A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* 299:2074–2076.
16. Gentleman RC, et al. (2004) Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol* 5:R80.
17. Suarez-Farinas M, Pellegrino M, Wittkowski KM, Magnasco MO (2005) Harshlight: A"corrective make-up" program for microarray chips. *BMC Bioinformatics* 6:294.
18. McDonald WH, Ohi R, Miyamoto DT, Mitchison TJ, Yates JR (2002) Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: Single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. *Int J Mass Spectrom* 219:245–251.
19. Thompson MR, et al. (2007) Dosage-dependent proteome response of *Shewanella oneidensis* MR-1 to acute chromate challenge. *J Proteome Res* 6:1745–57.
20. VerBerkmoes NC, et al. (2006) Determination and comparison of the baseline proteomes of the versatile microbe Rhodopseudomonas palustris under its major metabolic states. *J Proteome Res* 5:287–298.
21. Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Mass Spectrom* 5:976–989.
22. Tabb DL, McDonald WH, Yates JR (2002) DTASelect and Contrast: Tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* 1:21–6.
23. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: The yeast proteome. *J Proteome Res* 2:43–50.
24. Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 4:207–214.
25. Zybailov B, et al. (2006) Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. *J Proteome Res* 5:2339–2347.
26. Liu H, Sadygov RG, Yates JR (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 76:4193–201.
27. DeSantis TZ, Jr, et al. (2006) NAST: A multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res* 34:W394–399.
28. Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14:817–818.
29. Bjursell MK, Martens EC, Gordon JI (2006) Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, B*acteroides thetaiotaomicron*, to the suckling period. *J Biol Chem* 281:36269–36279.
30. Sonnenburg JL, et al. (2005) Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* 307:1955–1959.
31. Martens EC, Chiang HC, Gordon JI (2008) Mucosal glycan foraging enhances the fitness and transmission of a saccharolytic human gut symbiont. *Cell Host Microbe* 4:447–457.

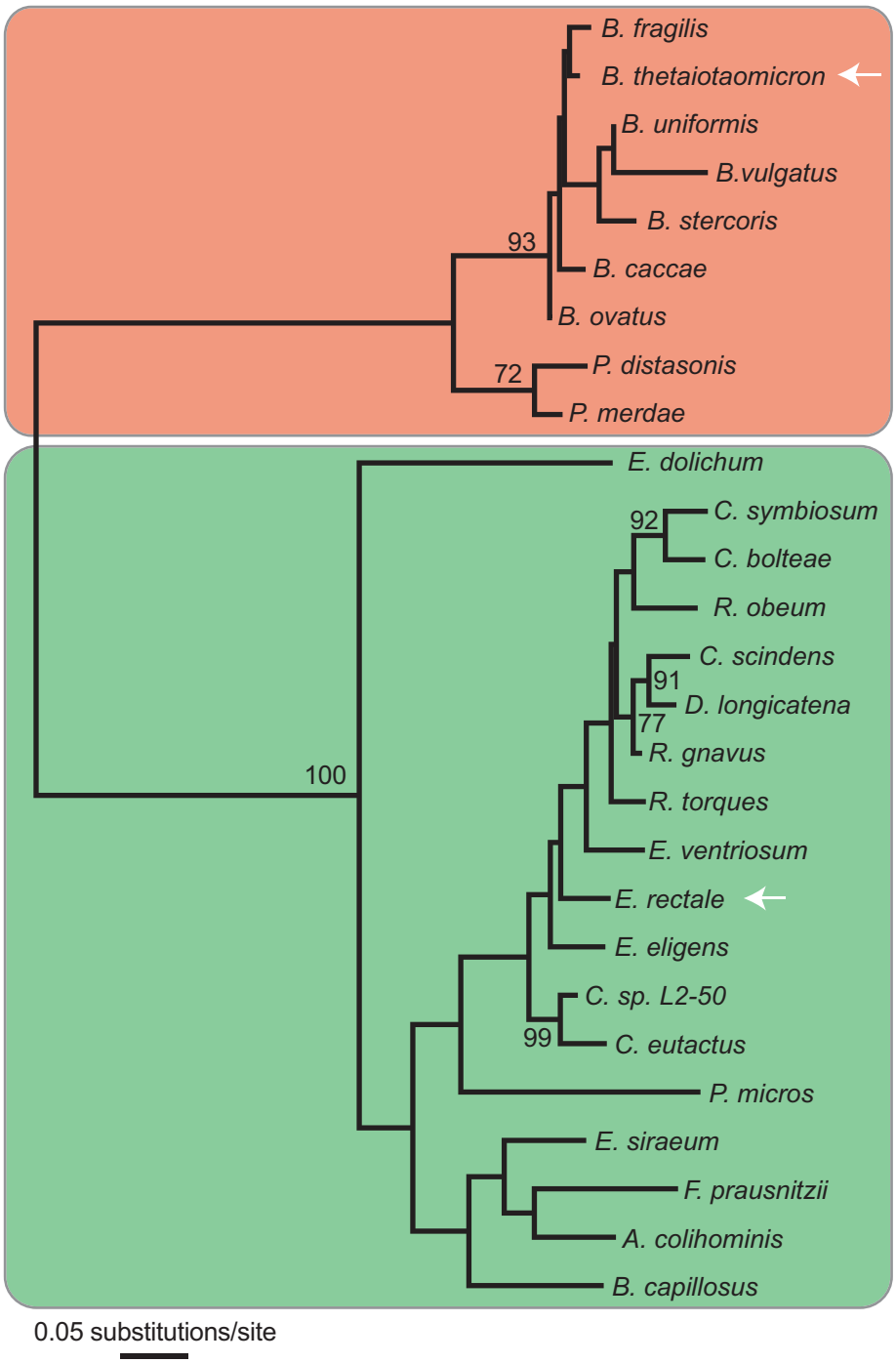


Fig. S1. Phylogenic relationships of human gut-associated Firmicutes and Bacteroidetes surveyed in the present study. A phylogeny based on 16S rRNA gene sequences showing the relationships between representatives from 2 dominant bacterial phyla in the gut microbiota. Green, Firmicutes; red, Bacteroidetes; arrows, organisms used for cocolonization studies described in the present study. 16S rRNA gene sequences were aligned with the NAST aligner (27). Likelihood parameters were determined using Modeltest (28) and a maximum-likelihood tree was generated using PAUP (www.paup.csi.fsu.edu). Bootstrap values represent nodes found in >70 of 100 repetitions.

	GO Category	Description	B. caccae	B. ovatus	B. stercoris	B. uniformis	B. fragilis	B. thetaiotaomicron	B. vulgatus	P. distasonis	P. merdae	C. bolteae	C. scindens	C. sp. L2-50	C. symbiosum	R. gnavus	R. obeum	R. torques	C. eutactus	D. longicatena	A. colihominis	B. capillosus	F. prausnitzii	P. micros	E. dolichum	E. siraeum	E. ventriosum	E. eligens	E. rectale
		Total genes	2093	2677	1857	2181	2313	2674	2296	2234	2060	3561	2099	1500	2602	2002	1993	1470	1529	1722	1972	2112	1557	977	1294	1352	1480	1563	1792
Poly-saccharide metabolism	0030246	carbohydrate binding	25	42	23	35	26	36	30	19	21	35	21	15	22	13	11	15	12	13	12	20	11	5	13	14	21	10	15
	0016798	glycoside hydrolase	97	194	48	124	97	162	116	62	70	51	16	28	22	46	29	29	40	27	13	35	31	4	25	46	30	23	41
	0008484	sulfuric ester hydrolase activity	23	26	13	20	20	32	16	20	15	1	3	1	0	2	2	2	3	1	2	5	1	0	0	0	4	2	3
Energy production	0016651	NAD(P)H oxidoreductase	17	17	16	16	16	17	16	10	9	11	4	3	3	3	4	2	1	2	3	4	1	2	1	2	1	1	1
	0016788	ester hydrolase	98	113	83	89	87	117	84	86	78	70	53	52	67	60	51	54	49	49	51	67	50	40	48	51	54	55	52
	0016655	oxidoreductase	15	15	14	14	14	15	14	8	7	8	2	2	1	1	1	1	0	0	1	0	0	1	0	1	0	0	0
Mobile elements	0000150	recombinase activity	2	4	4	2	4	3	5	5	5	21	21	6	9	16	11	11	7	6	21	36	33	5	7	4	11	2	13
	0004803	transposase activity	12	16	47	13	4	24	33	2	15	22	9	4	13	18	15	14	7	15	7	5	2	2	22	3	2	6	11
Oxygen sensitivity	0016209	antioxidant activity	15	18	8	8	15	11	18	11	11	7	2	2	6	3	4	3	1	5	2	5	1	3	2	0	0	1	2
Environ-mental sensing and regulation	0060089	molecular transducer activity	170	281	123	173	198	253	190	192	155	195	108	72	129	89	89	44	85	64	63	86	47	14	27	55	64	73	95
	0004871	signal transducer activity	170	281	123	173	198	253	190	192	155	195	108	72	129	89	89	44	85	64	63	86	47	14	27	55	64	73	95
	0003711	transcription elongation regulator	6	9	7	12	13	9	6	8	4	4	3	3	9	4	3	4	4	4	3	4	2	2	3	2	3	3	3
	0004673	protein histidine kinase activity	56	93	37	57	61	89	51	68	43	89	52	34	63	39	48	20	40	28	22	34	21	8	14	20	25	22	30
	0000155	two-component sensor activity	52	86	34	53	57	83	45	60	34	86	47	30	56	37	43	17	39	27	20	27	20	7	12	18	24	22	29
	0030528	transcription regulator activity	163	245	119	161	195	238	173	170	151	445	221	127	322	211	190	123	154	165	166	258	141	52	86	114	123	131	164
Transport	0004872	receptor activity	92	163	64	93	105	134	116	94	89	5	3	0	1	0	0	1	3	3	2	0	1	0	1	2	3	2	1
	0008565	protein transporter activity	37	41	25	35	41	43	32	33	33	12	7	5	15	5	7	5	6	5	10	6	8	3	5	8	4	6	10
	0022804	active transmembrane transporter	64	71	55	66	76	74	68	72	59	211	91	82	160	135	105	93	82	123	93	127	93	71	109	75	103	97	90
	0022891	substrate-specific transmembrane transporter	92	106	75	89	100	111	95	99	86	202	116	81	173	123	95	89	79	123	104	117	92	62	114	69	105	92	88
	0015291	secondary active transmembrane transporter	30	36	23	35	37	37	35	40	31	135	43	44	102	66	53	37	44	74	43	54	58	17	60	27	55	51	43
	0015399	primary active transmembrane transporter	35	35	32	31	39	37	33	32	28	73	48	37	58	66	53	54	38	49	50	72	35	54	39	48	47	47	47
	0015144	carbohydrate transmembrane transporter	10	14	7	8	9	18	9	12	9	59	11	10	11	30	9	18	7	20	10	25	20	4	53	7	27	12	14
	0015197	peptide transporter	0	0	0	0	0	0	0	0	0	29	8	0	20	3	1	0	0	0	9	4	3	1	2	0	2	0	2
Motility	0019861	flagellum	0	0	0	2	0	0	0	0	0	21	0	0	17	1	0	0	0	0	23	0	0	1	0	19	0	26	23
			Bacteroidetes									Firmicutes																	

Fig. S2. Genes involved in carbohydrate metabolism and energy production whose representation is significantly enriched or depleted in sequenced human gut-associated Firmicutes and Bacterodetes. The number of genes assigned to each GO term in each genome is shown. Significance is judged by a binomial test, with multiple hypothesis testing correction (see *Methods*) comparing the proportion of genes assigned to a GO term in one genome versus the average number assigned to the same GO term across all of the Firmicutes. The Firmicutes also use distinct mechanisms for environmental sensing and membrane transport. The Bacteroidetes employ a large number of paralogs of the SusC/D system to bind and import sugars (classified as receptors, GO:0004872), while the Firmicutes use ABC transporters and phosphotransferase systems (classified as active membrane transporters, GO:0022804). Red , enriched; blue, depleted; darker color $P < 0.001$ and light color $P < 0.05$ relative to the average of all Firmicute genomes excluding the one tested.

CAZy Family		*B. thetaiotaomicron*	*B. vulgatus*	*E. rectale*	*E. eligens*
GH2	various	32	25	3	2
GH20	hexosaminidase	20	8	0	0
GH43	furanosidase	31	22	2	3
GH92	α-1-2-mannosidase	23	9	0	0
GH76	α-1-6-mannosidase	10	0	0	0
GH97	α-glucosidase	10	7	0	0
GH18	chitinase/ glucosaminidase	12	2	1	1
GH28	galacturonase	9	13	0	3
GH29	α-fucosidase	9	8	0	0
GH1	6-P-ß-glucosidase	0	0	1	1
GH25	lysozyme	1	1	4	5
GH94	phosphorylase	0	0	3	1
PL9	pectate lyase	2	0	0	4
GH8	oligoxylanase	0	0	1	0
GH13	α-amylase	7	4	13	6
GH24	lysozyme	0	1	1	0
GH42	ß-galactosidase	1	1	2	0
GH53	endo-1,4- galactanase	1	0	2	0
GH77	amylomaltase	1	1	3	1
GH112	galacto-N-biose phosphorylase	0	0	1	0
GH10	xylanase	0	1	0	0
GH15	α-glycosidase	0	1	0	0
GH63	α-glucosidase	0	2	0	0
	Total GH	**255**	**167**	**52**	**30**
	Total PL	**17**	**7**	**0**	**7**

Fig. S3. Comparison of glycoside hydrolases and polysaccharide lyases repertoires of *E. rectale, E. eligens, B. vulgatus* and *B. thetaiotaomicron.* The number of genes in each genome in each CAZy GH or PL family are shown. Families that are significantly depleted relative to *B. thetaiotaomicron* are colored blue ($P < 0.001$), as judged by a binomial test followed by Benjamini-Hochberg correction. Families for which *B. thetaiotaomicron* has significantly more members are colored yellow. Families that are absent in *B. thetaiotaomicron* are highlighted in orange. *B. thetaiotaomicron* has a larger genome and a disproportionately larger assortment of GHs. Both Firmicutes have a reduced capacity to use host-derived glycans (hexosaminidases, mannosidases, and fucosidases; GH20, GH29, GH76). *E. rectale* has a large number of starch-degrading enzymes (GH13), while *E. eligens* has a capacity to degrade pectins (PL9, GH28). See Table S2 for a complete list of all CAZy enzymes among the sequenced gut Bacteroidetes and Firmicutes examined.

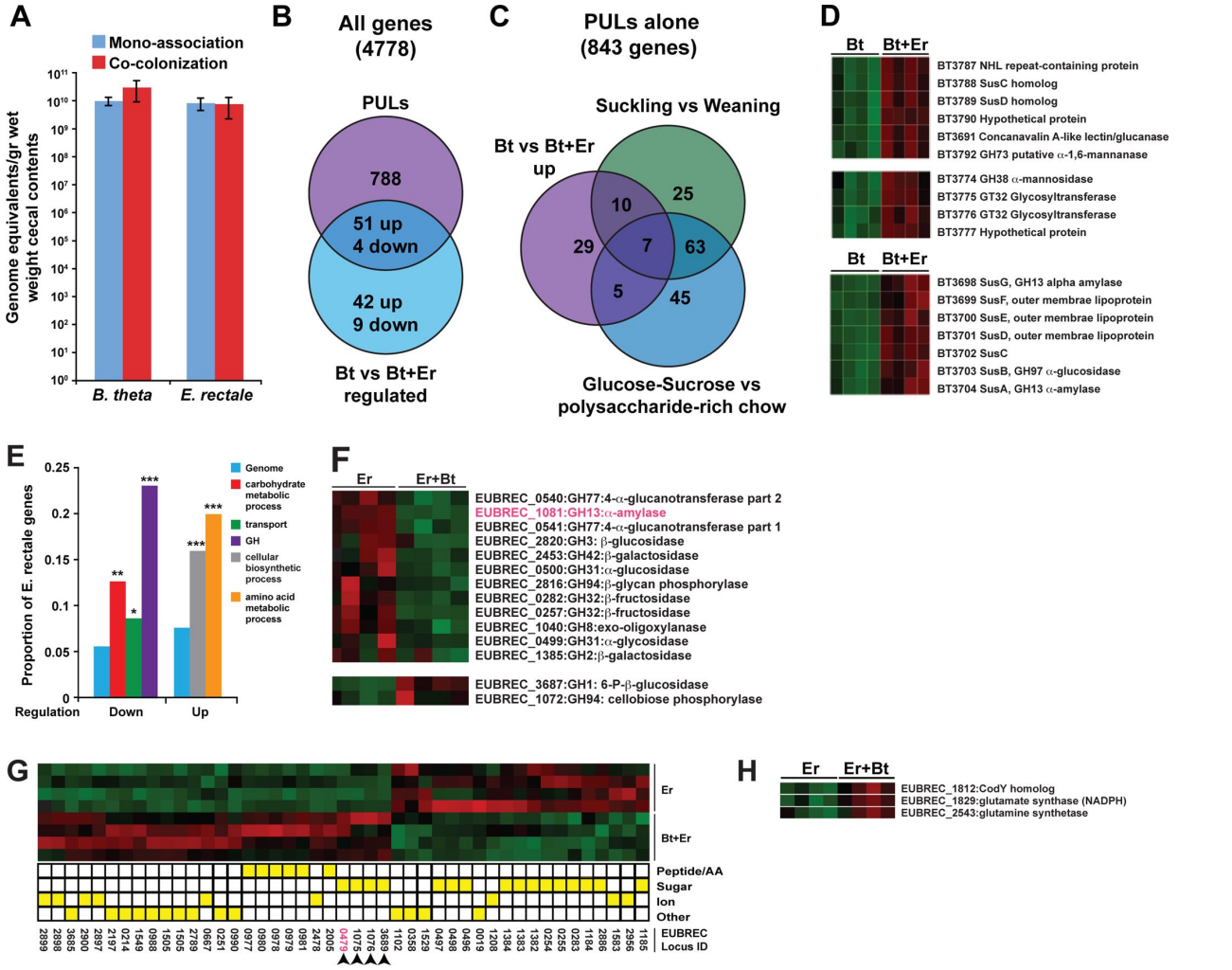


Fig. S4. Creation of a minimal human gut microbiota composed of a sequenced Firmicute (*E. rectale*) and a sequenced Bacteroidetes (*B. thetaiotaomicron*). (*A*) Levels of colonization of the ceca of 11-week-old male gnotobiotic mice colonized for 14 days with one or both organisms. Animals were given an irradiated low-fat plant polysaccharide-rich (LF/PP) chow diet ad libitum. *B. thetaiotaomicron* and *E. rectale* colonize the ceca of mice to similar levels in both mono- and biassociation. Error bars denote standard error of the mean of 2–3 measurements per mouse, 4 mice per group. Results are representative of 3 independent experiments. (*B*) Summary of genes showing up-regulation in *B. thetaiotaomicron* with cocolonization. 55 of the 106 genes are within PULs, and of these, 51 (93%) were up-regulated. (*C*) Summary of *B. thetaiotaomicron* PUL-associated genes up-regulated with cocolonization and their representation in datasets of genes up-regulated during the suckling-weaning transition (29), and when adult gnotobiotic mice are switched from a polysaccharide-rich diet to one devoid of complex glycans and containing simple sugars (glucose, sucrose (30)). The latter 2 datasets are composed of genes that are also up-regulated ≥10-fold relative to log-phase growth in minimal glucose medium (30). (*D*) Heat map of GeneChip data from 3 loci up-regulated by *B. thetaiotaomicron* upon colonization with *E. rectale*; 2 are involved in degradation of α-mannans that *E. rectale* cannot access; the third is the Starch utilization system (Sus) locus, which targets a substrate that both species can use. Maximal relative expression across a row is red; minimal is green. Differential expression was judged using the MAS5 algorithm and CyberT (see Table S4A and *Methods*). (*E*) Overview of the response of *E. rectale* to cocolonization with *B. thetaiotaomicron*. Genes assigned to GO terms for carbohydrate metabolism (GO:0005975), transporters (GO: GO:0006810) and predicted GHs are all significantly over-represented among down-regulated genes while genes with GO terms for biosynthesis (GO:0044249), in particular amino acid metabolism (GO:0006520), are significantly over-represented among up-regulated genes. All categories shown are significantly different from the genome as a whole. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ (binomial test). (*F*) Heat map from GeneChip data of all significantly regulated *E. rectale* GH genes showing that all but 2 are downregulated (both cytoplasmic phosphosugar processing enzymes) n = 4 mice/treatment group. (*G*) Heat map of all significantly regulated *E. rectale* genes assigned to the GO term for transporters (GO:0006810) showing that a number of simple sugar transporters are downregulated upon cocolonization, while peptide and amino acid transporters and 3 predicted simple sugar transporters (arrows; EUBREC_0479, a galactoside ABC transporter; EUBREC_1075–6, a lactose/arabinose transport system; and EUBREC_3689, a predicted cellobiose transporter) are up-regulated. (*H*) Heat map of selected global regulators from *E. rectale* shows that CodY, a repressor in other Firmicutes of stationary-phase genes such as those needed to access lower-energy carbon sources, is significantly up-regulated upon cocolonization, suggesting increased accessibility of nitrogen and/or carbon sources. Glutamine synthetase and glutamate synthase, are also up-regulated, consistent with the observed up-regulation of various amino acid and peptide transporters. Differentially regulated genes were identified using the MAS5 algorithm and Cyber-T (see Table S4B and *Methods*). Genes whose differential expression with cocolonization was further validated by qRT-PCR are highlighted with pink lettering (2 independent experiments, *n* = 4–5 mice per group, 2–3 measurements per gene; see Fig. 1).

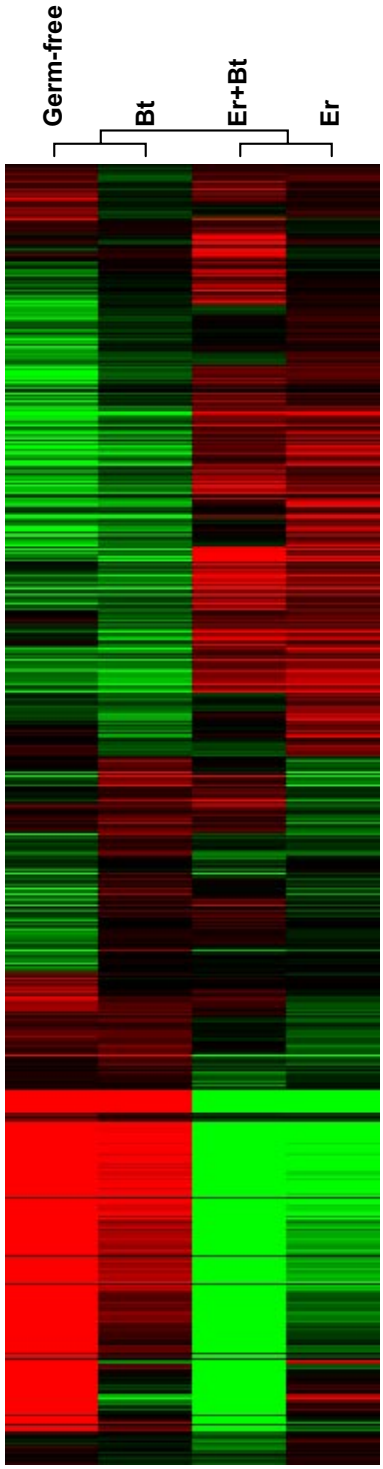


Fig. S5. Unsupervised hierarchical clustering of average GeneChip expression intensity values from probesets representing differentially expressed genes (>1.5-fold; <1% FDR) in the proximal colons of germ-free versus cocolonized animals. Clustering was performed using host expression data from all treatment groups: germ-free, *B. thetaiotaomicron* and *E. rectale* monoassociations, and *B. thetaiotaomicron-E. rectale* biassociation. Data from all GeneChips in a given treatment group were averaged ($n = 4$ animals/group; total of 694 probesets analyzed). Clustering was performed by cluster 3.0 software (www.bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/software.htm#ctv) and data were visualized using Tree View software. Each probeset is represented by a single row of colored bars. Maximal relative expression across a row is red; minimal is green.

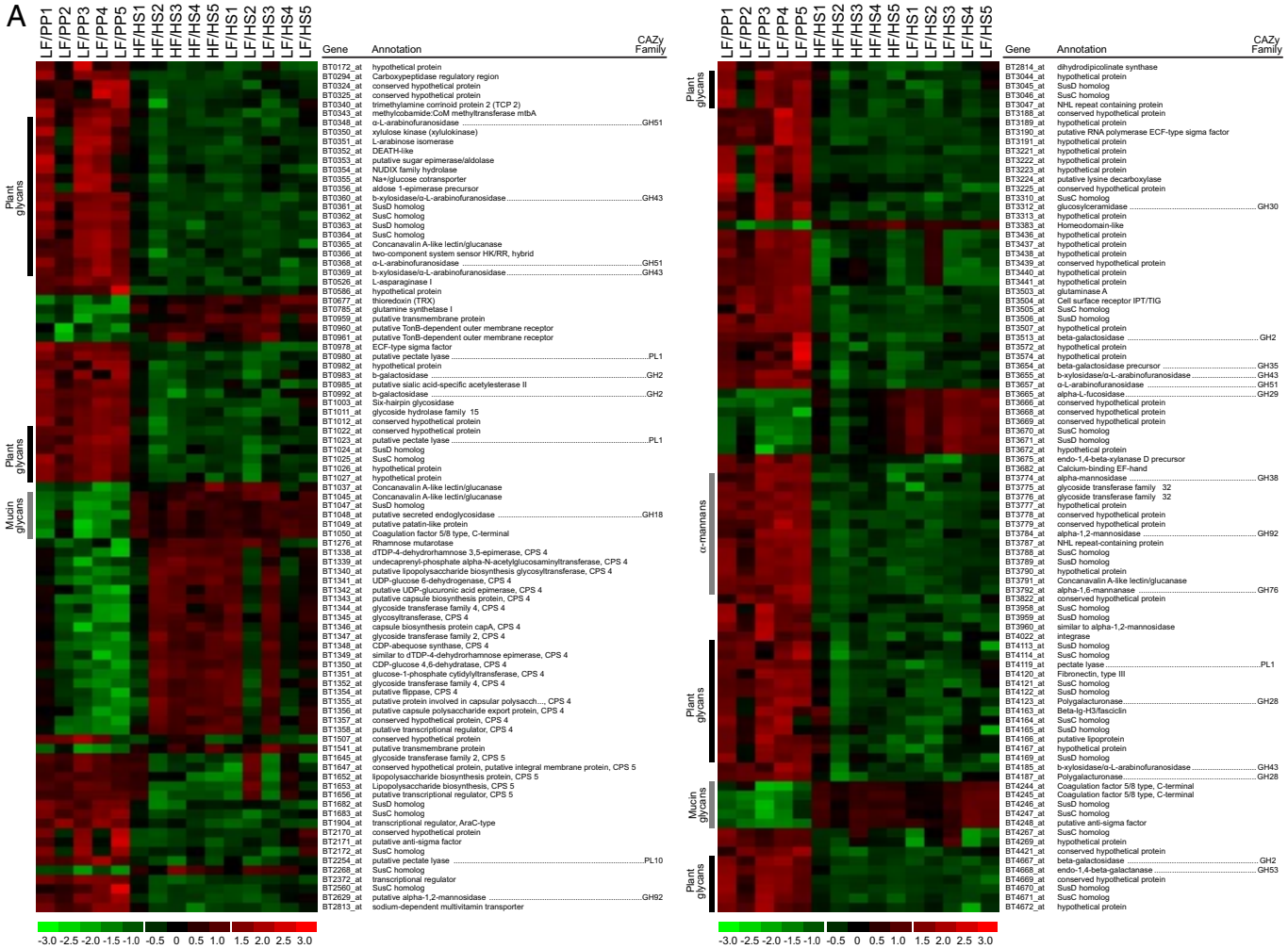
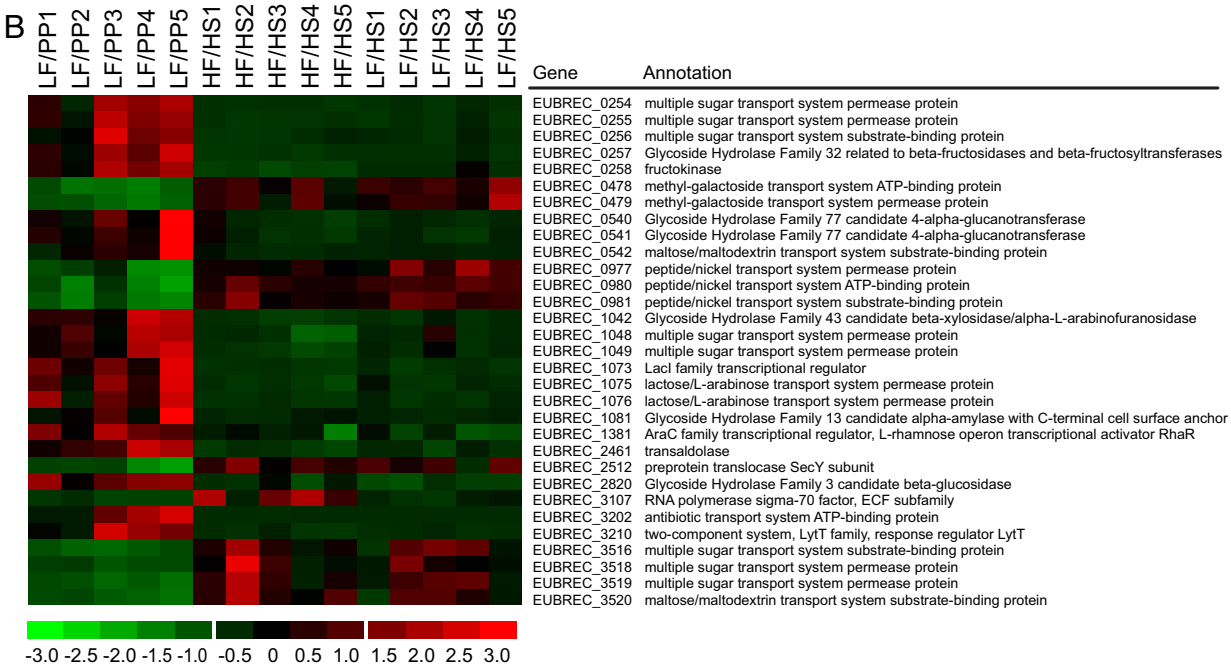


Fig. S6. Effects of low carbohydrate content diets on a *B. thetaiotaomicron–E. rectale* community gene expression. (*A*) Heat map of GeneChip data from *B. thetaiotaomicron* in the ceca of mice (cocolonized with *E. rectale*; $n = 4$ mice per group) that were either fed (*i*) a standard low-fat plant polysaccharide-rich diet (LF/PP), (*ii*) a high-fat, high sugar (HF/HS) ''Western style'' diet containing cellulose as the only complex plant polysaccharide or (*iii*) a control for diet (*iii*) that contained 4-fold less fat, high levels of simple sugars plus cellulose (LF/HS). Polysaccharide utilization loci (PULs) whose specificities are known are indicated (31). Genes predicted to be regulated by complex plant polysaccharides are also highlighted. (*B*) Heat map of GeneChip data from *E. rectale* in the ceca of mice cocolonized with *B. thetaiotaomicron* fed as described in *A*. All genes were defined as significantly differentially expressed (>2 fold, PPDE >0.95). Genes encoding hypothetical proteins were not included in the heat map. Differential expression was judged using the MAS5 algorithm and Cyber-t. Maximal relative expression across a row is red; minimal is green.



Gene	Annotation
EUBREC_0254	multiple sugar transport system permease protein
EUBREC_0255	multiple sugar transport system permease protein
EUBREC_0256	multiple sugar transport system substrate-binding protein
EUBREC_0257	Glycoside Hydrolase Family 32 related to beta-fructosidases and beta-fructosyltransferases
EUBREC_0258	fructokinase
EUBREC_0478	methyl-galactoside transport system ATP-binding protein
EUBREC_0479	methyl-galactoside transport system permease protein
EUBREC_0540	Glycoside Hydrolase Family 77 candidate 4-alpha-glucanotransferase
EUBREC_0541	Glycoside Hydrolase Family 77 candidate 4-alpha-glucanotransferase
EUBREC_0542	maltose/maltodextrin transport system substrate-binding protein
EUBREC_0977	peptide/nickel transport system permease protein
EUBREC_0980	peptide/nickel transport system ATP-binding protein
EUBREC_0981	peptide/nickel transport system substrate-binding protein
EUBREC_1042	Glycoside Hydrolase Family 43 candidate beta-xylosidase/alpha-L-arabinofuranosidase
EUBREC_1048	multiple sugar transport system permease protein
EUBREC_1049	multiple sugar transport system permease protein
EUBREC_1073	LacI family transcriptional regulator
EUBREC_1075	lactose/L-arabinose transport system permease protein
EUBREC_1076	lactose/L-arabinose transport system permease protein
EUBREC_1081	Glycoside Hydrolase Family 13 candidate alpha-amylase with C-terminal cell surface anchor
EUBREC_1381	AraC family transcriptional regulator, L-rhamnose operon transcriptional activator RhaR
EUBREC_2461	transaldolase
EUBREC_2512	preprotein translocase SecY subunit
EUBREC_2820	Glycoside Hydrolase Family 3 candidate beta-glucosidase
EUBREC_3107	RNA polymerase sigma-70 factor, ECF subfamily
EUBREC_3202	antibiotic transport system ATP-binding protein
EUBREC_3210	two-component system, LytT family, response regulator LytT
EUBREC_3516	multiple sugar transport system substrate-binding protein
EUBREC_3518	multiple sugar transport system permease protein
EUBREC_3519	multiple sugar transport system permease protein
EUBREC_3520	maltose/maltodextrin transport system substrate-binding protein

Fig. S6. continued

Other Supporting Information Files

Table S1
Table S2
Table S3
Table S4
Table S5
Table S6
Table S7
Table S8