SUPPLEMENTAL DATA

EXPERIMENTAL PROCEDURES

Preparation of DNA from the cecal microbiota – Cecal contents were frozen at -80^{o}C immediately after sacrifice. An aliquot (~10mg) of each sample was then suspended, while frozen, in a solution containing 500 μl of extraction buffer [200 mM Tris (pH 8.0), 200 mM NaCl, 20 mM EDTA], 210 μl of 20% SDS, 500 μl of a mixture of phenol:chloroform:isoamyl alcohol (pH 7.9, 25:24:1), and 500 μl of a slurry of 0.1 mm-diameter zirconia/silica beads (BioSpec Products, Bartlesville, OK). Microbial cells were subsequently lysed by mechanical disruption with a bead beater (BioSpec Products) set on high for 2 min at RT, followed by extraction with phenol:chloroform:isoamyl alcohol (pH 7.9, 25:24:1), and precipitation with isopropanol. DNA obtained from ten separate 10 mg frozen aliquots of each cecal sample were pooled (≥200μg DNA) and used to construct plasmid libraries (pOTw13) for 3730xl capillary-based metagenomic sequencing (see below).

16S rRNA sequence-based surveys of the distal gut (cecal) mouse microbiota - Five replicate PCR reactions were performed for each cecal DNA sample. Each 25 μl reaction contained 50-100 ng of purified DNA, 10 mM Tris (pH 8.3), 50 mM KCl, 2 mM $MgSO_4$, 0.16 μM dNTPs, 0.4 μM of the bacteria-specific primer 8F (5'-AGAGTTTGATCCTGGCTCAG-3'), 0.4 μM of the universal primer 1391R (5'-GACGGGCGGTGWGTRCA-3'), 0.4 M betaine, and 3 units of Taq polymerase (Invitrogen). Cycling conditions were 94°C for 2 min, followed by 35 cycles of 94°C for 1 min, 55°C for 45 sec, and 72°C for 2 min, with a final extension period of 20 min at

72°C. Replicate PCRs were pooled and concentrated (Millipore; Montage PCR filter columns). Full-length 16S rRNA gene amplicons (1.3kb) were then gel-purified using the Qiaquick kit (Qiagen), subcloned into TOPO TA pCR4.0 (Invitrogen), and the ligated DNA transformed into *E. coli* TOP10 (Invitrogen). For each mouse, 384 colonies containing cloned amplicons were processed for sequencing. Plasmid inserts were sequenced bi-directionally using vector-specific primers plus the internal primer 907R (5'-CCGTCAATTCCTTTRAGTTT-3').

16S rRNA gene sequences were edited and assembled into consensus sequences using the PHRED and PHRAP software packages within the Xplorseq program (Papineau *et al.*, 2006). Sequences that did not assemble were discarded and bases with PHRED quality scores <20 were trimmed. Sequences were checked for chimeras using Bellerophon version 2 (Huber, 2004) and sequences with greater than 95% identity to both parents were removed (n=535; 13% of aligned sequences). The final dataset (n=8,511 16S rRNA gene sequences; for sequence designations see **Table S3**) was aligned using the on-line version of the NAST multi-aligner [DeSantis *et al.*, 2006; minimum alignment length=1250 nucleotides (500 for *Rag1-/-* data); percent identity >75]. Hypervariable regions were masked using the lanemaskPH filter provided within the ARB database (Ludwig *et al.*, 2004), and the aligned sequences added to the ARB neighbor-joining tree (based on pairwise distances with the Olsen correction), using the parsimony insertion tool. A phylogenetic tree containing all 16S rRNA gene sequences was then exported from ARB, clustered using online UniFrac (Lozupone *et al.*, 2006) without abundance weighting, and visualized with TreeView (Page, 1996). A distance matrix of all 16S rRNA gene sequences was imported into DOTUR (Schloss and

Handlesman, 2005) for phylotype binning and measurements of diversity (e.g., the Shannon index).

Taxonomic assignment of shotgun sequencing reads – Quality-trimmed reads were assigned to reference genomes by comparison with the NCBI non-redundant database (NR version 4/19/07; BLASTX e-value < 10^{-5}; BLASTX parameters '–F F'). Sequences were assigned to the taxonomic group (division, class, genus, etc.) that would include all significant hits using MEGAN (under the default parameters, only reads with a BLAST score ≥ 10% of the top score were included; see Huson *et al.*, 2007). Reads containing a 16S rRNA fragment were identified by BLASTN comparison of each microbiome to the RDP database (version 9.33; Cole *et al.*, 2005). 16S rRNA gene fragments were then aligned using the NASTA multi-aligner (DeSantis *et al.*, 2006) with a minimum template length of 400 bases and a minimum percent identity of 75%. The resulting alignment was then imported into an ARB neighbor-joining tree and hypervariable regions masked using the lanemaskPH filter (Ludwig *et al.*, 2004).

Transcriptional profiling - A 10mg aliquot of frozen cecal contents from a mouse fed the Western diet (sample 'Western 3') was immersed in 1ml of RNAProtect (Qiagen), vortexed, centrifuged for 10 min at 5000 x g, and the supernatant was removed. Microbial cells in the pellet were subsequently lysed by mechanical disruption with a bead beater (BioSpec Products) set on high for 2 min at RT in a solution containing 500 μl of extraction buffer. RNA was extracted with phenol:chloroform:isoamyl alcohol (pH 4.5, 125:24:1), precipitated with isopropanol, and further purified with (i) the RNeasy Mini Kit (Qiagen), (ii) on-column digestion with DNAseI (Qiagen), (iii) an additional

DNAse treatment (DNAfree kit, Ambion), and (iv) passage through a RNeasy column (Qiagen).

A modification of the protocol included with the MessageAmpII-bacteria Kit (Ambion) that was developed at MIT (Frias-Lopez *et al.*, in press), was used for mRNA-enriched cDNA synthesis. cDNA was purified (Qiaquick, Qiagen) and subcloned into pSMART (10G Supreme Cells, Lucigen). Plasmid inserts from 384 randomly picked colonies were sequenced (single unidirectional reads) using vector specific primers and an ABI 3730xl instrument. Sequences were trimmed based on quality score and to remove vector sequences (Applied Biosystems; KB Basecaller), and to remove poly(A) tails. Only sequences with a final length ≥180 bases were analyzed (average of 430 nucleotides). Sequences were annotated based on BLASTX (see above) and BLASTN comparisons against the NCBI nucleotide database (version 9/26/07; BLASTN parameters '–F F'). 16S rRNA sequences were annotated based on their best-BLAST hit to 16S rRNA genes of known taxonomic origin (e-value < 10^{-25}). Although rRNA gene fragments were the dominant sequence (90.6% of the high-quality reads), the library had a lower abundance of rRNA transcripts than comparable libraries created directly from total distal gut community RNA (99%; P.J. Turnbaugh and J.I. Gordon, unpublished data).

Whole genome sequencing and annotation - A draft assembly of the *Eubacterium dolichum* strain ATCC29143 genome was generated from ABI 3730xl paired-end reads of inserts in whole genome shotgun plasmid libraries (35,683 reads; average read length of 569 nucleotides, representing ~9X coverage), as well as from

reads produced from one run of the 454 FLX pyrosequencer (425,423 reads with mean length of 250 nucleotides, representing ~49X coverage).

The Newbler *de novo* shotgun sequence assembler was used to assemble 454 FLX sequences based on flowgram signal space. This process includes overlap generation, contig layout, and consensus generation. The resulting contigs were then broken into linked sequences to generate pseudo paired-end reads, and aligned with 3730xl reads using PCAP (Huang *et al.*, 2003). To minimize potential assembly/contamination errors in the draft genomes, only contigs greater than 2kb were used. Genes were predicted using MetaGene (Noguchi *et al.*, 2006). Each predicted gene sequence was translated, and the resulting protein sequence assigned InterPro numbers using InterProScan (version 4.3; Mulder *et al.*, 2005). Each gene was annotated based on the output of InterProScan and BLASTP comparisons versus the KEGG database (version 40; Kanehisa *et al.*, 2004) and the STRING database (version 7; von Mering *et al.*, 2007), in addition to experimentally validated metabolic pathway maps in the MetaCyc database (http://metacyc.org; Caspi *et al.*, 2006).

For KEGG pathway analysis, the relative abundance each pathway was calculated for each genome (number of genes assigned to a given pathway divided by the total number of pathway assignments). The relative abundance was then converted into a z-score based on the mean and standard deviation of the given pathway across all microbiomes. KEGG pathways were clustered using Cluster3.0 (deHoon *et al.*, 2004). Single linkage hiearchical clustering via Euclidean distance was performed, and the results visualized (Treeview Java applet; Saldanha, 2004).

Analysis of end-products of bacterial fermentation – Short-chain fatty acids (SCFAs) were measured in cecal samples obtained from mice fed Western, FAT-R, or CARB-R diets (n=3-5 mice/group; two aliquots per mouse). The procedure, described in an earlier publication (Samuel and Gordon, 2006), involved double diethyl ether extraction of deproteinized cecal contents spiked with isotope-labeled internal SCFA standards (Isotec: [$^{2}H_3$]- and [2-^{13}C]acetate, [$^{2}H_5$]propionate, and [$^{13}C_4$]butyrate), derivatization of SCFAs with N-tert-butyldimethylsilyl-N-methyltrifluoracetamide (MTSTFA), and GC-MS analysis of the resulting TBDMS-derivatives using a gas chromatograph (Model 6890; Hewlett–Packard) interfaced to a mass spectrometer detector (Model 5973; Agilent Technologies).

Lactate levels were quantified using a microanalytic approach: cecal samples were quick frozen in liquid nitrogen, stored at -80°C, and lyophilized at -35°C. 1-5 mg of dried cecal material was homogenized in 0.4 ml 0.2 M NaOH at 1°C. Alkali extracts were prepared by heating an 80 μl aliquot for 20 min at 80°C and adding 80 μl of 0.25 M HCl and 100 mM Tris base. Acid extracts were prepared by adding 20 μl 0.7 M HCl to a separate 60 μl aliquot, heating for 20 min at 80°C, and neutralizing with 40 μl of 100 mM Tris base. The Bradford method was used to determine the protein content of the alkali extracts (BioRad). Cecal lactate levels were determined using a combination of pyridine nucleotide coupled enzymatic reactions with the Lowry oil well technique and enzymatic cycling amplification (Passoneau and Lowry, 1993). A 0.2μl aliquot (25-100ng protein) from the acid extracts was added to 2μl of reagent containing 50mM 2-Amino-2-methanol-1-proponal buffer pH 9.9, 2mM glutamate pH 9.9, 0.2mM NAD+, 50ug/ml beef heart lactate dehydrogenase (Sigma; specific activity 500 units/mg protein) and

50μg/ml pig heart glutamate pyruvate transaminase (Roche; spec. act. 80units/mg protein). Following a 30 min incubation at 24°C the reaction was terminated with the addition of 1μl 0.15M NaOH and heated 20 min at 80°C. Once the samples cooled to 24°C, a 1μl aliquot was transferred to 0.1ml NAD cycling reagent and amplified 5000 fold. Lactate standards, 5 to 10μM, were carried throughout all steps.

Statistical methods – χ^2 tests were performed on the number of gene assignments to a given KEGG or STRING orthologous group in each microbiome relative to the number of gene assignments to all other groups. Xipe (Rodriguez-Brito, 2006; version 2.4) was employed for bootstrap analyses of KEGG pathway enrichment and depletion, as described previously (Turnbaugh *et al.*, 2006), using the parameters sample size = 10,000 and confidence level = 0.90. ANOVA was performed using a model comparison approach (Judd and McClelland, 1989), implemented with the linear regression function in Excel (version 11.0, Microsoft). Student's t-tests were utilized to identify statistically significant differences between two groups. Data are represented as mean±SEM unless otherwise indicated. The p-value associated with a given correlation coefficient (R^2) was generated by a permutation analysis, as described previously (Ley *et al.*, 2006b). Briefly, the values were scrambled randomly and an R^2 generated 10,000 times; the resulting distribution of R^2 values was used to assess the probability of obtaining the observed R^2.

SUPPLEMENTAL REFERENCES

Caspi, R. *et al.* (2006). MetaCyc: A multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res. *34*, D511-D516.

de Hoon, M.J., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. Bioinformatics *20*, 1453-1454.

DeSantis, T. Z., Hugenholtz, P., Keller, K., Brodie, E.L., Larsen, N., Piceno, Y.M., Phan, R., and Andersen, G.L. (2006). NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic Acids Res. *34*, W394-399.

Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., and DeLong, E.F. (2007). Measuring microbial community gene expression in ocean surface waters. Proc. Natl. Acad. Sci. USA, in press.

Hooper S.D., and Bork, P. (2005). Medusa: a simple tool for interaction graph analysis. Bioinformatics. *21*, 4432-4433.

Huang, X., Wang, J., Aluru, S., Yang, S.P., and Hillier, L. (2003). PCAP: a whole-genome assembly program. Genome Res. *13*, 2164-2170.

Huber, T., Faulkner, G., and Hugenholtz, P. (2004). Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics *20*, 2317-2319.

Ludwig, W. *et al.* (2004). ARB: a software environment for sequence data. Nucleic Acids Res. *32*, 1363-1371.

Page, R.D. (1996). TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. *12*, 357-358.

Papineau, D., Walker, J.J., Mojzsis, S.J., and Pace, N.R. (2005). Composition and structure of microbial communities from stromatolites of Hamelin Pool in Shark Bay, Western Australia. Appl. Environ. Microbiol. *71*, 4822-4832.

Passonneau, J.V. and Lowry, O.H. (1993). Enzymatic Analysis: A Practical Guide (Totawa, NJ: Humana Press).

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics *20*, 3246-3248.

Samuel, B.S. and Gordon, J.I. (2006). A humanized gnotobiotic mouse model of host-archael-bacterial mutualism. Proc. Natl. Acad. Sci. USA *103*, 10011-10016.

Sonnenburg, J.L., Chen, C.T., and Gordon, J.I. (2006). Genomic and metabolic studies of the impact of probiotics on a model gut symbiont and host. PLoS Biol. *4*, e413.

SUPPLEMENTAL FIGURE LEGENDS

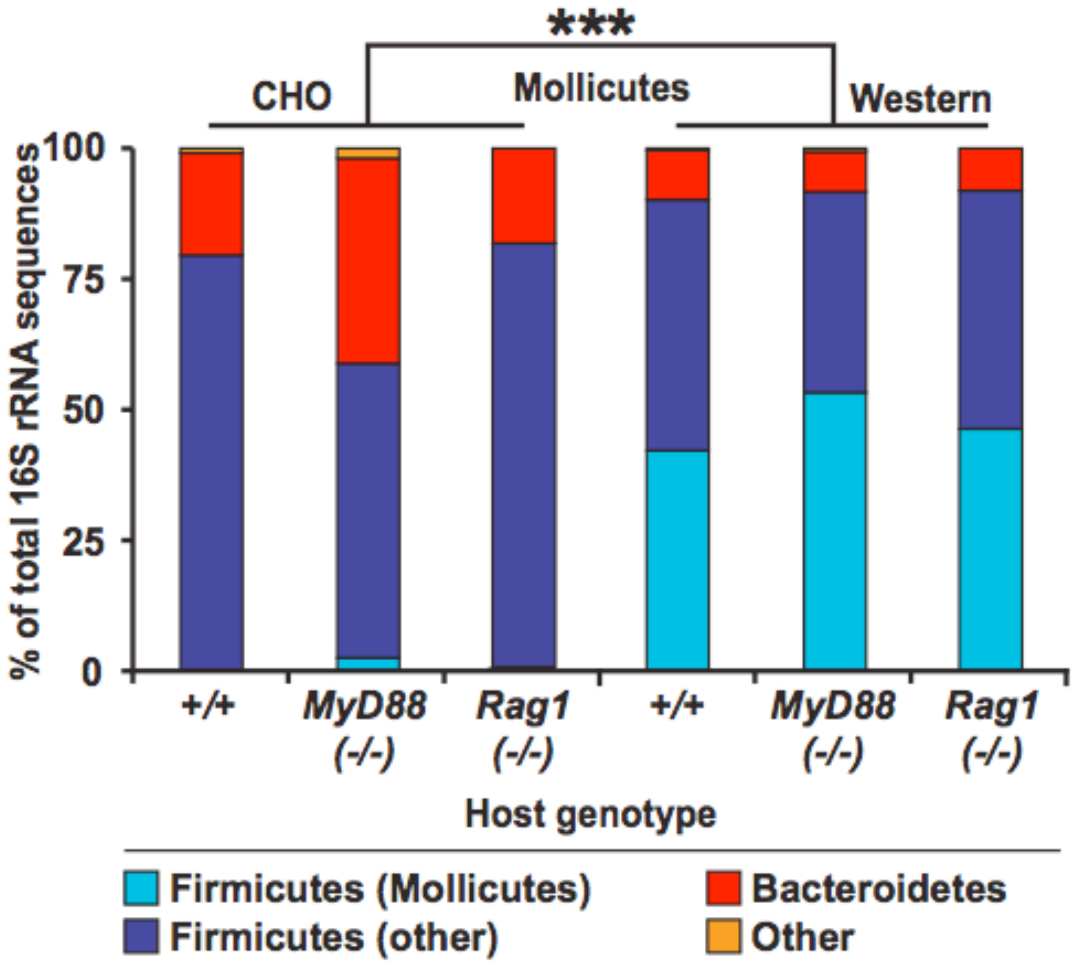

Figure S1 – The Mollicute bloom occurs in conventionally-raised wild-type C57BL/6J mice as well as in mice without an intact innate or adaptive immune system. Wild-type (+/+), *MyD88 -/-*, or *Rag1 -/-* C57BL/6J mice were weaned onto a standard low-fat polysaccharide-rich (CHO) or high-fat/high-sugar (Western) diet. 16S rRNA gene sequence-based surveys were performed; sequences were aligned (NAST; DeSantis *et al.*, 2006), and inserted into an ARB neighbor-joining tree (Ludwig *et al.*, 2004). Asterisks indicate significant differences (Student's t-test $p<0.001$).

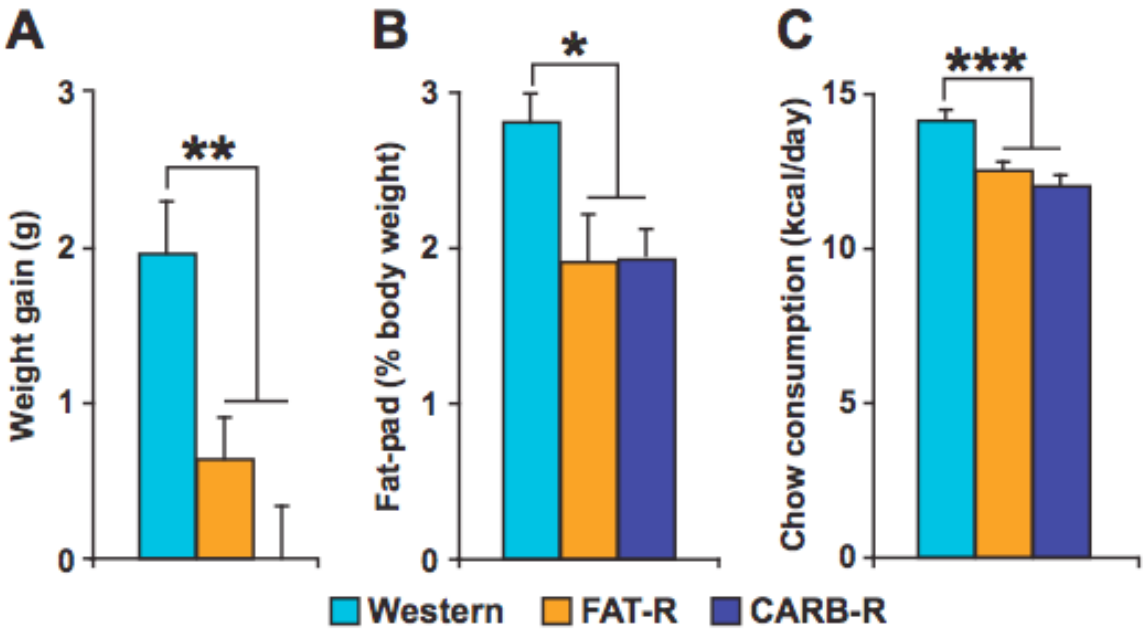

Figure S2 – Mice with diet-induced obesity that are switched to a FAT-R or CARB-R diet exhibit stabilization of weight, decreased caloric intake and reduced adiposity. (A) Weight gain (g) and **(B)** percentage epididymal fat-pad weight to body weight in wild type C57BL/6J mice that were initially weaned onto a Western diet for 8 weeks, and then maintained on the Western diet, or switched to a FAT-R or CARB-R diet for four weeks (n=5-6 mice/treatment group). Weight was monitored during the four-week period. **(C)** Chow consumption (kcal/d) is decreased in mice switched to a FAT-R or CARB-R diet. Data are represented as mean±SEM. Asterisks indicate significant differences (ANOVA of FAT-R or CARB-R versus Western, *$p<0.05$, **$p<0.01$, ***$p<0.0001$).

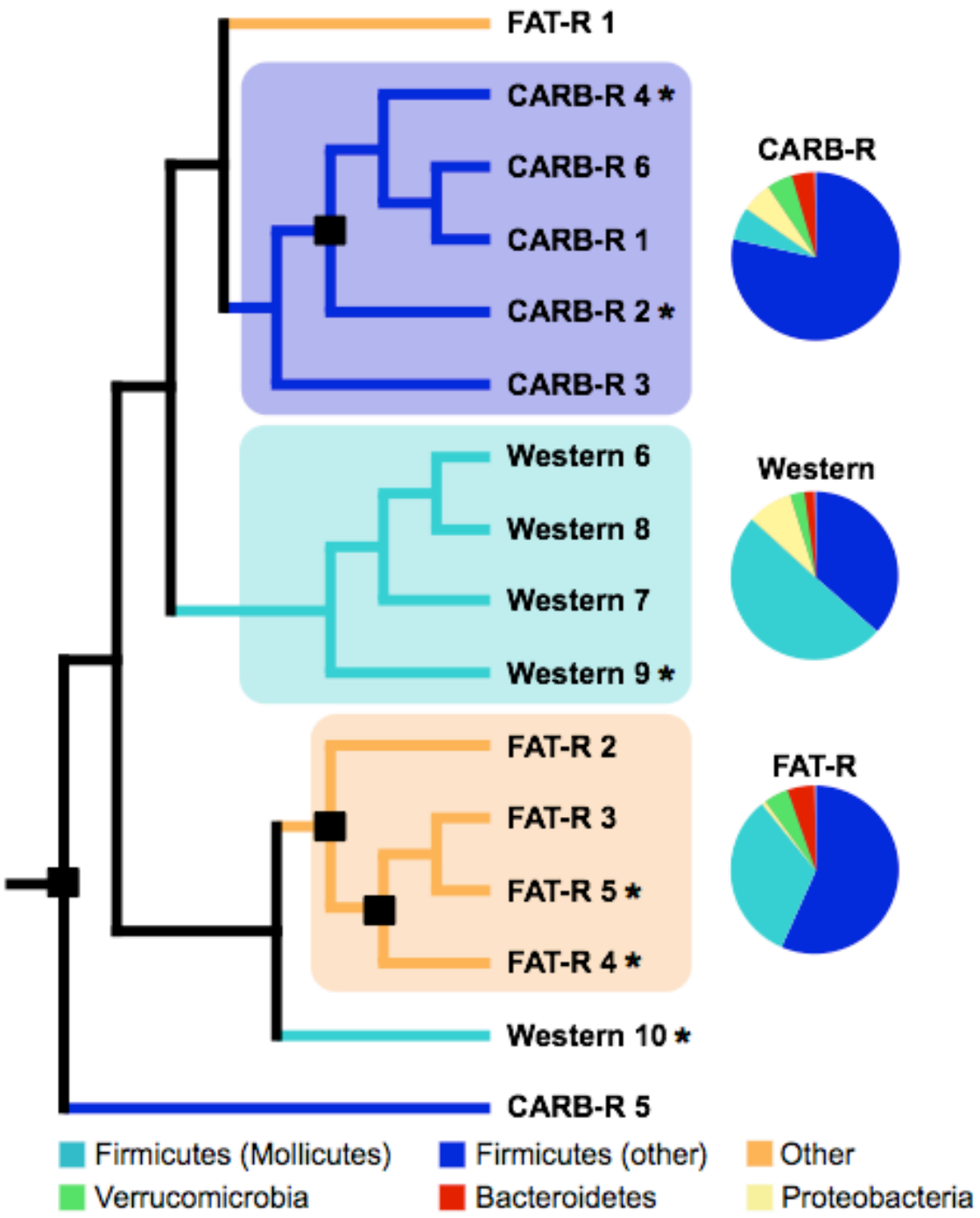

Figure S3 – Switching from a Western to FAT-R or CARB-R diet results in a division-wide increase in the relative abundance of Bacteroidetes, and a decrease in the relative abundance of Mollicutes. UniFrac-based analysis of bacterial community membership shows an impact of diet on gut microbial ecology: cecal communities

analyzed from two families of C57BL/6J wild-type mice (**Table S3**) generally cluster based on host diet (Western, FAT-R, and CARB-R). The average relative abundance (% of total 16S rRNA gene sequences) of bacterial lineages within the cecal microbiota of all mice fed a Western, FAT-R, or CARB-R diet is displayed as pie charts. Black boxes indicate nodes that were reproduced in >50% of all jackknife replications (n=126 sequences were randomly re-sampled). Asterisks indicate cecal samples that were analyzed by whole community shotgun sequencing.

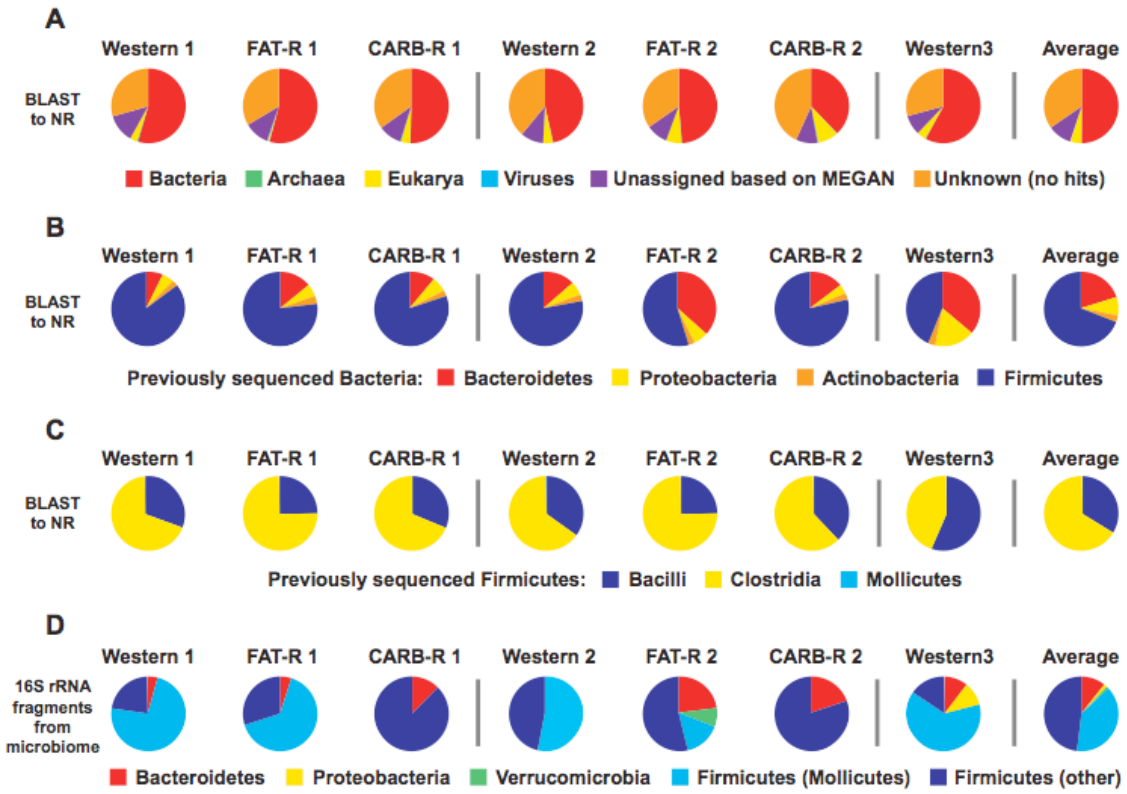

Figure S4 – Taxonomic assignments of metagenomic sequencing reads from seven cecal microbiome datasets based on BLAST homology searches, and by alignment of 16S rRNA gene fragments. (A) The cecal microbiome is dominated by sequences homologous to Bacteria. Sequencing reads were trimmed based on quality and vector sequence and the resulting datasets were used as queries against the NCBI non-redundant database (e-value<10^{-5}). Sequences were assigned to the lowest taxonomic group that would include all significant hits, using MEGAN (Huson *et al.*, 2007). Pie charts are shown for each individual dataset and for the average of all datasets. Colors indicate assignments to bacteria (red), archaea (green), eukarya (yellow), viruses (blue), sequences that could not be confidently assigned to a group (purple), and sequences with no significant BLASTX matches (orange). **(B)** Relative abundance of microbiome sequences homologous to genomes from four bacterial divisions: Bacteroidetes (red),

Proteobacteria (yellow), Actinobacteria (orange), and Firmicutes (blue). All divisions observed at >1% relative abundance are shown. **(C)** Relative abundance of microbiome sequences homologous to genomes from bacterial classes within the Firmicutes division: Bacilli (dark blue), Clostridia (yellow), and Mollicutes (light blue). **(D)** Taxonomic assignments of 16S rRNA gene fragments obtained from cecal microbiome datasets. 16S rRNA gene fragments were identified by querying the Ribosomal Database Project (RDP) database (version 9.33; BLASTN e-value < 10^{-5}; Cole *et al.*, 2005). 16S rRNA gene fragments were aligned with NAST (DeSantis *et al.*, 2006) and added to an ARB neighbor-joining tree (Ludwig *et al.*, 2004). 16S rRNA gene fragments from the Bacteroidetes (red), Proteobacteria (yellow), Verrucomicrobia (green), Mollicutes (light blue), and other Firmicutes (dark blue) are shown.

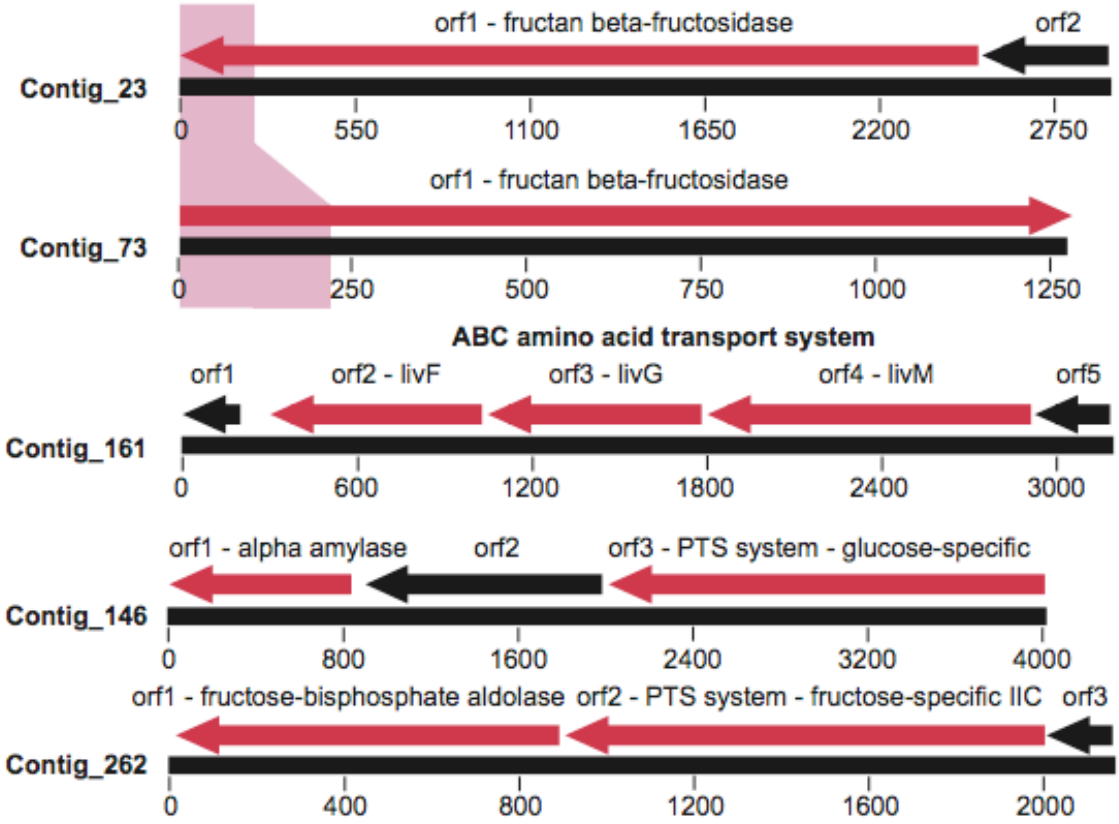

Figure S5 - Assembly of metagenomic sequence data reveals physical linkage between the Mollicute phosphotransferase system (PTS) and other genes involved in carbohydrate metabolism. The pooled mouse gut microbiome dataset was assembled using ARACHNE (n=7 combined datasets; Batzoglou *et al.*, 2002; see **Tables S6** and **S7** for assembly statistics). The contig length is shown as a solid black bar. Arrows indicate predicted proteins. Functional assignments were derived from the NCBI annotations and verified by BLASTP comparisons of each predicted protein with the STRING-extended COG database (von Mering *et al.*, 2007) and the KEGG database (Kanehisa *et al.*, 2004), in addition to Hidden Markov Model (HMM)-based protein domain searching with InterProScan (Mulder *et al.*, 2005). Contigs 23 and 73 are >98% identical over the region

in pink (234/238 nucleotides): they are likely different ends of the same gene that were not joined due to the relatively stringent assembly parameters employed.

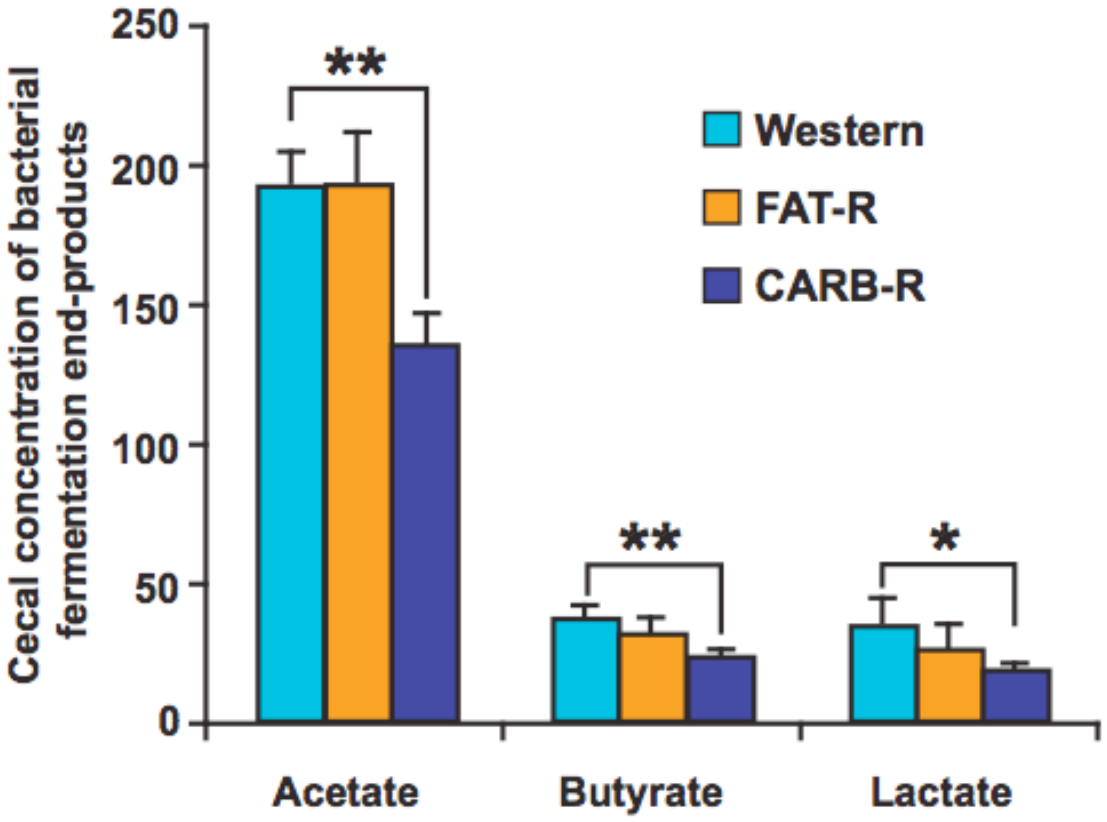

Figure S6 – Concentration of bacterial fermentation end-products in the ceca of Western, FAT-R, and CARB-R mice. Acetate and butyrate levels (μmol per g wet weight cecal contents) were measured by gas chromatography mass spectrometry. Lactate levels (mM per kg protein) were measured using established microanalytic methods (see Methods above). Data are represented as mean±SEM. Asterisks indicate significant differences (Student's t-test of Western versus CARB-R, *$p<0.05$, **$p<0.01$).

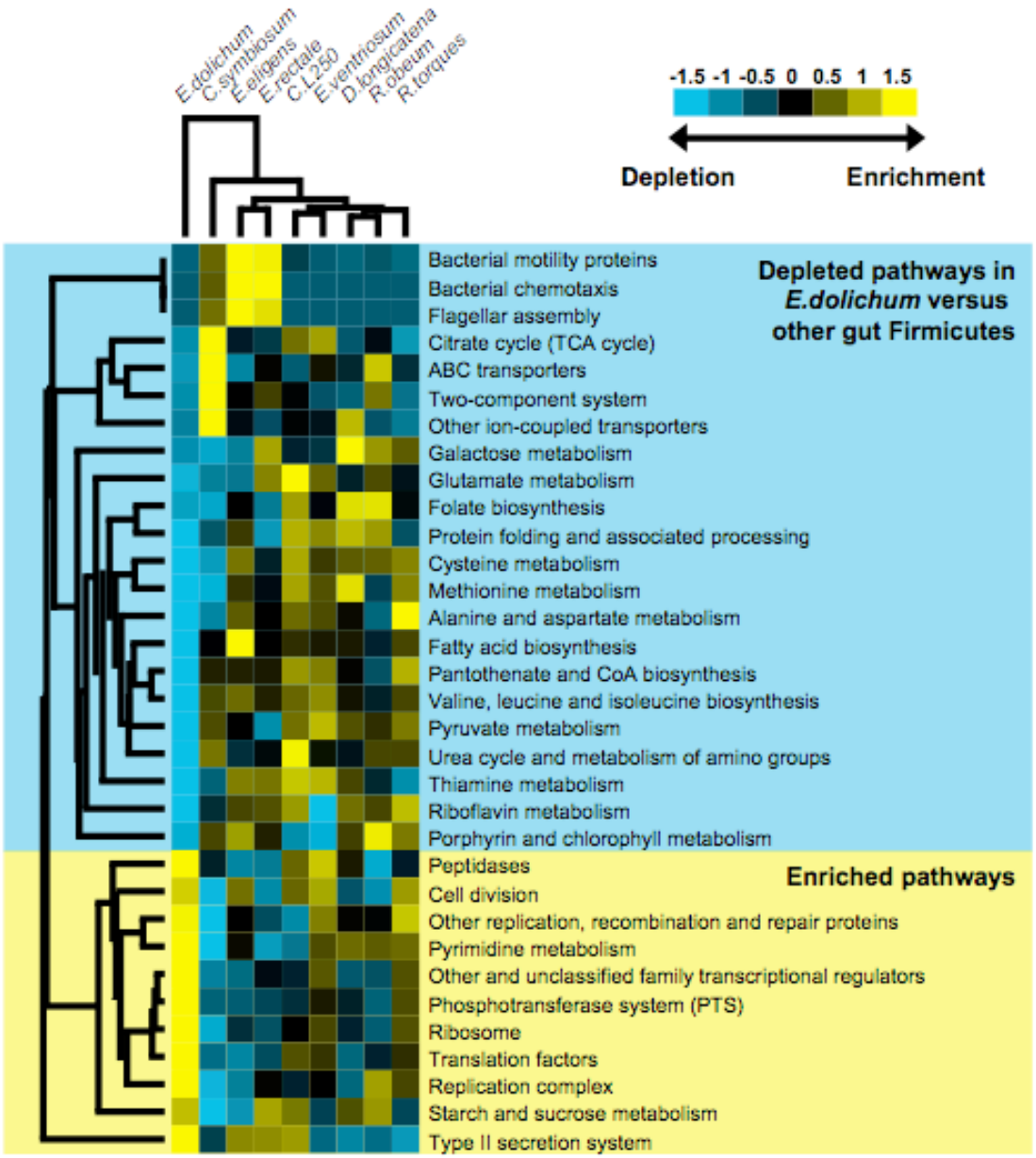

Figure S7 – KEGG metabolic pathways significantly enriched in the human gut-derived *Eubacterium dolichum* strain DSM 3991 genome relative to eight human gut-associated Firmicutes. Pathways whose relative representation is significantly different between the *E. dolichum* genome and the pooled gut Firmicute genomes (n=8) were identified using a bootstrap comparison of the abundance of sequences assigned to all KEGG pathways (xipe version 2.4; confidence level = 0.98, sample size = 10,000;

Rodriguez-Brito *et al.*, 2006). The relative abundance of all KEGG pathways with significantly different representation found at a relative abundance >0.6% in at least two microbiome datasets was transformed into a z-score and clustered by genome and pathway using a Euclidean distance metric (de Hoon *et al.*, 2004). Enrichment (yellow) and depletion (blue) are defined as a relative abundance greater or less than the mean for all datasets (i.e. a z-score greater or less than zero, respectively). For full strain names see **Figure 6**.

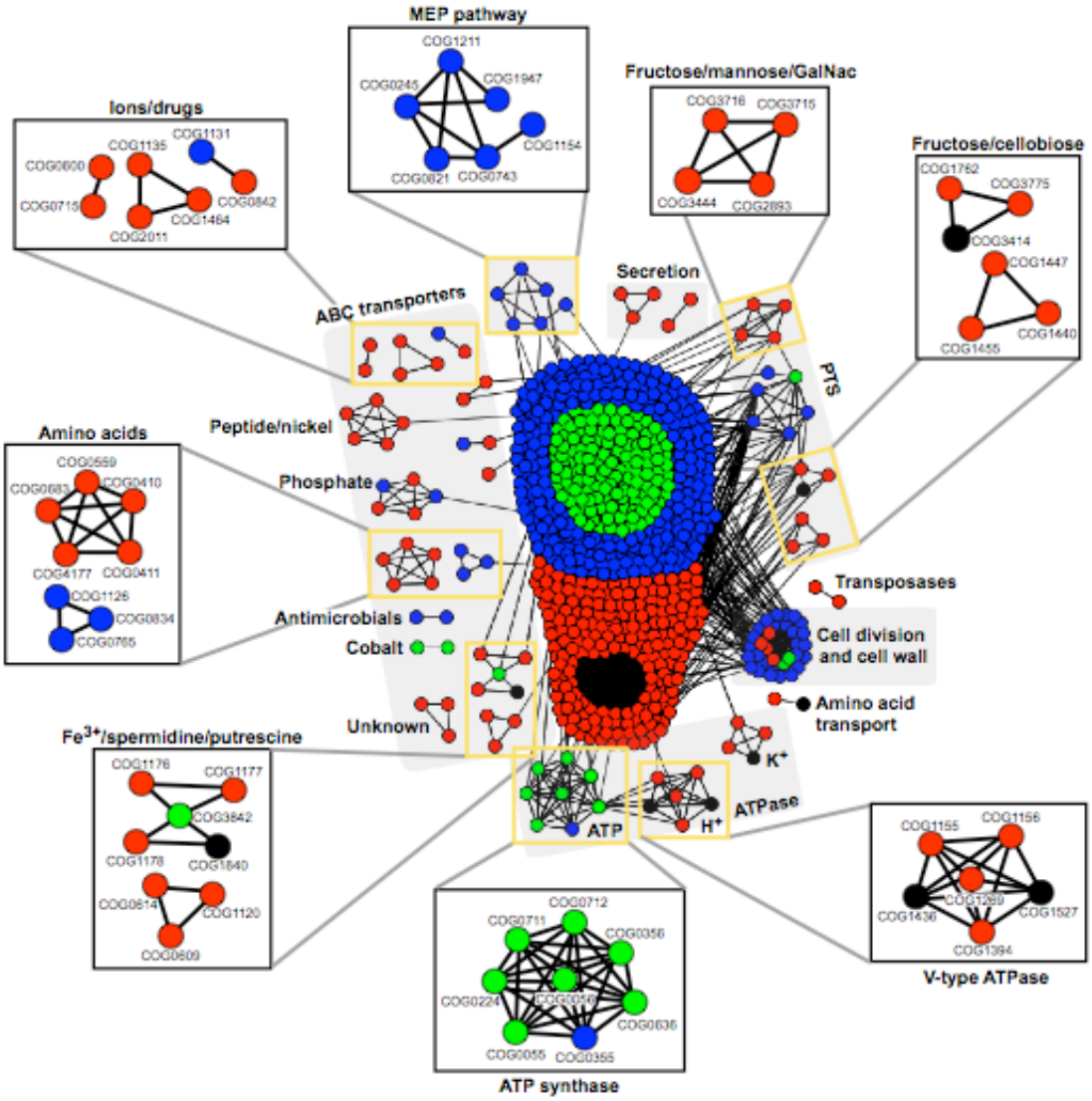

Figure S8 – STRING-based protein network analysis of the predicted *E. dolichum* proteome. MetaGene (Noguchi *et al.*, 2006) was used to predict proteins from the *E. dolichum* deep draft assembly. Proteins were subsequently assigned to COGs based on homology (BLASTP e-value<10^{-5}; STRING database version 7, von Mering *et al.*, 2007). Annotated COG interactions were used to organize the protein network, including interactions based on neighborhood, gene fusion, co-occurrence, homology, co-expression, experiments, databases, and text mining (Medusa *Java* appet; Hooper and Bork, 2005). Nodes, each representing a different orthologous group, are colored as

follows: green, present in all analyzed Firmicute genomes (including the mycoplasma); blue, present in all recently sequenced gut Firmicute genomes; red, present in the Western diet-associated cecal microbiome (based on BLAST homology searches, e-value<10^{-5} and the deposited annotations in the STRING database, version 7). 89% of the COGs found in the *E.dolichum* genome were also found in the Western diet microbiome. Most of the COGs in green are involved in essential cellular functions such as transcription and translation (56% of the COG category assignments are to 'Information storage and processing'). Some clusters of interest are highlighted, including the phosphotransferase system (PTS), the 2-methyl-D-erythritol 4-phosphate pathway for isoprenoid biosynthesis (MEP), cell wall biosynthesis, ABC transporters, and V-type ATPases for H^+ import.

SUPPLEMENTAL TABLES

Supplementary Table 1: Protein, carbohydrate, and fat composition of various mouse chow diets

Diet	Protein[a]	CHO[a]	Fat[a]	kcal/g
CHO[b]	23.2	60.7	16.1	3.74
Western	18.7	40.7	40.7	4.49
FAT-R	18.7	60.0	21.3	3.95
CARB-R	48.3	11.2	40.5	4.31

[a]values represent percentage of total kcal

[b]B&K Universal autoclavable chow diet (Sonnenburg *et al.*, 2006)

Supplementary Table 2: Percent weight of chow ingredients

Ingredient	Western	FAT-R	CARB-R
Casein	23.6	20.8	59.9
DL-Methionine	0.354	0.354	0.000
Sucrose, Cane	18.3	32.0	0.000
Corn Starch	16.0	16.0	0.000
Maltodextrin (Lo-Dex)	12.0	12.0	11.0
Vegetable Oil	10.0	5.00	10.0
Beef Tallow	10.0	4.10	8.80
Cellulose (Fiber)	4.00	4.00	4.00
Mineral Mix (AIN-93G-MX)	4.13	4.13	4.13
Calcium Phosphate Dibasic	0.472	0.472	0.472
Vitamin Mix (Teklad 40060)	1.18	1.18	1.18
Ethoxyquin (Antioxidant)	0.002	0.002	0.002
Calcium Carbonate	0.000	0.000	0.500

Supplementary Table 3: 16S rRNA gene-sequence libraries

Figure label	ARB label	Host	Host diet	16S gene sequences
CARB 1	WD1	CONV-D wt	CHO	96
CARB 2	WD2	CONV-D wt	CHO	343
CARB 3	WD3	CONV-D wt	CHO	267
CARB 4	WD4	CONV-D wt	CHO	207
CARB 5	WD5	CONV-D wt	CHO	216
Western 1	WD6	CONV-D wt	Western	222
Western 2	WD7	CONV-D wt	Western	256
Western 3	WD8	CONV-D wt	Western	221
Western 4	WD9	CONV-D wt	Western	220
Western 5	WD10	CONV-D wt	Western	221
Donor 1	WD11	CONV-R wt	-	194
CARB-R 1	MD4	CONV-R wt DIO, family 2	CARB-R	185
CARB-R 2	MD8	CONV-R wt DIO, family 1	CARB-R	233
CARB-R 3	MD9	CONV-R wt DIO, family 1	CARB-R	184
CARB-R 4	MD21	CONV-R wt DIO, family 2	CARB-R	259
CARB-R 5	MD23	CONV-R wt DIO, family 1	CARB-R	516
CARB-R 6	MD26	CONV-R wt DIO, family 2	CARB-R	138
FAT-R 1	MD18	CONV-R wt DIO, family 2	FAT-R	241
FAT-R 2	MD19	CONV-R wt DIO, family 2	FAT-R	203
FAT-R 3	MD24	CONV-R wt DIO, family 1	FAT-R	177
FAT-R 4	MD25	CONV-R wt DIO, family 1	FAT-R	162
FAT-R 5	MD27	CONV-R wt DIO, family 2	FAT-R	127
Western 6	MD2	CONV-R wt DIO, family 2	Western	263
Western 7	MD6	CONV-R wt DIO, family 1	Western	126
Western 8	MD7	CONV-R wt DIO, family 1	Western	176
Western 9	MD20	CONV-R wt DIO, family 2	Western	233
Western 10	MD22	CONV-R wt DIO, family 1	Western	193
CARB 6	myd1	CONV-R MyD88 -/-	CHO	241
CARB 7	myd2	CONV-R MyD88 -/-	CHO	260
CARB 8	myd3	CONV-R MyD88 -/-	CHO	266
Western 11	myd4	CONV-R MyD88 -/-	Western	223
Western 12	myd5	CONV-R MyD88 -/-	Western	231
-	rag2	CONV-R Rag1 -/-	CHO	66
-	rag3	CONV-R Rag1 -/-	CHO	103
-	rag4	CONV-R Rag1 -/-	Western	84
-	rag5	CONV-R Rag1 -/-	Western	111
	rag6	CONV-R Rag1 -/-	Western	94
-	CRWD2	CONV-R wt	Western	272
-	CRWD4	CONV-R wt	CHO	265
-	CRWD5	CONV-R wt	CHO	167
-	CRWD6	CONV-R wt	Western	225

Supplementary Table 4: Nomenclature used to designate microbiome datasets obtained from the cecal microbiota of C57BL/6J mice

Figure label	Microbiome label	Host family	Host state	Host diet	16S rRNA survey label	ARB label
Western 1	WEST1	1	CONV-R	Western	Western 10	MD22
FAT-R 1	FATR1	1	CONV-R	FAT-R	FAT-R 4	MD25
CARB-R 1	CARBR1	1	CONV-R	CARB-R	CARB-R 2	MD8
Western 2	WEST2	2	CONV-R	Western	Western 9	MD20
FAT-R 2	FATR2	2	CONV-R	FAT-R	FAT-R 5	MD27
CARB-R 2	CARBR2	2	CONV-R	CARB-R	CARB-R 4	MD21
Western 3	WEST3	-	CONV-D	Western	Western 3	WD8

Supplementary Table 5: Microbiome sequencing statistics

Microbiome	Average read length	Forward reads[a]	Sequence (Mb)
Western 1	668	9,072	6.1
FAT-R 1	586	10,681	6.3
CARB-R 1	603	10,773	6.5
Western 2	633	10,997	7.0
FAT-R 2	723	10,893	7.9
CARB-R 2	591	10,244	6.1
Western 3	734	11,705	8.6
TOTAL	-	**74,365**	**48**

[a]trimmed according to quality and vector sequence

Supplementary Table 6: Microbiome assembly statistics

Sample	Input reads	Mean trimmed read length	Fraction assembled (%)	Number of contigs	N50 contig length	Max contig length
Western 1	11136	612	0.9	17	1306	2159
FAT-R 1	12288	582	0.2	6	1218	1458
CARB-R 1	12288	590	0.0	2	1246	1246
Western 2	11904	573	2.6	37	1782	7376
FAT-R 2	11904	622	0.8	23	1428	3451
CARB-R 2	11520	575	0.1	4	1071	1236
Western 3	13440	627	6.7	107	1884	11022
All microbiomes	84480	598	3.9	387	1738	11990

Supplementary Table 7: Read placements in contigs and BLAST results

	Western 1	Western 2	Western 3	FAT-R 1	FAT-R 2	CARB-R 1	CARB-R 2	BBH[a]	e-value
contig 23	5	5	9	4	1	0	0	*S.mutans*	0
contig 73	1	0	2	0	0	0	0	*S.mutans*	4E-91
contig 146	2	4	4	3	1	0	0	*E.faecalis*	0
contig 161	2	1	4	3	3	0	0	*E.dolichum*	3E-97
contig 262	1	5	0	3	4	0	0	*L.monocytogenes*	1E-119

[a]Best-BLAST-Hit: BLASTX versus KEGG database (version 40) plus recently sequenced gut Firmicutes

Supplementary Table 8: *E.dolichum* draft genome sequencing statistics

Total contig number	51
Total contig bases	2209242
Average contig length	43318
Maximum contig length	453733
N50 contig length	291535
N50 contig number	3
Major contig (> 2000 bp) number	17
Major contig bases	2181491
GC content	38