Supporting Information

Brulc *et al*. 10.1073/pnas.0806191105

SI Text

Module Detection in *C. thermocellum* Simulation. A comparative analysis of the *C. thermocellum* genome and other metagenomes, revealed that although dockerins appear more difficult to detect in the simulated 454 short read query than are CBMs and GHs, there is variation present in their representation within different metagenomes. For example, in the case of GH48 (amino acid length 851 from *R. flavefaciens*), it is detected once in only 1 of the bovine microbiomes. Additionally, GH94 (cellobiose phosphorylases) that are predicted to be easily detectable because of their large size (>600 aa in length), are totally absent from the bovine microbiomes, but are very abundant in the termite hindgut as are the GH23 G-type lysozymes. Furthermore, we only detect GH29 (α-L-fucosidase), GH30 (glucosylceramidase/ß-1,6-glucanase/xylosidase), GH32 (invertase), GH78 (α-L-rhamnosidase), and GH97 (α-glucosidase) in the bovine rumen microbiomes. So, although longer genes are obviously more likely to be found more often, the variation in gene length among these families (amino acid length varying between 154 and 735 for GH modules, 45–187 for CBM modules, 150–160 for cohesin modules, and 64–87 for dockerin modules) does not fully account for differences in the sampling. Therefore, in addition to merely being longer, genes also have to be more abundant to be found more frequently. So, although the 454 short reads do detect dockerins and CBMs when they are present, their underrepresentation maybe because of a lower likelihood of detection by using pyrosequence short reads.

Supporting Methods. ***Rumen sampling.*** Samples of whole rumen contents were obtained from 3 fistulated 5-yr old Angus Simmental Cross steers (samples 8, 64, and 71) averaging ≈500 kg of that were maintained in open front barns at Illinois State University and housed singularly during the period of study. The steers were fistulated at 10 months of age and the fistulas are approximately 5 inches in diameter. The steers were placed on a NRC restricted diet of medium-quality grass-legume hay (13% crude protein, 60% neutral detergent fiber (NDF-cellulose, hemicellulose and lignin), 40% acid detergent fiber (ADF-cellulose and lignin), 20% hemicellulose, 0.5% calcium, 0.25% phosphorus, and 0.5% trace mineralized salt) at maintenance intake based on 2001 Dairy NRC for grass-legume hay (1). They were fed once a day for the entire length of the study (a total of 8 weeks including 2 weeks before sampling). Previous to this study, these steers were fed a corn silage diet (48% NDF, 28% ADF, 20% hemicellulose). The NRC grass-legume hay diet has more easily degradable fiber components, but contains less soluble more fermentable carbohydrates than the corn silage diet. Approximately 3 liters of whole rumen digesta (fiber-adherent (FA) and liquid associated microbes), was collected from the dorsal third rumen 6 weeks after the beginning of the study, 1 hour after the morning feeding. The rumen samples obtained appeared to be the same consistency and the fiber portions from each of the animals were slightly degraded but remained mainly intact. Samples were then partitioned into FA fractions and liquid fractions before DNA extraction (2, 3). The pooled sample (PL) (derived from a mixture of liquid fractions from each of the three steers) was thoroughly mixed before the DNA extraction (2, 4). Samples were stored at −80 °C until DNA extraction.

Bacterial 16S rRNA gene amplification for clone library construction. The universal bacterial primers 8FPL (AGTTTGATCCTGGCTCAG) and 1492RPL (GGYTACCTTGTTACGACTT) were used for PCR amplification of the 16S rDNA gene. Each PCR contained: 5 μl of 10X Ex Taq buffer (Takara Bio Inc.), 4 μl of 200 μm dNTP mixture, 0.3 μl of 1.25units/50 μl of Ex Taq DNA polymerase, 1.0 μl of each of 10 mM primer, 2 μl of sample and distilled water for a total of 50 μl of reaction. The PCR amplification was conducted on a BIO-RAD MJ Mini Personal Thermal Cycler (BIO-RAD Laboratories) and the conditions were 95 °C (5 min), followed by 30 cycles of 95 °C (30 sec), 57 °C (30 sec), 72 °C (30 sec) and a final extension of 72 °C (8 min).

Cloning and sequencing of 16S rRNA gene amplicons. 16S rRNA gene PCR products were cloned into pCR-4-TOPO vectors (Invitrogen) and electroporated into electrocompetent *E. coli* DH10B (Invitrogen). Transformants were selected by plating onto selective SOB/amp agar plates that were incubated overnight. Colonies were picked randomly and grown by overnight incubation in SOB/amp media. Plasmid template DNA from each transformant was prepared by a modified alkaline lysis method. The nucleotide sequences were determined by cycle sequencing by using BigDye Terminator (Applied Biosystems) and 3.2 pmol of M13F and M13R sequencing primers. Sequences were generated by using ABI 3730xl sequencers (Applied Biosystems) and trimmed to remove vector sequence. A total of 3,617 sequences were generated for the four microbiomes (805 for PL, 996 for FA-8, 857 for FA-64, and 959 for FA-71) and aligned in greengenes.

Multidimensional scaling statistical analyses. The comparison of 16S rRNA gene composition of the metagenomic libraries and the near full-length 16S rRNA gene libraries were conducted using Nonmetric Multidimensional Scaling (NMDS) (5). Briefly, each bacterial 16S rRNA gene sequence from each of the 8 libraries (1 metagenomic and 1 full length library for each of the 3 FA and one PL samples) was classified to the Phylum level by using the BLASTX algorithm with an expected cutoff of 1×10^{-5} (6), and the library's percentage contribution from each Phylum was calculated by dividing the number of sequences in each Phylum by the total number of 16S rRNA gene sequences in the library. The Bray-Curtis similarity coefficient [coefficient S_{17} of (7)] was then calculated for each possible pair of libraries by using this Phylum-level assessment of library composition resulting in an 8-by-8 similarity matrix used to conduct NMDS. NMDS seeks to graphically represent the ranked similarity between all pairs of samples by plotting the most similar samples closest together, the least similar samples farthest apart, and so on. NMDS was conducted with the software Primer 5 for Windows (8), using 20 random starting configurations, with the optimal two-dimensional solution having a final stress of 0.

The similar NMDS approach used to examine the similarity/differences in the abundances of CAZy proteins between the 4 bovine metagenomes and the microbial community from the termite hindguts (9) and a metagenome constructed from a bacterial species known to degrade cellulose, *Clostridium thermocellum*, began with a pyrosequence simulation of the *C. thermocellum* genome. The complete *C. thermocellum* genome (3,843,301 Mbp) was randomly sampled with replacement to generate a simulated metagenome. The fragments varied in length from 29 to 481 bp, and the relative abundance of each fragment length was modeled on the abundance of that sized fragment in the real 454 data (supporting information (SI) Fig. S9), Thus, there were most fragments that were 105 bp long, and very few fragments <90 bp or longer than 120 bp. A total of 29,999,657 Mbp (7.8057-fold coverage of the C. *thermocellum* genome) was generated by this sampling. The fragments gener-

ated were annotated and analyzed exactly as for the metagenomes. They were processed using the MG-RAST pipeline, and sequences similar to the GHs, CBMs, dockerins or cohesins were manually inspected for similarity. The number of sequences showing similarity to each of the 71 CAZy proteins in the six metagenomes was identified (described in *SI Methods*). A hierarchal clustering analysis was conducted on the number of sequences showing similarity to each protein using a χ^2 analysis (Table S3) (preformed on SPSS 15). The resultant 6-by-6 dissimilarity matrix was used to conduct a NMDS (SPSS v15). The NMDS was conducted with a single random start, minimum stress set at 0.0001 and 100 iterations.

1. Council NR (2001) *Nutrient Requirements of Dairy Cattle* (Natl Acad Press, Washington, DC), pp 1–381.
2. Larue R, Yu ZT, Parisi VA, Egan AR, Morrison M (2005) *Environ Microbiol* 7:530–543.
3. Yu Z, Morrison M (2004) *Biotechniques* 36:808–812.
4. Dehority B, Grubb J (1980) *Appl Environ Microbiol* 39:376–381.
5. Kruskal JB (1964) *Psychometrika* 29:1–27.
6. Edwards R, *et al.* (2006) *BMC Genomics* 7:57–69.
7. Legendre P, Legendre L (1998) *Numerical ecology* (Elsevier, Amsterdam) 2nd English Ed, pp 1–853.
8. Clarke KR, Gorley RN (2001) *PRIMER-E* (PRIMER-E Ltd, Plymouth, UK).
9. Warnecke F, *et al.* (2007) *Nature* 450:560–565.
10. Dinsdale EA, *et al.* (2008) *Nature* 452:629–632.

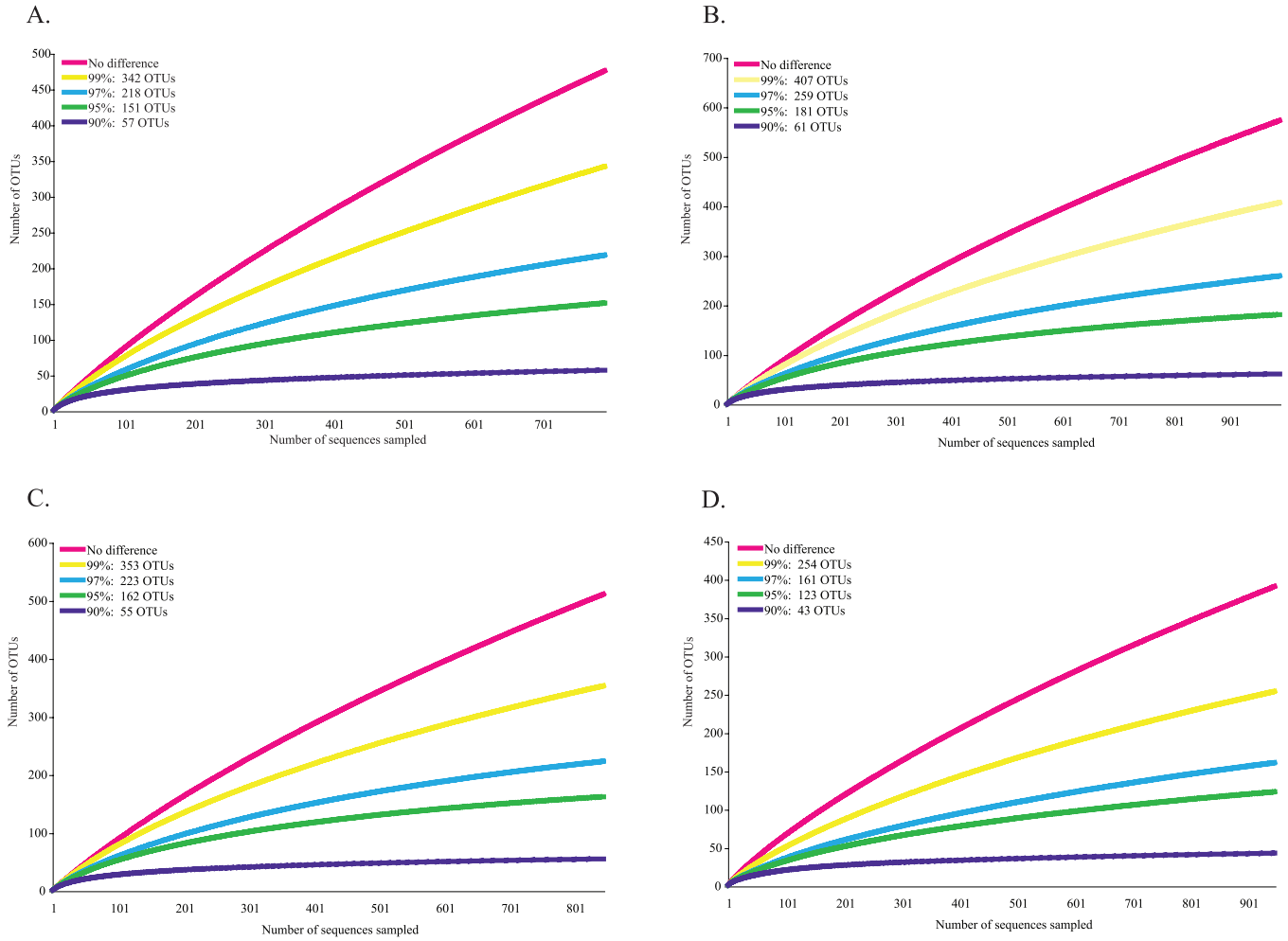

Fig. S1. Rarefaction curves of microbiomes (*A*) PL (*B*) FA from bovine 8 (*C*) FA from bovine 64, and (*D*) FA from bovine 71 at various phylotype cutoffs (90%, 95%, 97%, 99%, and no difference).

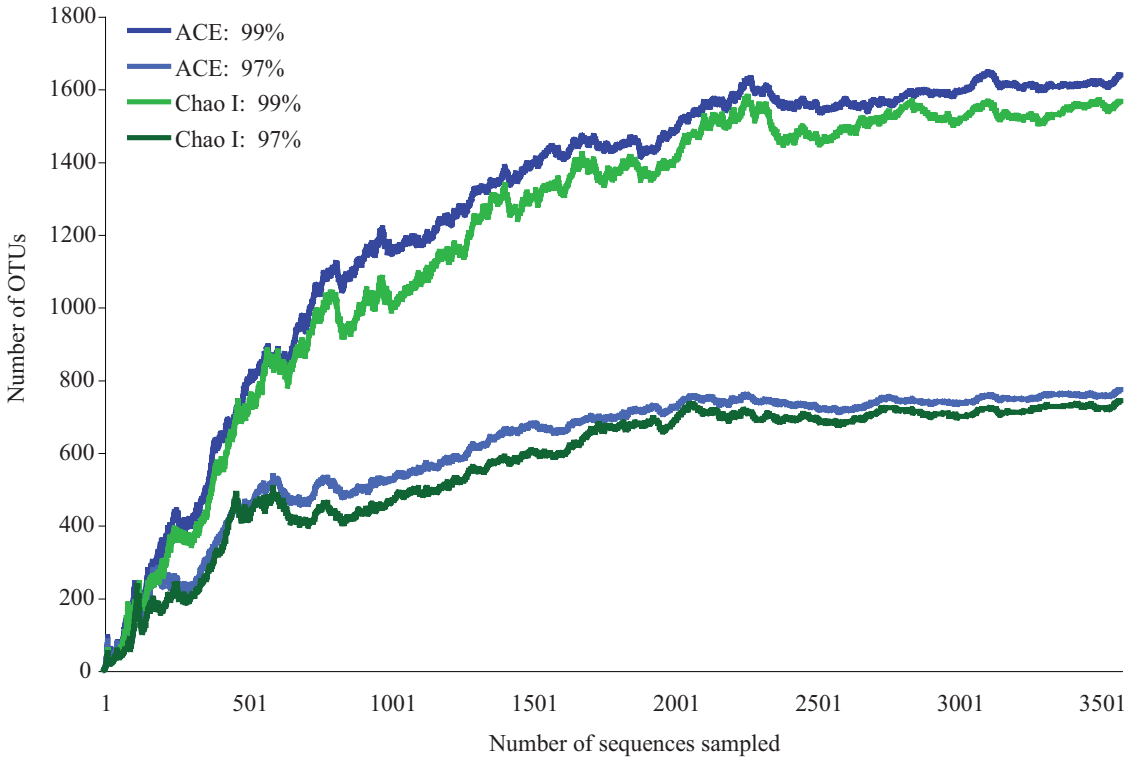

Fig. S2. ACE and ChaoI curves derived from the total number of sequences in full-length 16S libraries at phylotype cutoffs 99% and 97%.

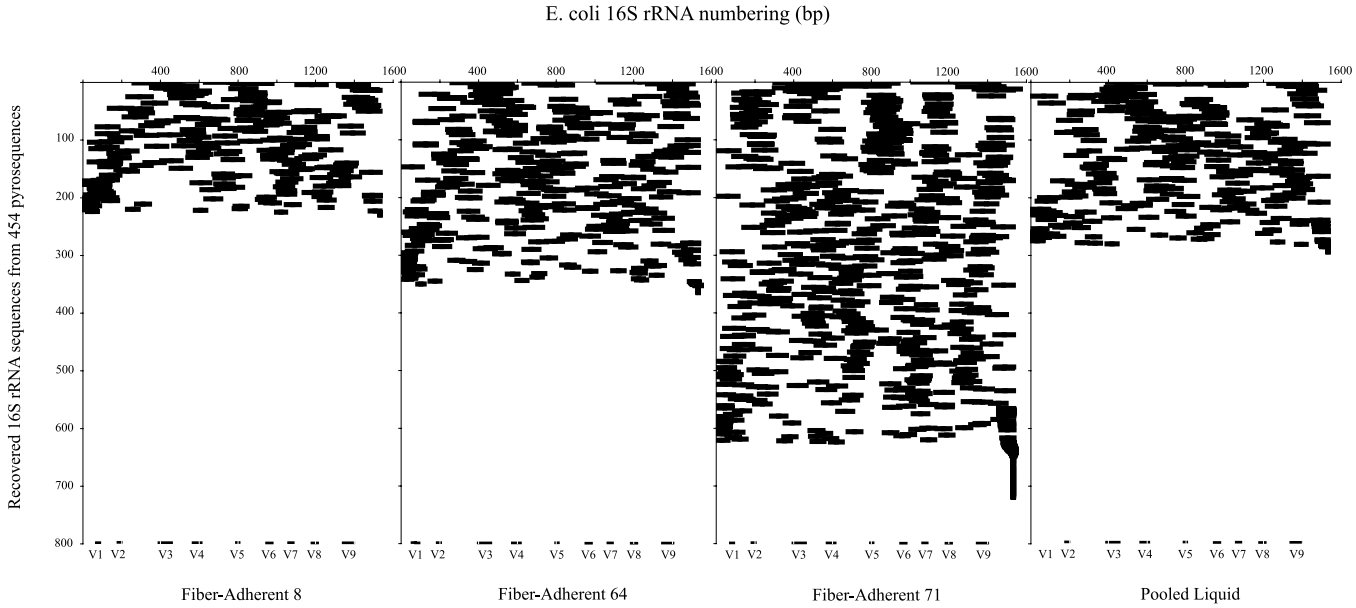

Fig. S3. Distribution of 16S rRNA gene hits from FA (bovines 8, 64, and 71) and PL microbiomes mapped onto the consensus SSU rRNA gene molecule. 16S rRNA gene hits are limited to those with an E value of 0.01 and a length hit of >50 nt.

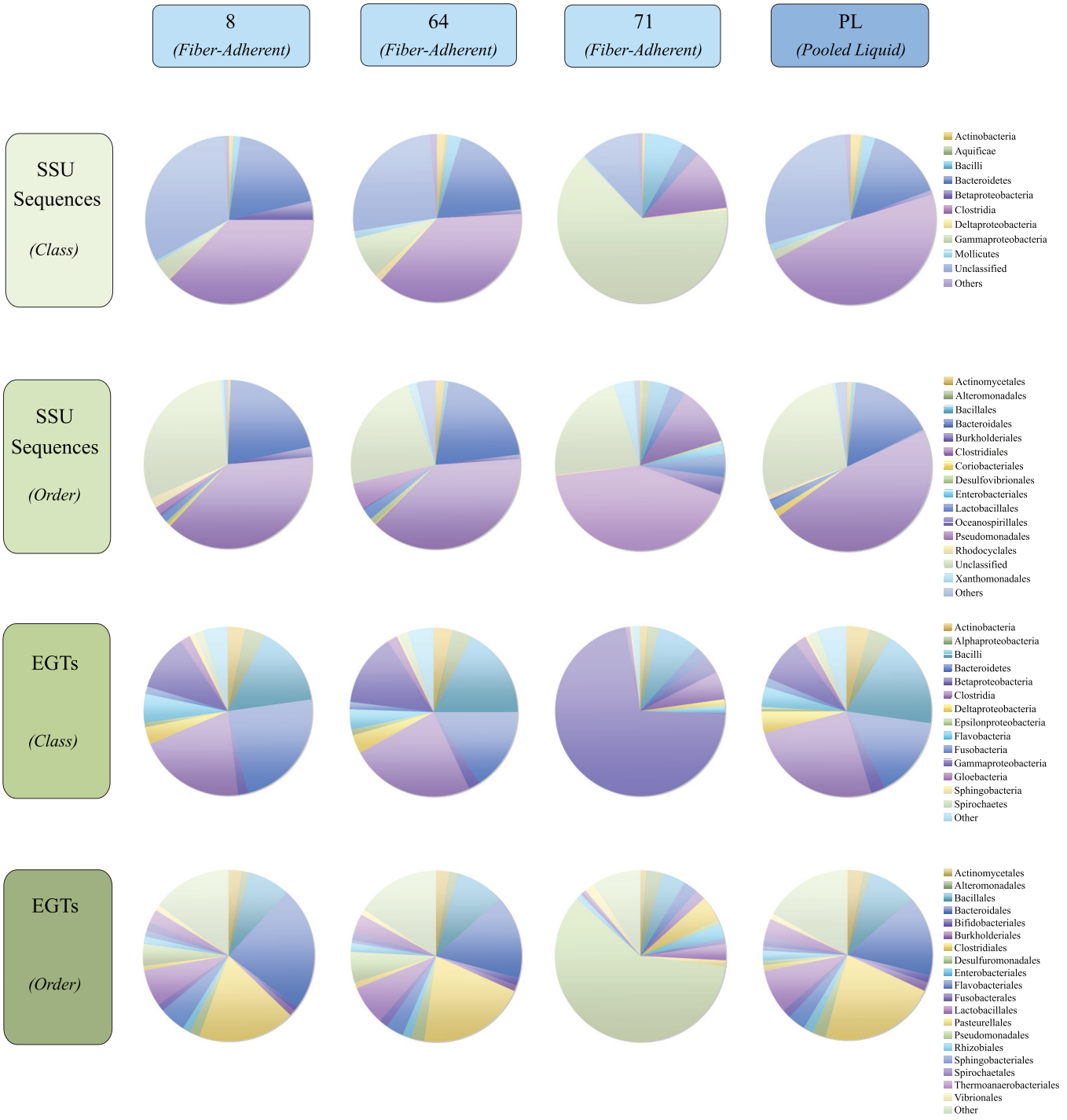

Fig. S4. Bacterial phylogenetic composition of pyrosequences from 4 bovine rumen samples. The percentage of SSU rRNA gene sequences or EGTs in either bacterial classes or bacterial orders from the FA (bovines 8, 64, and 71) and PL rumen microbiomes is shown. Bacterial classes or orders comprising <1% of the total amount of hits were grouped into the category ''others''.

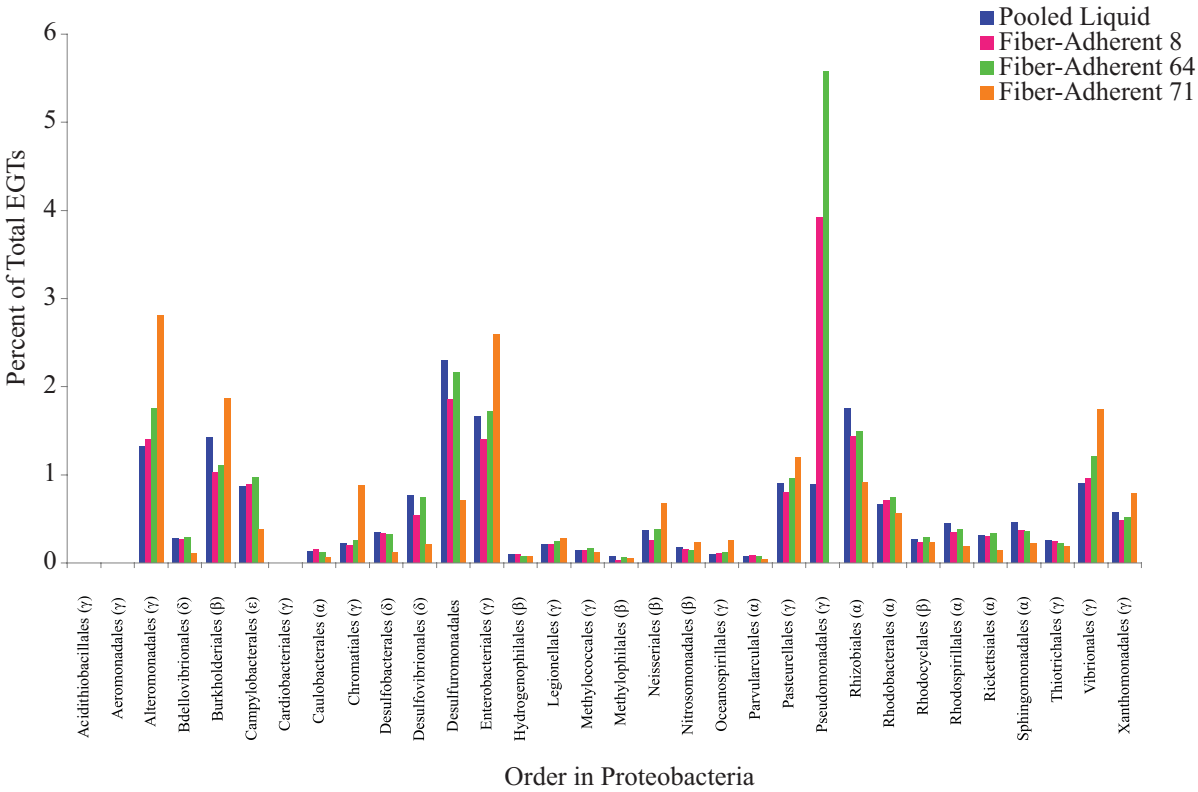

Fig. S5. Phylogenetic composition of proteobacteria EGTs from 4 pyrosequenced bovine rumen samples. The percentage of total EGTs in each order of proteobacteria from the FA (bovines 8, 64, and 71) and PL rumen microbiomes is shown (α-alphaproteobacteria, β-betaproteobacteria, δ-deltaproteobacteria, ε-epsilonproteobacteria, γ-gammaproteobacteria).

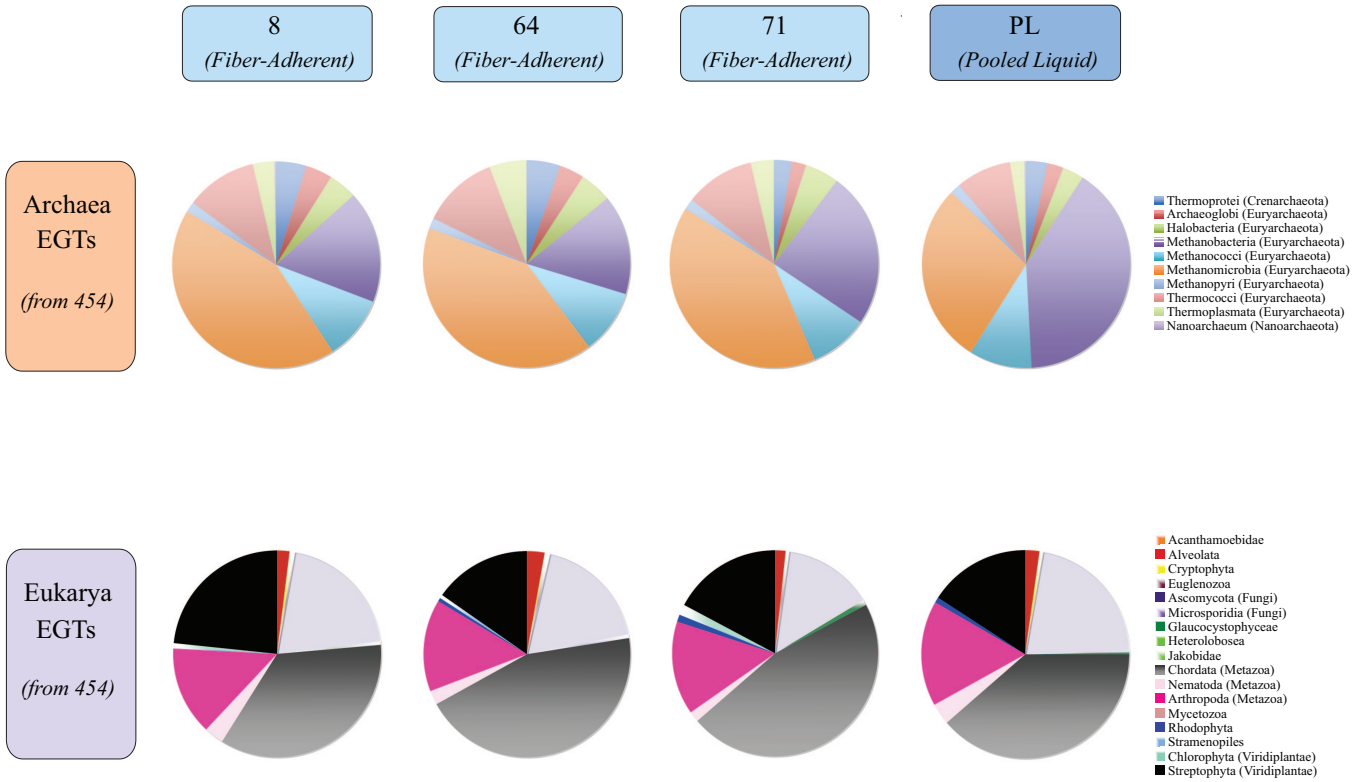

Fig. S6. Phylogenetic composition of archaeal and eukaryotic EGTs from four pyrosequenced bovine rumen samples. The percentage of EGTs in each of the archaeal class or eukaryotic division from the FA (bovines 8, 64, and 71) and PL rumen microbiomes is shown. The majority of Archaeal EGTs correspond to methanogenic classes with the largest proportion corresponding to the PL-microbiome. No archaeal SSU rRNA gene phylotypes were identified in the FA-microbiome from bovine 8, although 4 sequences similar to Methanobacteria were found in the other 2 FA-microbiomes. Additionally, 14 Methanobacteria SSU 16S rRNA gene sequences were found in the PL-microbiome. In the Eukaryotic comparison, fungal rRNA gene sequences were not identified in any of these samples; however, 19% of eukaryotic EGTs appear to be fungal.

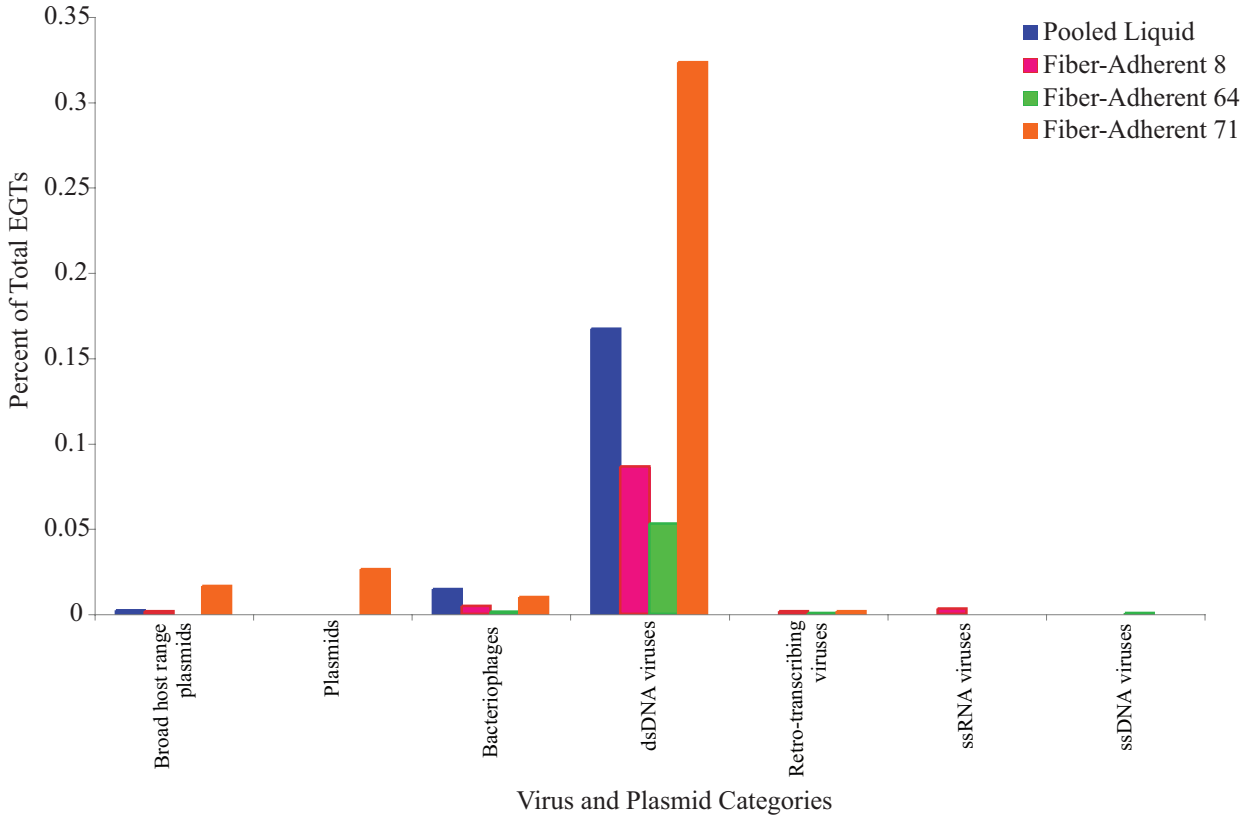

Fig. S7. Phylogenetic composition of the virus and plasmid EGTs from four pyrosequenced bovine rumen samples. The percentage of total EGTs in each of the virus and plasmid category from the FA (bovine 8, 64, and 71) and PL rumen microbiomes is shown.

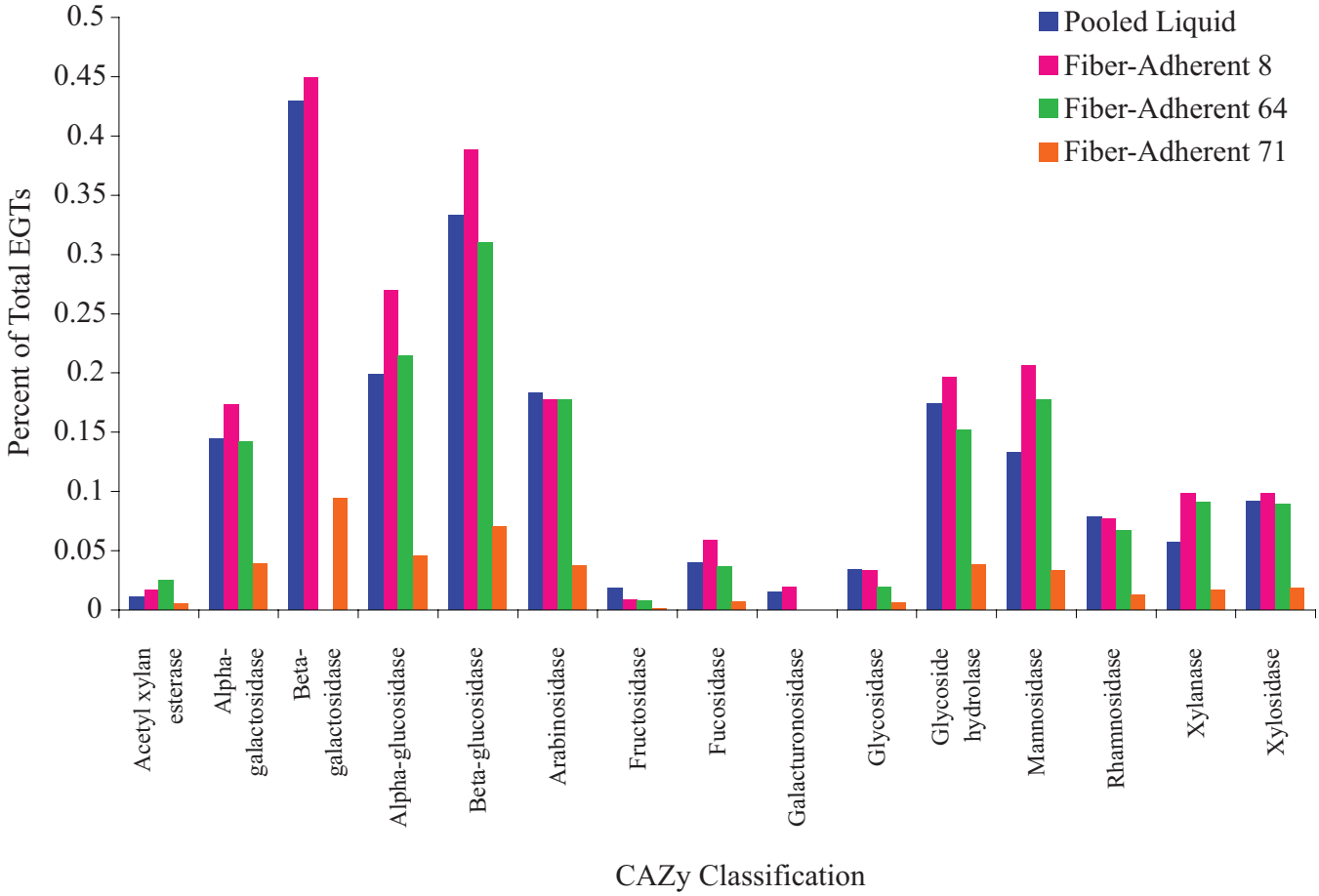

Fig. S8. Composition of plant cell wall degrading enzyme EGTs from four pyrosequenced bovine rumen samples. The percentage of total EGTs in each of the plant cell wall degrading enzyme categories from the FA (bovine 8, 64, and 71) and PL rumen microbiomes is shown.

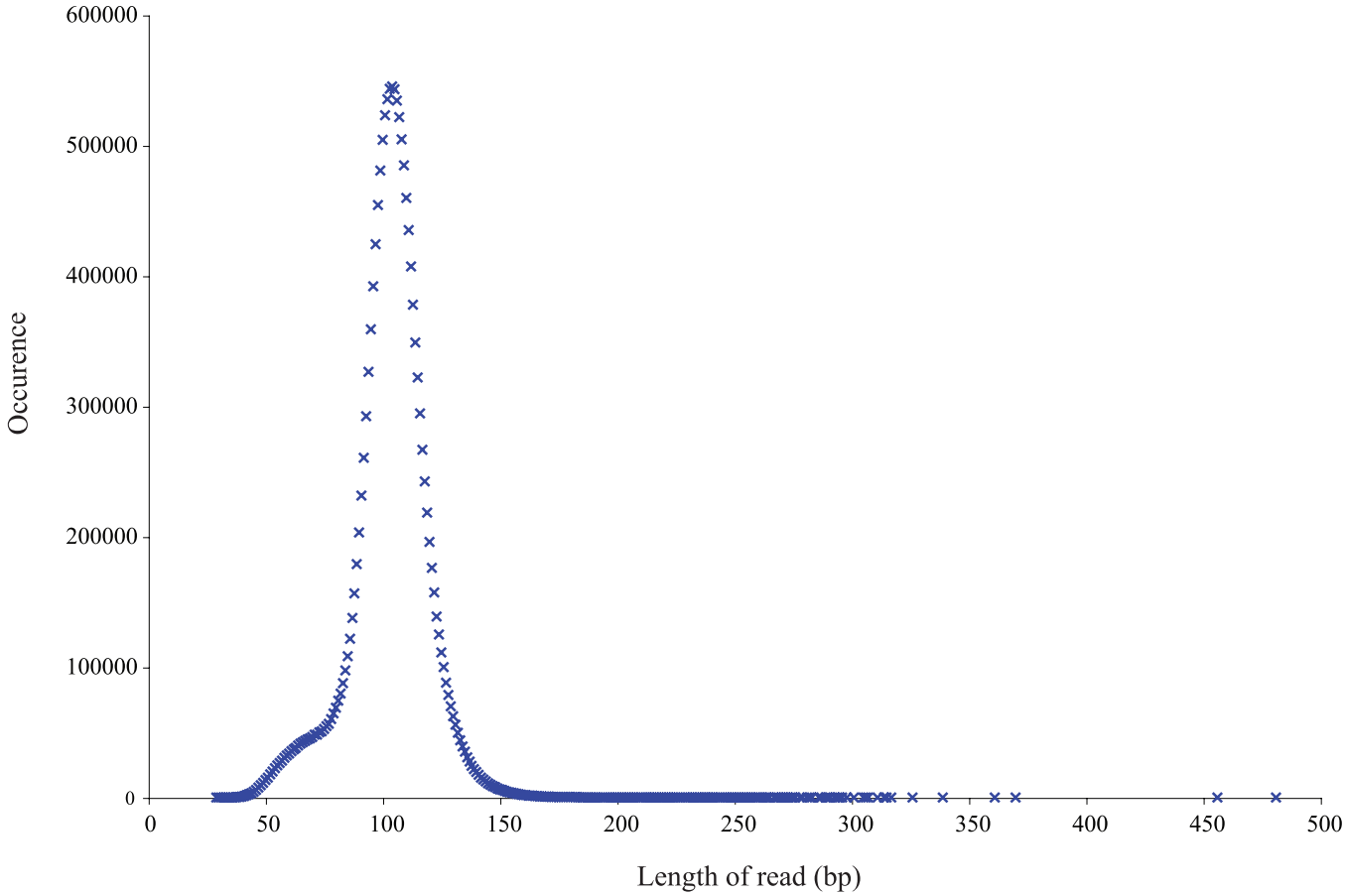

Fig. S9. Distribution of 454 simulated fragments from the *C. thermocellum* genome (3,843,301 Mbp) used for analysis presented in Fig. 1*C* and Table 2.

Table S1. Consistency comparison for top BLAST hits at both the Order and Family taxonomic levels for FA (bovines 8, 64, and 71) and PL rumen microbiomes A

Taxonomic level	Parameter	Fiber-adherent fractions			Liquid fraction
		8	64	71	PL
Order					
Top 2 BLAST hits					
	Total number of pegs	6100	7636	46425	6566
	Number of pegs (%) at consistency fraction:				
	0.5	957 (15.7)	1221 (16.0)	1638 (3.5)	1269 (19.3)
	1.0	5143 (84.3)	6415 (84.0)	44787 (96.5)	5297 (80.7)
Top 3 BLAST hits					
	Total number of pegs	3894	4606	29347	3835
	Number of pegs (%) at consistency fraction:				
	0.33	95 (2.4)	123 (2.7)	171 (0.6)	142 (3.7)
	0.67	2378 (61.1)	2311 (50.2)	1716 (5.9)	2221 (57.9)
	1.00	1421 (36.5)	2172 (47.2)	27460 (93.6)	1472 (38.4)
Top 5 BLAST hits					
	Total number of pegs	1512	1803	3308	1709
	Number of pegs (%) at consistency fraction:				
	0.20	3 (0.2)	5 (0.3)	7 (0.2)	1 (0.1)
	0.40	74 (4.9)	76 (4.2)	125 (3.8)	84 (4.9)
	0.60	864 (57.1)	833 (46.2)	489 (14.8)	753 (44.1)
	0.80	96 (6.4)	161 (8.9)	290 (8.8)	145 (8.5)
	1.00	475 (31.4)	728 (40.4)	2397 (72.5)	726 (42.5)
Top 10 BLAST hits					
	Total number of pegs	261	458	986	435
	Number of pegs (%) at consistency fraction:				
	0.10				0
	0.20	2 (0.8)		1 (0.1)	2 (0.5)
	0.30	4 (1.5)	6 (1.3)	14 (1.4)	8 (1.8)
	0.40	8 (3.1)	17 (3.7)	33 (3.4)	14 (3.2)
	0.50	14 (5.4)	11 (2.4)	75 (7.6)	11 (2.5)
	0.60	12 (4.6)	23 (5.0)	59 (6.0)	20 (4.6)
	0.70	2 (0.8)	11 (2.4)	32 (3.3)	9 (2.1)
	0.80	17 (6.5)	30 (6.6)	62 (6.3)	25 (5.8)
	0.90	31 (11.9)	70 (15.3)	82 (8.3)	61 (14.0)
Family					
Top 2 BLAST hits					
	Total number of pegs	6100	7636	46425	6566
	Number of pegs (%) at consistency fraction:				
	0.5	2736 (44.9)	3795 (50.0)	40572 (87.4)	2885 (43.9)
	1.0	3364 (55.2)	3841 (50.0)	5853 (12.6)	3681 (56.1)
Top 3 BLAST hits					
	Total number of pegs	3894	4606	29347	3835
	Number of pegs (%) at consistency fraction:				
	0.33	397 (10.2)	547 (11.9)	1199 (4.1)	625 (16.3)
	0.67	3068 (78.8)	3416 (74.2)	26661 (90.9)	2602 (67.9)
	1.00	429 (11.0)	643 (14.0)	1487 (5.1)	608 (15.9)
Top 5 BLAST hits					
	Total number of pegs	1512	1803	3308	1709
	Number of pegs (%) at consistency fraction:				
	0.20	28 (1.8)	47 (2.6)	141 (4.3)	53 (3.1)
	0.40	210 (13.9)	256 (14.2)	721 (21.8)	276 (16.2)
	0.60	955 (63.2)	978 (54.2)	1336 (40.4)	879 (51.4)
	0.80	116 (7.7)	178 (9.9)	344 (10.4)	190 (11.1)
	1.00	203 (13.4)	344 (19.1)	766 (23.2)	311 (18.2)
Top 10 BLAST hits					
	Total number of pegs	261	458	986	435
	Number of pegs (%) at consistency fraction:				
	0.10	0	0	3 (0.3)	3 (0.7)
	0.20	16 (6.1)	20 (4.4)	100 (10.1)	24 (5.5)
	0.30	31 (11.9)	35 (7.6)	113 (11.5)	45 (10.3)
	0.40	16 (6.1)	47 (10.3)	91 (9.2)	31 (7.1)
	0.50	19 (7.3)	28 (6.1)	64 (6.5)	26 (6.0)
	0.60	31 (11.9)	63 (13.8)	125 (12.7)	77 (17.7)
	0.70	22 (8.4)	53 (11.6)	84 (8.5)	46 (10.6)
	0.80	26 (10.0)	47 (10.3)	58 (5.9)	27 (6.2)
	0.90	44 (16.9)	67 (14.6)	79 (8.0)	70 (16.1)
	1.00	56 (21.5)	98 (21.4)	269 (27.3)	86 (19.8)
	1.00	171 (65.5)	290 (63.3)	628 (63.7)	285 (65.5)

For each protein-encoding gene (peg) in the rumen microbiome samples, the consistency of the taxonomy of the top number of hits is listed. Consistency was measured by determining the number of times the taxonomies for each peg was the same out of all possibilities (for example, a consistency fraction of 0.5 indicates that for the top 2 BLAST hits, one taxonomy is the same and one is different, whereas if the fraction were 1.0, both taxonomies for these hits are the same). Results appear very consistent among the top hits. For example, $<$4% of the sequences have three different taxonomic Orders among the top three BLAST hits, and $>$60% of the sequences with 10 BLAST hits all came from the same taxonomic Order. The consistency was less at the family level, with only 20% of the sequences with 10 BLAST hits coming from the same family.

Table S2. Percentage of total EGTs in subsystems statistically more likely to be present in the FA (bovines 8, 64, and 71) and PL rumen microbiomes

Classification	Subclassification	Subsystem	Metagenome			
			PL	80F6	640F6	710F6
Amino acids and derivatives						
	Branched-chain amino acids	HMG CoA synthesis				0.44
		Isoleucine degradation				0.13
		Leucine degradation and HMG-CoA metabolism				7.37
		Valine degradation				7.37
Carbohydrates						
	Central carbohydrate metabolism	Dehydrogenase complexes				1.38
		Glyoxylate bypass				7.14
		Pyruvate metabolism II: Acetyl-CoA, acetogenesis from pyruvate				8.79
		TCA cycle				1.36
		Glycogen metabolism	6.89	0.76	0.72	
		Pyruvate:ferredoxin oxidoreductase			0.43	
	Di- and oligosaccharides	Beta-glucoside metabolism		0.60		
		Maltose and maltodextrin utilization	12.70	1.33	0.14	
		Sucrose metabolism	2.21	1.76	1.59	
	Fermentation	Acetyl-CoA fermentation to butyrate				11.89
		Butanol biosynthesis				0.83
	Monosaccharides	Lactose utilization	0.43	0.47	0.54	
		L-Arabinose utilization	3.98	0.66	0.46	
		Xylose utilization	5.28	0.53	0.55	
	One-carbon metabolism	Serine-glyoxylate cycle				17.24
	Organic acids	Glyoxylate synthesis				4.28
		Methylcitrate cycle				6.96
		Propionate-CoA to succinate module				0.71
Cell wall and capsule						
	Capsular and extracellular polysacchrides	Rhamnose containing glycans		0.89		
Clustering-based subsystems						
	Uncategorized	EC49–61				2.87
Cofactors, vitamins, prosthetic groups, pigments						
	NAD and NADP	NAD and NADP cofactor biosynthesis global				0.90
	Quinone cofactors	Ubiquinone biosynthesis				0.46
	Tetrapyrroles	Coenzyme B12 biosynthesis	7.22			
Experimental subsystems						
	Uncategorized	271-Bsub	6.48	6.37	6.64	
		Isocitrate methylcitrate isopropylmalate disambiguation				1.74
Fatty acids and lipids						
	Fatty acids	Fatty acid oxidation pathway				1.86
	Uncategorized	Polyhydroxybutyrate metabolism				1.60
Membrane transport(ers)						
	ABC transporters	ABC transporter dipeptide (TC 3.A.1.5.2)	4.83			
		ABC transporter oligopeptide (TC 3.A.1.5.1)	6.66			
	Uni- Sym- and antiporters	Na(+) H(+) antiporter				2.75
Metabolism of aromatic compounds						
	Metabolism of central aromatic intermediates	Homogentisate pathway of aromatic compound degradation				0.49
	Peripheral pathways for catabolism of aromatic compounds	n-Phenylalkanoic acid degradation				11.94
Miscellaneous						
	Uncategorized	ABC transporter maltose			0.55	
Nitrogen metabolism						
	Uncategorized	Nitrate and nitrite ammonification				4.28
Respiration						
	Electron donating reactions	Hydrogenases	6.93	0.65	0.78	
	Sodium ion-coupled energetics	Na^+ translocating decarboxylases and related biotin-dependent enzymes		0.60	0.63	
	Uncategorized	Biogenesis of cbb3-type cytochrome c oxidases			2.54	
		Biogenesis of c-type cytochromes				3.25
Stress response						
	Uncategorized	Copper homeostasis				5.90

Classification	Subclassification	Subsystem	Metagenome			
			PL	80F6	640F6	710F6
Sulfur metabolism						
	Inorganic sulfur assimilation	Inorganic sulfur assimilation				3.98
	Uncategorized	Galactosylceramide and sulfatide metabolism	0.57	0.66	0.70	
Virulence						
	Resistance to antibiotics and toxic compounds	Tetracycline resistance, ribosome protection type	1.38	1.19	0.13	

These subsystems are more frequently found from these four samples with a sample size of 5,000 proteins, 20,000 repeated samples, and a $P < 0.02$. The subsystems are annotated across genomes and are based on biochemical pathways, fragments of pathways, and clusters of genes that function together, or any group of genes considered to be related. Consistent with a previous analysis of multiple biomes (10), microbes associated with bovine rumen are dominated by carbohydrate metabolism, and are sparsely populated with genes for respiration.

Table S3. Dissimilarity matrix comparing *Clostridium thermocellum* "simulated" pyrosequenced genome and other microbiomes

	Chi-square between sets of frequencies					
Case	1: *C. thermocellum* 454 simulation	2: Pooled liquid	3: Fiber-adherent 8	4: Fiber-adherent 64	5: Fiber-adherent 71	6: Termite hindgut
1: *C. thermocellum* 454 Simulation	0.000	47.917	47.080	47.866	42.165	33.735
2: Pooled liquid	47.917	0.000	7.547	7.940	9.394	30.259
3: Fiber-adherent 8	47.080	7.547	0.000	7.109	10.137	30.150
4: Fiber-adherent 64	47.866	7.940	7.109	0.000	8.599	31.195
5: Fiber-adherent 71	42.165	9.394	10.137	8.599	0.000	24.985
6: Termite hindgut	33.735	30.259	30.150	31.195	24.985	0.000

Table S4. Nitrogen Metabolism subsystem composition of the FA and PL bovine rumen metagenomes

Subclassification	Subsystem	Metagenome			
		PL	80F6	640F6	710F6
Allantoin degradation					
	2-hydroxy-3-oxopropionate reductase (EC 1.1.1.60)			1	1
	Allantoinase (EC 3.5.2.5)				2
	Dihydroorotase (EC 3.5.2.3)	28	22	28	47
	Glycerate kinase (EC 2.7.1.31)	1		2	3
	Ureidoglycolate dehydrogenase (EC 1.1.1.154)		2		
Ammonia assimilation					
	Ammonium transporter	12	7	10	52
	Ammonium transporter family	5	4	7	1
	Ammonium/methylammonium permease	1	1		1
	Ferredoxin-dependent glutamate synthase (EC 1.4.7.1)				1
	Glutamate synthase [NADPH] large chain (EC 1.4.1.13)	5	8	3	3
	Glutamate synthase [NADPH] small chain (EC 1.4.1.13)	1	4	3	1
	Nitrogen regulation protein NR(I)				29
	Nitrogen regulation protein NR(II) (EC 2.7.3.−)			1	3
	Nitrogen regulatory protein P-II	2	2	2	1
Cyanate hydrolysis					
	Cyanate ABC transporter, ATP-binding protein	1	1	2	
	Cyanate ABC transporter, permease protein				1
	Cyanate ABC transporter, substrate binding protein			1	1
	Cyanate hydratase (EC 4.2.1.104)			1	
Denitrification					
	Copper-containing nitrite reductase (EC 1.7.2.1)				2
	Nitric oxide reductase activation protein NorQ	1	1		1
	Nitric-oxide reductase (EC 1.7.99.7), quinol-dependent		1	4	52
	Nitric-oxide reductase subunit B (EC 1.7.99.7)	1			
	Nitrous oxide reductase maturation protein NosF (ATPase)		6	1	4
Dissimilatory nitrite reductase					
	Cytochrome cd1 nitrite reductase (EC:1.7.2.1)				1
	Heme d1 biosynthesis protein NirJ	1	1		1
	Heme d1 biosynthesis protein NirL	1			
	Uroporphyrinogen-III methyltransferase (EC 2.1.1.107)	2			
Nitrate and nitrite ammonification					
	Assimilatory nitrate reductase large subunit (EC:1.7.99.4)	2	1		6
	Cytochrome c nitrite reductase, small subunit NrfH	4		4	3
	Cytochrome c-type protein NapC				2
	Cytochrome c552 precursor (EC 1.7.2.2)	1			2
	Nitrate ABC transporter, ATP-binding protein	3			2
	Nitrate/nitrite response regulator protein	1	1	3	26
	Nitrate/nitrite sensor protein (EC 2.7.3.−)		1		40
	Nitrate/nitrite transporter	1	3	6	175
	Nitrite reductase [NAD(P)H] large subunit (EC 1.7.1.4)		1	1	7
	Nitrite reductase [NAD(P)H] small subunit (EC 1.7.1.4)				1
	Nitrite reductase probable [NAD(P)H] subunit (EC 1.7.1.4)	3	5	7	2
	Nitrite reductase probable electron transfer 4Fe-S subunit (EC 1.7.1.4)	2	1		1
	NrfC protein			1	
	Periplasmic nitrate reductase component NapE				1
	Respiratory nitrate reductase alpha chain (EC 1.7.99.4)		3	11	175
	Respiratory nitrate reductase beta chain (EC 1.7.99.4)		1	1	49
	Respiratory nitrate reductase delta chain (EC 1.7.99.4)		2		46
	Respiratory nitrate reductase gamma chain (EC 1.7.99.4)		1		29
Nitric oxide synthase					
	Manganese superoxide dismutase (EC 1.15.1.1)				2
	Putative cytochrome P450 hydroxylase				1
Nitrogen fixation					
	FeMo cofactor biosynthesis protein NifB	3	1		1
	Mo/Fe-nitrogenase-specific transcriptional regulator NifA	2	2	3	1
	Nitrogenase (iron-iron) transcriptional regulator		1		1
Nitrosative stress					
	Anaerobic nitric oxide reductase flavorubredoxin	1			1
	Anaerobic nitric oxide reductase transcription regulator norR	3	1	3	1
	Ferredoxin 3 fused to uncharacterized domain	16	8	7	6
	Hcp transcriptional regulator HcpR (Crp/Fnr family)	2			1
	Nitric oxide reductase FlRd-NAD(+) reductase (EC 1.18.1.−)		1	1	14

The number of environmental gene tags (EGTs) in each of the Nitrogen Metabolism subsystems is shown. The BLASTX cutoff for an EGT is 1×10^{-5}.