Applied and Environmental Microbiology, June 2010, p. 3545–3553 Vol. 76, No. 11
0099-2240/10/$12.00 doi:10.1128/AEM.02689-09

Diversity of Bacteria and Glycosyl Hydrolase Family 48 Genes in Cellulolytic Consortia Enriched from Thermophilic Biocompost▽

Javier A. Izquierdo,[1,2] Maria V. Sizova,[1,2]† and Lee R. Lynd[1,2]*

Thayer School of Engineering, Dartmouth College, Hanover, New Hampshire 03755, and BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831

Received 4 November 2009/Accepted 30 March 2010

The enrichment from nature of novel microbial communities with high cellulolytic activity is useful in the identification of novel organisms and novel functions that enhance the fundamental understanding of microbial cellulose degradation. In this work we identify predominant organisms in three cellulolytic enrichment cultures with thermophilic compost as an inoculum. Community structure based on 16S rRNA gene clone libraries featured extensive representation of clostridia from cluster III, with minor representation of clostridial clusters I and XIV and a novel *Lutispora* species cluster. Our studies reveal different levels of 16S rRNA gene diversity, ranging from 3 to 18 operational taxonomic units (OTUs), as well as variability in community membership across the three enrichment cultures. By comparison, glycosyl hydrolase family 48 (GHF48) diversity analyses revealed a narrower breadth of novel clostridial genes associated with cultured and uncultured cellulose degraders. The novel GHF48 genes identified in this study were related to the novel clostridia *Clostridium straminisolvens* and *Clostridium clariflavum*, with one cluster sharing as little as 73% sequence similarity with the closest known relative. In all, 14 new GHF48 gene sequences were added to the known diversity of 35 genes from cultured species.

The exploration and understanding of cellulose fermentation capabilities in nature could inform and enable industrial processes converting cellulosic biomass to fuels and other products. Enrichment of microbial communities that can utilize cellulose is useful in this context for the identification of novel organisms, novel metabolisms, and novel functions. Of particular interest are communities that can utilize cellulose at high temperatures and under anaerobic conditions, featuring high rates of solubilization under conditions where the energy and the reducing power of substrates are conserved in potentially useful fermentation products.

Some evidence indicates that cocultures may be able to utilize cellulose more fully and produce higher concentrations of ethanol than pure cultures of model cellulolytic organisms such as *Clostridium thermocellum* and *Clostridium straminisolvens* (16, 20, 34). An initial step toward understanding the functional roles of community members in cooperative cellulose degradation is answering the question of what organisms are present in cellulolytic consortia obtained from nature. Currently, diversity estimation methods applied to cellulolytic communities range from traditional methods targeting the 16S rRNA gene (4, 12) to complex metagenomic analyses targeting the breadth of functional genes present in genomes of mixed cultures and the environment (3).

From a functional gene standpoint, cellulase systems are complex assemblages of multifunctional glycosyl hydrolases. Even particularly relevant families, such as family 5 and family 9, tend to include hydrolases with multiple substrate specificities, deep evolutionary roots, and extensive sequence diversity within the same organism (19). However, family 48 glycosyl hydrolases include a select group of cellulosomal and unbound cellulases thought to play an essential role in cellulose solubilization by model cellulolytic clostridia (5, 7, 15), actinobacteria (6, 13), and anaerobic fungi (31). One key feature of this family of glycosyl hydrolases (mostly exoglucanases) is their ability to enhance cellulose solubilization in synergistic interactions with family 9 glycosyl hydrolases (2, 13). But unlike the latter, and with the notable exception of CelS and CelY in *Clostridium thermocellum*, family 48 hydrolases are present mostly in single copies in the genomes of cellulolytic microbes, making family 48 hydrolase genes a desirable target for primer design and molecular characterization.

In this paper we describe the enrichment of microbial communities from a thermophilic compost pile and provide an assessment of diversity in stable cellulolytic enrichments by addressing total bacterial diversity using the 16S rRNA gene as well as introducing a novel method to assess functional diversity in cellulolytic consortia by targeting glycosyl hydrolase family 48 (GHF48) genes.

MATERIALS AND METHODS

Sampling site. Compost samples were collected at the Middlebury College composting facility in May 2008. Samples were collected with a T-shaped steel coring tube (diameter, 2 cm) from specific locations in the compost pile where temperatures were determined to be the highest, 40 to 50 cm below the surface of the compost pile. Temperatures at each of these hot spots ranged from 52 to 72°C, and samples were designated CO-4, CO-5, and CO-6.

Enrichment protocol and media. Three samples of approximately 15 g of compost were used as inocula in bottles with 100 ml of mineral medium containing 1g of Avicel (PH-105; FMC Corp., Philadelphia, PA) as the carbon source. Bottles were flushed with nitrogen gas on site immediately after sampling and inoculation, thus ensuring strict anaerobic conditions from primary sampling. The primary enrichment (PE) medium consisted of the following, in grams

* Corresponding author. Mailing address: Thayer School of Engineering, 8000 Cummings Hall, Dartmouth College, Hanover, NH 03755. Phone: (603) 646-2231. Fax: (603) 646-2277. E-mail: lee.lynd@dartmouth.edu.

† Present address: Department of Biology, Northeastern University, Boston, MA 02115.

▽ Published ahead of print on 9 April 2010.

per liter: KH_2PO_4, 2.08; K_2HPO_4, 2.22; $MgCl_2 \times 6H_2O$, 0.1; NH_4Cl, 0.4; $CaCl_2 \times 2H_2O$, 0.05. Primary enrichment cultures were incubated at 55°C upon arrival at the laboratory and were transferred to maintenance (M) medium after 4 to 6 days of incubation. M medium consisted of 3 g of Avicel/liter, 1.04 g of KH_2PO_4/liter, 1.11 g of K_2HPO_4/liter, 2.5 g of $NaHCO_3$/liter, 0.2 g of $MgCl_2 \times 6H_2O$/liter, 0.4 g of NH_4Cl/liter, 0.05 g of $CaCl_2 \times 2H_2O$/liter, 0.05 g of $FeCl_2 \times 4H_2O$/liter, 1 ml of SL-10 trace elements/liter (1), 0.5 g of L-cysteine HCl/liter, and 4 ml of 0.025% resazurin/liter. Vitamins were added in the following concentrations (in milligrams per liter): pyridoxamine dihydrochloride, 0.8; *p*-aminobenzoic acid (PABA), 0.4; D-biotin, 0.2; vitamin B_{12}, 0.2; thiamine-HCl, 0.05; folic acid, 0.2; pantothenic acid calcium salt, 0.5; nicotinic acid, 0.5; pyridoxine-HCl, 0.1; thioctic acid, 0.5; riboflavin, 0.05. Phosphates and other minerals were prepared and autoclaved separately to avoid precipitation and chemical interactions during autoclaving. Vitamins were sterilized by filtration. A separate stock solution (100×) of L-cysteine HCl, $FeCl_2 \times 4H_2O$, $MgCl_2 \times 6H_2O$, NH_4Cl, and $CaCl_2 \times 2H_2O$ was flushed with N_2 immediately after dissolving and was autoclaved. Serum bottles containing sterile medium were placed in an anaerobic glove box, cooled down, mixed with a reducing agent solution, closed with sterile rubber stoppers, and capped with aluminum seals.

Batch cultivation in fermentors. After 10 consecutive transfers, anaerobic cultivation of each enrichment culture was carried out at 55°C in 1.6-liter Biostat Aplus fermentors (Sartorius Stedim, Gottingen, Germany), with a working volume of 1 liter and without a pH control. Fermentors were equipped with Norprene tubing (Cole Palmer Instrument Company, Vernon Hills, IL) to minimize oxygen diffusion. Cultures were grown in the same maintenance (M) medium, with 3 g liter^{-1} Avicel (PH-105; FMC Corp., Philadelphia, PA) and with reducing agents and vitamins added separately after the autoclaving of fermentor vessels. Cultures were stirred at 100 rpm and were sparged with 100 ml min^{-1} nitrogen gas. Each fermentation experiment was performed in duplicate and with a 2% (vol/vol) inoculum.

Analytical methods. Samples were taken every 3 h from each fermentation experiment for determination of the residual dry weight of cellulose, pH, pellet nitrogen, and fermentation products. Residual cellulose levels were determined by dry weight measurements, using triplicate 3-ml samples after filtration (0.22-μm-pore size Millipore Isopore GTTP filters), washing, and drying for 24 h at 60°C. The concentrations of fermentation products and soluble sugars (acetate, ethanol, formate, lactate, succinate, cellobiose, and glucose) were analyzed by high-performance liquid chromatography using an Aminex HPX-87H column (Bio-Rad, Hercules, CA). Pellet nitrogen was measured in centrifuged pellet samples by using a TOC-V combustion analyzer coupled with a TNM-1 total-nitrogen module (Shimadzu Corporation, Columbia, MD) and comparing the results to a 1-g liter^{-1} glycine standard.

DNA extraction and PCR amplification of the 16S rRNA gene. After 10 consecutive transfers using M medium with Avicel as the growth substrate, DNA was extracted from enrichment cultures using the ZR soil microbe DNA kit (Zymo Research Corp., Orange, CA) with minor modifications. Triplicate 1.5-ml samples from each enrichment were subjected to DNA extraction. Ballistic exposure times in a Mini-BeadBeater (Biospec, Bartlesville, OK) were increased to 5 min, and bead-beating speed was lowered to 2,500 rpm to maximize the DNA yield. The DNA obtained from each set of triplicate extractions was pooled for each enrichment. The extracted DNA was used in triplicate PCR amplifications targeting the 16S rRNA gene, using the universal oligonucleotide primers 8F and 1492R, designed to anneal to conserved regions of bacterial 16S rRNA genes (18). PCR amplifications were prepared with 10× buffer, 0.25 mM each deoxynucleoside triphosphate (dNTP), 2 U *Taq* polymerase (all from New England Biolabs), and 0.5 μM each primer (Integrated DNA Technologies, Coralville, IA) for final volumes of 30 μl, using 30 cycles of 94°C (30 s), 56°C (30 s), and 72°C (30 s) with an initial denaturation at 95°C (5 min) and a final extension at 72°C (5 min). As positive controls, cultures of *Clostridium thermocellum* DSM 1313, *Clostridium cellulolyticum* H10, and *Clostridium straminisolvens* DSM 16021 (kindly provided by Mascoma Corporation), *Clostridium clariflavum* DSM 19732 (obtained from the DSMZ), and *Clostridium thermocellum* ATCC 27405 were subjected to DNA extraction using the GenElute bacterial genomic DNA kit (Sigma, St. Louis, MO) according to the manufacturer's instructions.

Design of GHF48 primers and PCR amplification of GHF48 genes. A total of 35 sequences, representing all bacterial and fungal GH48 amino acid sequences, were aligned and edited using BioEdit, version 7.0.5 (11). An alignment of the clostridial sequences was further used in the primer design for amino acid positions Ile64 to Ala443, surrounding the catalytic domain of *C. thermocellum* CelS (10). Four sets of primers were tested, and PCR amplification was optimized using a PCR gradient with genomic DNA of *C. thermocellum* ATCC 27405, *C. thermocellum* DSM 1313, *C. straminisolvens*, *C. clariflavum*, *C. cellulollyticum*, and *C. acetobutylicum* as individual positive controls. The final protocol utilized two degenerate primers, GH48F (5′GGAATTCCKTAYCAYTCHATWGAAA-3′) and GH48R (5′-CATGCCTGVAWWCCRAACCA-3′). Optimal PCR conditions utilized 30 cycles of 94°C (30 s), 52°C (30 s), and 72°C (30 s), with an initial denaturation at 95°C (5 min) and a final extension at 72°C (5 min), and the same concentrations of reagents used for 16S rRNA gene PCR.

Clone library generation. Clone libraries based on the 16S rRNA gene and the GH48 family gene were constructed as previously described (14). Briefly, after the final round of PCR amplification, pooled amplified fragments were purified and cloned into a pGEM-T Easy vector (Promega, Madison, WI). Two vector-specific primers were used for the amplification of the DNA inserts from individual clones: pGemF (5′-GCA AGG CGA TTA AGT TGG G-3′) and pGemR (5′-ATG ACC ATG ATT ACG CCA AG-3′). PCR amplification used the same cycles used for 16S rRNA gene amplification. Clones were screened by agarose gel electrophoresis in order to detect those with inserts of the correct size. A clone library was created for each enrichment, and each library was screened for a total of 100 clones. Sequencing was performed by Agencourt Bioscience Corporation (Beverly, MA). The amplified clonal 16S rRNA and GHF48 gene sequences were aligned with representative sequences obtained from GenBank using the ClustalX software package (33). BioEdit, version 7.0.5, was used for manual editing of the sequences and, in the case of GH48, their translation into amino acid sequences (11). Rarefaction curves and operational taxonomic unit (OTU) definition at 99% sequence similarity were determined using the DOTUR software package (26). For the construction of phylogenetic trees, distances between pairs of nucleotide sequences were calculated based on the minimum-evolution criterion in the MEGA software package, version 3.1 (17). Bootstrap confidence values were obtained based on 1,000 replicates.

Nucleotide sequence accession numbers. The nucleotide sequences of both the 16S rRNA genes and glycosyl hydrolase family 48 genes from the clone libraries have been deposited in the GenBank database under accession numbers GQ265308 to GQ265351 and GQ487568 to GQ487569.

RESULTS

Dynamics of cellulose utilization and fermentation. After 8 to 10 consecutive transfers, anaerobic compost-derived enrichments CO-4, CO-5, and CO-6 exhibited consistent fermentation rates at 3 and 5 g liter^{-1} Avicel at 55°C. In batch fermentor experiments using 3 g liter^{-1} Avicel, between 86% and 97% of the cellulose was utilized in 60-h fermentations (Fig. 1A, C, and E). Cellulose utilization slowed down in all fermentations as the pH dropped from 8.4 to 6.5 in CO-4 and to 7.0 in CO-5 and CO6. Fermentation products accumulated between 20 and 40 h of incubation and consisted mostly of acetate, ethanol, and formate, with no residual sugars detected. Acetate and ethanol were the dominant products in the CO-5 and CO-6 fermentations, yielding between 0.4 and 0.5 g liter^{-1}. However, formate was equally dominant in CO-4 enrichment fermentations, at around 0.5 g liter^{-1}, with higher acetate and ethanol production. Biomass production, as evidenced by changes in pellet nitrogen levels, peaked between 30 and 40 h of incubation for the CO-5 and CO-6 enrichments at around 65 mg N liter^{-1}, while the CO-4 enrichment reached a maximum of 40 mg N liter^{-1} at around 30 h of incubation (Fig. 1B, D, and F). Based on the products measured, we were able to account for most of the carbon in fermented carbohydrates: at least 94.1% was accounted for in CO-4, 80.2% in CO-5, and 80.5% in CO-6.

16S rRNA gene diversity. Different levels of diversity were detected across enrichments at a sequence similarity cutoff of 99%. The CO-4 enrichment had the lowest diversity, with a total of 3 representative operational taxonomic units (OTUs) after 100 clones were surveyed. The CO-5 enrichment displayed slightly higher diversity, with a total of 9 OTUs, while CO-6 had the highest, with a total of 18 OTUs. Rarefaction analysis of 16S rRNA gene libraries (Fig. 2) showed that while

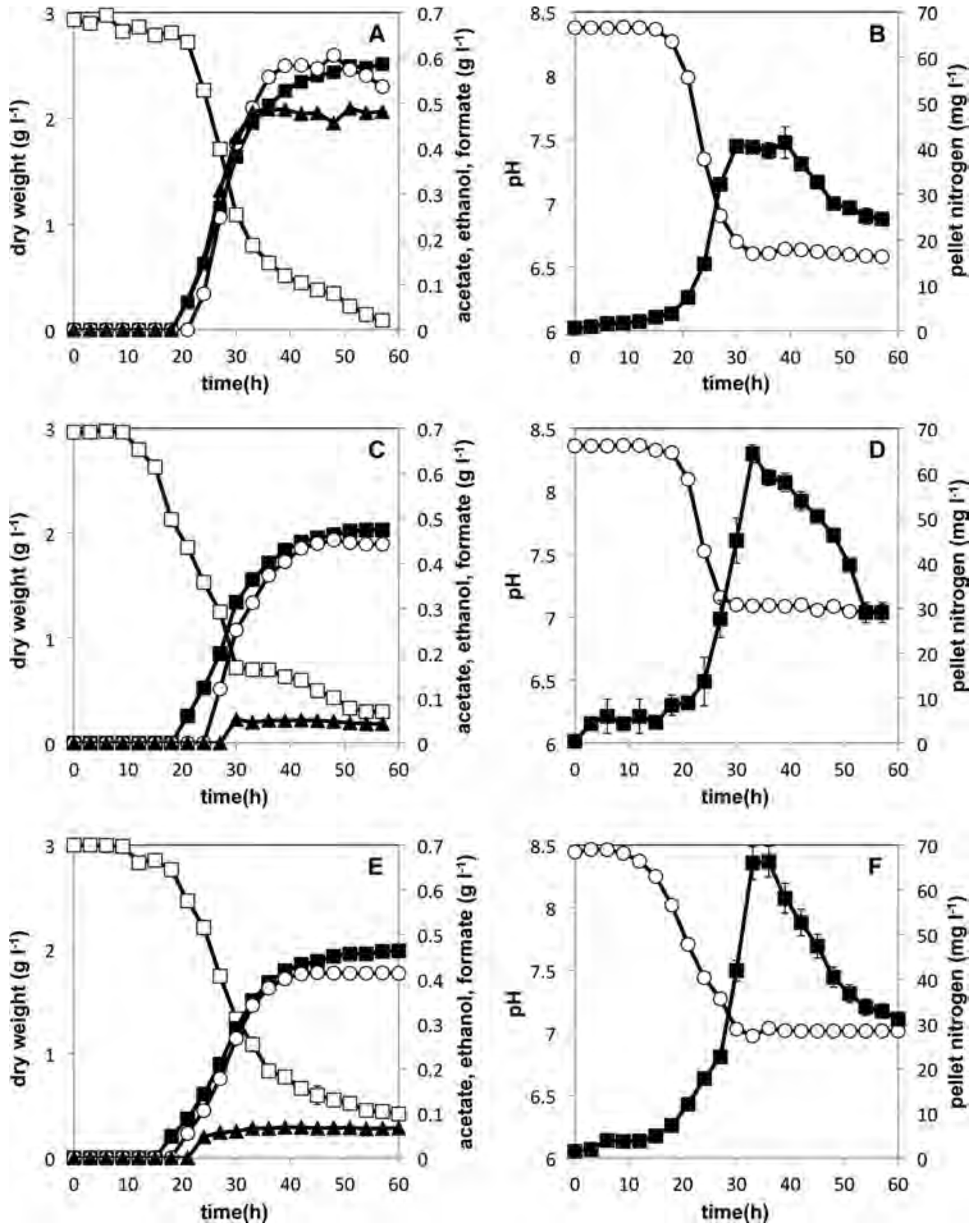

FIG. 1. Dynamics of residual dry weight of cellulose and product formation (A, C, and E) and changes in pH and total pellet N levels (B, D, and F) for the CO-4 (A and B), CO-5 (C and D), and CO-6 (E and F) enrichment cultures during 3-g liter^{-1} Avicel fermentations. (A, C, and E) Symbols: open squares, dry weight of cellulose; filled squares, acetate production; open circles, ethanol production; filled triangles, formate production. (B, D, and F) Symbols: open circles, changes in pH; filled squares, total pellet N. For the dry weight (of cellulose) and pellet N measurements, each data point represents the mean ± standard deviation calculated from triplicate samples.

the CO-4 and CO-5 enrichments had both attained a plateau in the number of retrievable OTUs, the CO-6 enrichment was still yielding new OTUs after 100 clones. However, additional sampling of 50 more clonal sequences from the CO-6 enrichment contributed only two new sequences as singletons (occurrence of 1 clone per OTU) and, given the repetition of previously found OTUs, reduced the slope of the curve (Fig. 2).

All of the 16S rRNA gene sequences retrieved from the CO-4, CO-5, and CO-6 enrichments belonged to the family *Clostridiaceae* (Fig. 3). The great majority of these sequences belonged to cluster III clostridia and were closely related to *Clostridium straminisolvens*, *Clostridium stercorarium*, and *Clostridium clariflavum*. However, differences in the distribution of sequences across clostridial clusters were also observed between enrichments. All OTUs from the CO-4 enrichment grouped within cluster III, as did 7 out of 8 OTUs from the CO-5 enrichment. However, not only did the CO-6 enrichment have the largest number of detected OTUs, but they were spread across clusters I, III, and XIV and a novel cluster of which the novel clostridium *Lutispora thermophila* (29) is the main representative.

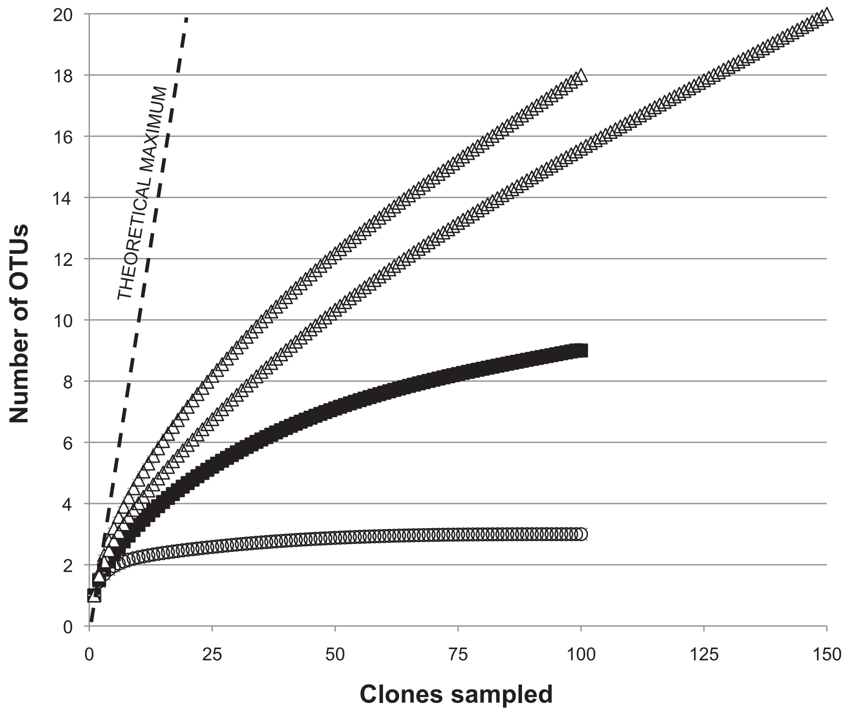

FIG. 2. Rarefaction curves for clone libraries constructed from the CO-4 (open circles), CO-5 (filled squares), and CO6 (open triangles) enrichments. The two curves constructed for the CO-6 enrichment represent two calculations, one using 100 clones and one using 150 clones.

Clostridium straminisolvens-like sequences were consistently found in all three enrichments, although in different proportions. The most abundant 16S rRNA gene sequences retrieved in the CO-5 and CO-6 enrichments accounted for 68% and 45% of each clone library, respectively, and were almost identical to *Clostridium straminisolvens*. On the other hand, the most abundant OTU in the CO-4 enrichment was most similar to *Clostridium clariflavum* (previously sp. EBR-02E-0045 [GenBank accession no. AB186359]), obtained from a thermophilic methanogenic reactor (28, 30). Underscoring the differences between the CO-4 enrichment and the others, this particular sequence was not retrieved in the libraries constructed for the CO-5 and CO-6 enrichments.

In addition, a number of sequences grouped most closely to novel clostridia. Two OTUs from the CO-6 enrichment, representing 1% to 3% of this clone library each, were determined to be most similar to the newly isolated *Lutispora thermophila*, which is a noncellulolytic organism from a thermophilic methanogenic reactor (29). Likewise, a minor component of the CO-5 enrichment was most closely related to a proposed novel clostridium within cluster XIV, *Clostridium islandicum* (GenBank accession no. EF088328). A larger number of sequences in the CO-6 enrichment (representing 1% to 24% of this particular clone library) did not have any close cultured representatives, although they fell within cluster III and seemed most similar to *C. stercorarium*.

PCR amplification of glycosyl hydrolase family 48 gene sequences. Primers targeting a structural region essential to the function of family 48 glycosyl hydrolases were prepared based on translated amino acid sequences of *C. thermocellum* and *C. cellulolyticum* controls. This region includes the acidic residue responsible for protonating the glycosidic oxygen, identified as Glu87 in *C. thermocellum* CelS (10) and Glu55 in *Clostridium cellulolyticum* CelF (21); the Trp184, Asp255, and Tyr351 residues, associated with the water nucleophile involved in the catalytic mechanism of *Clostridum thermocellum* CelS (10); and the Asp230, Tyr299, Trp310, and Trp312 residues in the catalytic tunnel of *C. cellulolyticum* CelF (21).

An alignment of the PCR-amplified fragment with all the other sequences of GHF48 genes available in GenBank as of August 2009 and subsequent phylogenetic analyses revealed a grouping similar to the clustering observed in whole-gene alignments (Fig. 4). Sequences from the *Fungi* and the classes *Clostridia*, *Actinobacteria*, and *Bacilli* formed clear clusters, with the class *Clostridia* as the most divergent taxon. Known representatives of the *Gammaproteobacteria* (*Hahella chejuensis*) and the *Chloroflexi* (*Herpetosiphon aurantiacus*) each formed distinct, separate branches adjacent to the *Actinobacteria* cluster. A single exception to the class-based clustering was presented by the Cel48 sequence from *Myxobacter* sp. AL-1 (a deltaproteobacterium), which grouped with sequences from the *Bacilli*. It should also be noted that the GHF48 gene of *Ruminococcus albus* (*cel48A*) is distinctly different from those of all other glycosyl hydrolases from this family, including fellow clostridia and fellow member of the *Ruminococcaceae* family *Bacteroides cellulosolvens*.

Amplicons obtained from *C. thermocellum* ATCC 27405 and *C. thermocellum* DSM 1313 required further cloning in order to separate coamplified copies of the *celY* and *celS* genes prior to sequencing. For both strains, PCR amplification produced an even mixture (approximately 50:50) of the two genes, as observed in randomly selected clones used for amplification,

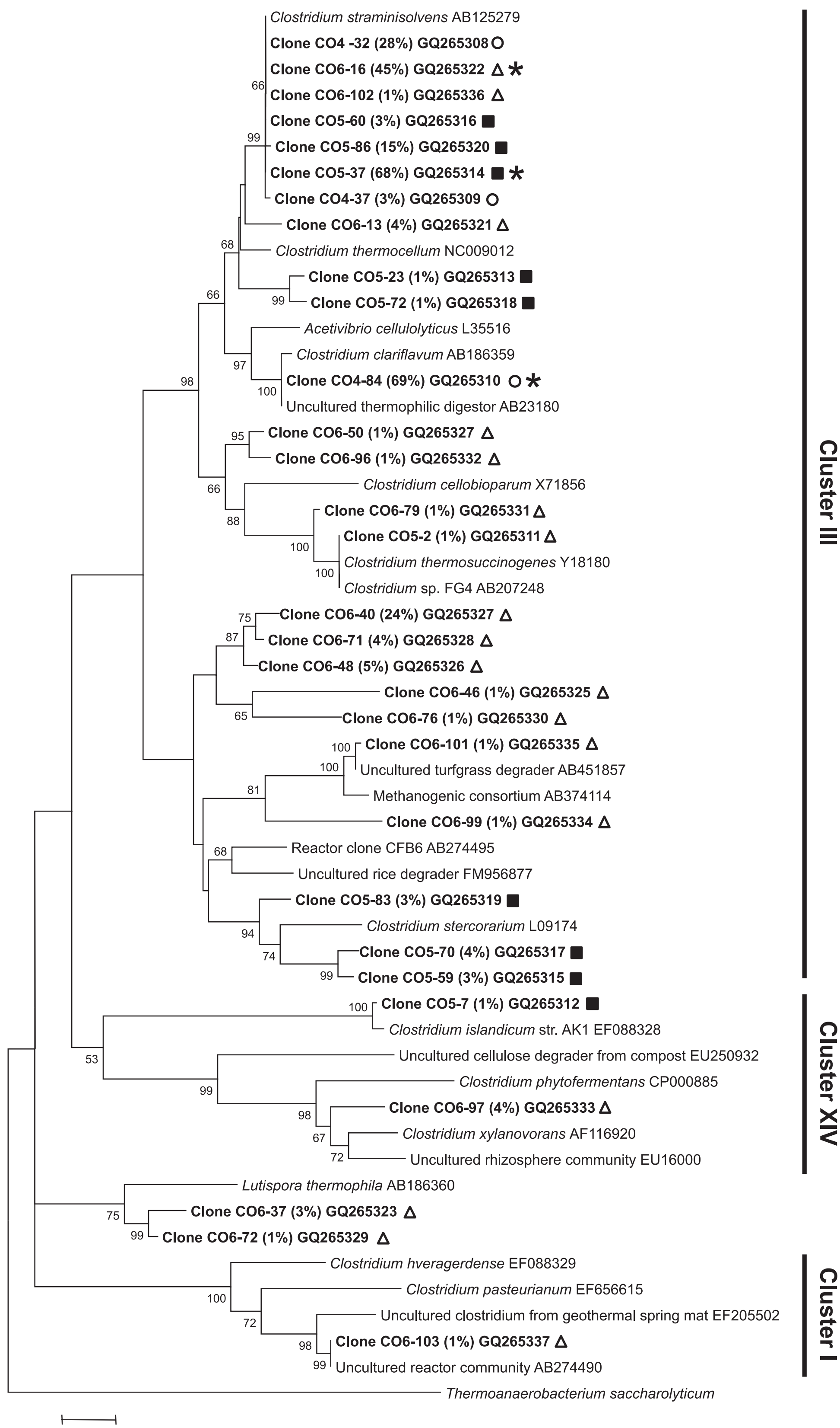

FIG. 3. Phylogenetic tree of 16S rRNA gene diversity in cellulolytic enrichments. Phylogenetic relationships were inferred by minimum evolution analysis of nucleotide sequences obtained from clone libraries from the CO-4 (open circles), CO-5 (filled squares), and CO6 (open triangles) enrichments. Bootstrap values are shown for 1,000 replicates. To indicate specific abundance, the percentage of each clone's contribution to its library is specified in parentheses. The most abundant clones in each library are marked with asterisks.

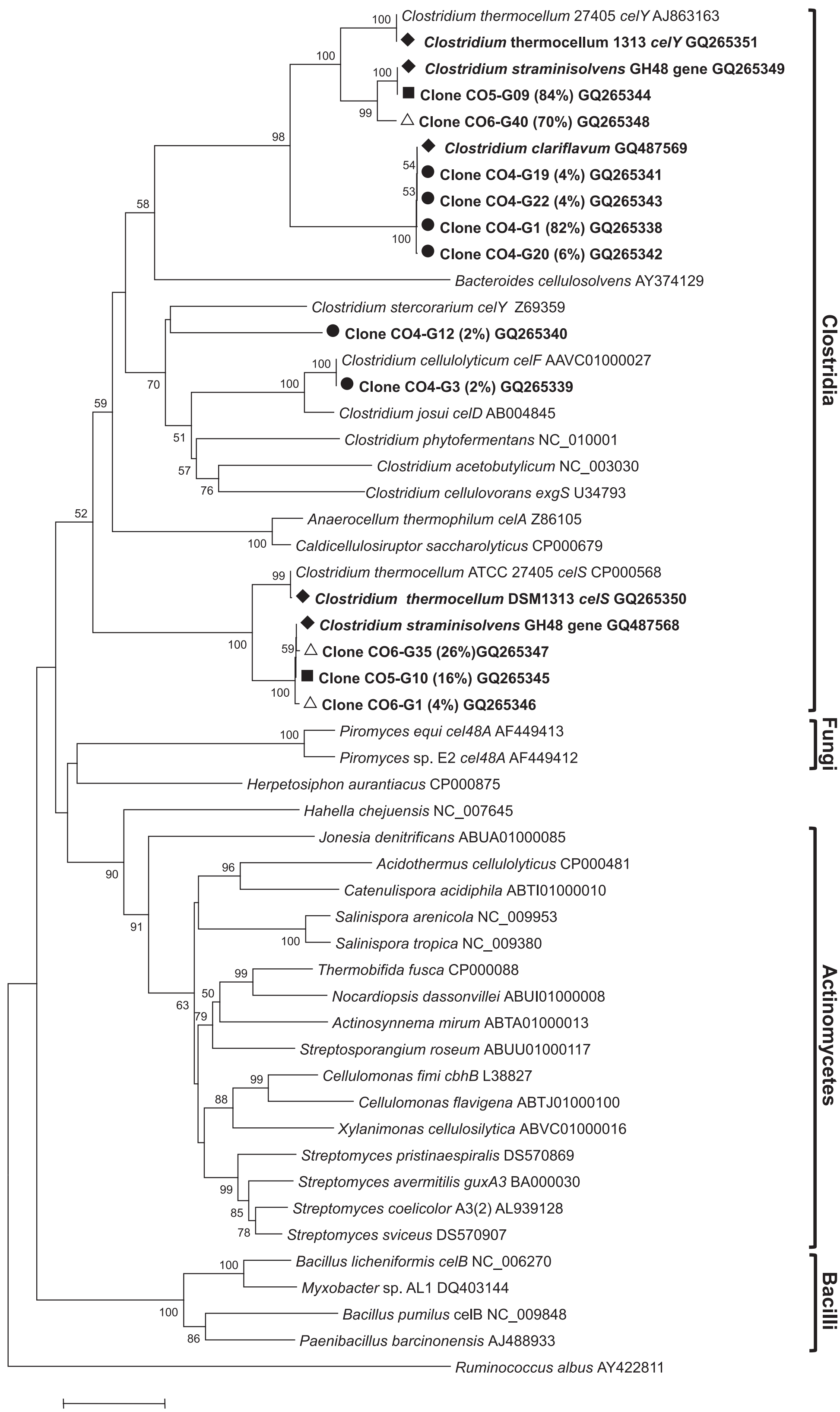

FIG. 4. Phylogenetic tree of glycosyl hydrolase family 48 (GHF48) genes in cellulolytic enrichments. Phylogenetic relationships were inferred by minimum-evolution analysis of nucleotide sequences obtained from clone libraries from the CO-4 (filled circles), CO-5 (filled squares), and CO6 (open triangles) enrichments. Controls using type cultures sequenced in-house are labeled with filled diamonds. Bootstrap values are shown for 1,000 replicates. To indicate specific abundance, the percentage of each clone's contribution to its library is specified in parentheses.

thus demonstrating the ability of these primers to amplify both cellulosomal (*celS*) and noncellulosomal (*celY*) copies of GHF48 genes present in the same organism.

Diversity of family 48 glycosyl hydrolases in enrichments. All sequences retrieved from GHF48 gene libraries from enrichments belonged to the clostridia (Fig. 4). As a reference, we amplified the GHF48 genes in *Clostridium straminisolvens* and *Clostridium clariflavum*, given their prevalence in 16S rRNA libraries. Two distinctly different sequences obtained from *C. straminisolvens* were novel GHF48 gene fragment sequences, with 84% sequence similarity to *Clostridium thermocellum celY* and 89% sequence similarity to *Clostridium thermocellum celS*, respectively. Unlike *C. thermocellum*, where the *celS*/*celY* sequence retrieval ratio was close to 50:50, in *C. straminisolvens* the ratio favored *celY*-like sequences, with a *celS*/*celY* sequence retrieval ratio close to 30:70. Similarly, dominant sequences in the CO-5 and CO-6 enrichments (representing 70% to 84% of the clone libraries) were most similar to the novel *Clostridium straminisolvens celY*-like GHF48 gene reported here. The remaining GHF48 gene sequences from the CO-5 and CO6 enrichments, accounting for 16% and 30% of each clone library, were almost identical to the *celS*-like gene identified in *C. straminisolvens.* The dominant sequences from the CO-4 enrichment and a novel GHF48 gene sequence from *Clostridium clariflavum* grouped very closely together (>99% sequence similarity), with only 78% amino acid sequence similarity to their closest match found in *C. thermocellum celY*. Although they were retrieved in much lesser proportions, one sequence from the CO-4 enrichment was identical to *C. cellulolyticum celF*, and one was most closely related to *C. stercorarium celY*. Aside from the CO-4 GHF48 gene sequence that was identical to *C. cellulolyticum celF*, all the other GHF48 gene sequences reported here (including those of the *Clostridium straminisolvens* and *Clostridium clariflavum* type cultures) represent novel cellulase genes.

DISCUSSION

Anaerobic enrichments from compost with cellulose as the sole carbon source facilitate the growth of consortia with various levels of community complexity and fermentation capabilities. By examining both the 16S rRNA and the glycosyl hydrolase family 48 gene composition, we have been able to characterize these communities and their dominant members at both the taxonomic and the functional level.

The breadth of 16S rRNA gene diversity in our enrichments, beyond the dominant sequences closely related to *Clostridium straminisolvens* and *Clostridium clariflavum*, opens the question of what other functional roles are played by other clostridia within these communities. This is particularly true for enrichments resulting from multiple transfers such as those described here, where the presence of noncellulolytic clostridia has not been diluted after several transfers with cellulose as the sole carbon source. Of particular interest are the uncultivated members with moderately high abundances (such as 16S rRNA gene clone CO6-40), where the isolation of pure cultures from a mixed consortium would enable further understanding of the different functional roles within both simple and complex cellulolytic communities. Obtaining stable cellulolytic enrichments with various levels of diversity and their taxonomic characterization are the first steps toward this goal. Additionally, the development of a molecular marker targeting an essential functional gene for these enrichments, as is the case for glycosyl hydrolase family 48 genes, further expands the possibilities for future comparisons between communities and community members from a transcriptomic point of view. However, it should be noted that fermentation experiments revealed subtle yet important differences in product and cell biomass formation. The CO-4 enrichment, dominated by *C. clariflavum*-like sequences, produces comparatively larger amounts of formate with lower overall cell biomass production and is able to continue cellulose utilization beyond pH 7.0 to a final pH of 6.5. This is consistent with previous observations for *C. clariflavum* DSM 19732, where formate production is evident, in contrast to the findings for its closest relatives, *C. thermocellum* and *C. straminisolvens* (30). On the other hand, the more diverse CO-5 (9 OTUs) and CO-6 (18 OTUs) enrichments are characterized by similar product profiles dominated by acetate and ethanol, higher cell biomass formation, and a final pH of 7.0, indicating that the *C. straminisolvens* dominance, and not the total diversity or community structure, may be an important factor in the ability of these communities to utilize cellulose.

Glycosyl hydrolase family 48 genes appear to be promising targets as functional marker genes for environments in which active anaerobic cellulose degradation occurs. Our work confirms the consistency in GHF48 gene taxonomic grouping and demonstrates substantial scope for expanding known exoglucanase diversity. The phylogenetic clustering observed with the primers used in this study, as with whole GHF48 gene sequences, revealed clear taxonomic clustering for representative members of the *Clostridia*, *Actinobacteria*, *Bacilli*, *Gammaproteobacteria*, *Chloroflexi*, and *Fungi* (Fig. 4). The exceptional grouping of *Myxobacter* sp. AL-1, a deltaproteobacterium, with sequences from the *Bacilli* is interesting from an evolutionary point of view but is not unexpected. Similarities between *Myxobacter* sp. AL-1 and different members of the *Bacilli* (in particular *Bacillus licheniformis*), in terms of unique structural similarities in family 48 glycosyl hydrolases (25) but also in terms of sequence similarities of other types of cellulases (22), have been reported previously.

Analysis based on both 16S rRNA and GHF48 gene sequences indicated that *Clostridium straminisolvens*- and *Clostridium clariflavum*-like organisms were major components in the microbial communities present in the enrichments studied. This paper is the first report of *Clostridium straminisolvens* and *Clostridium clariflavum* possessing glycosyl hydrolase family 48 genes, and *Clostridium straminisolvens* is the first organism after *C. thermocellum* to possess two distinctly different glycosyl hydrolase family 48 genes (NCBI accession no. GQ265349 and GQ487568). We also found that these genes are closely associated with the noncellulosomal *celY* gene and the cellulosomal *celS* gene in *C. thermocellum.* Given their close association with *C. thermocellum* counterparts, we speculate that the CelY-like cellulase in *C. straminisolvens* may act as a noncellulosomal exoglucanase in synergy with glycosyl hydrolase family 9 endoglucanases, as has been reported for *C. thermocellum* and *C. stercorarium* (2). Likewise, the presence of a CelY-like gene in *C. clariflavum* may indicate a similar mechanism. Further exploration of adjacent domains will be useful not only for determining the presence of cellulose-binding domains and modes of action but also for gaining a better understanding of how conserved this family might actually be and whether horizontal gene transfer plays a role in the observed diversity in GHF48 genes from environmental strains.

A variety of novel GHF48 gene sequences have been retrieved from the cellulolytic enrichments described in this study. Aside from the dominant *C. straminisolvens* and *C. clariflavum* sequences, two sequences associated with *C. cellulolyticum* and *C. stercorarium* were also retrieved from cellulolytic enrichments and may belong to minor cellulolytic members of the community. It should, however, be noted that the mere presence of this gene does not necessarily confer functionality. For example, a GHF48 gene present in *C. acetobutylicum*, although it is expressed, does not confer cellulolytic activity on this organism (24).

With a variety of metagenomic studies addressing the diversity of glycosyl hydrolases in different environments (3, 9, 27), only one previous effort has targeted the diversity of cellulases beyond cultivated representatives by focusing on a specific family, describing the diversity and abundance of GHF5 in aquatic environments (8). More quantitative analyses of the relevance of GHF48 genes in cellulolytic microbial communities are needed, given the low abundance of these particular cellulases reported in metagenomic studies of complex cellulolytic microbial communities (3), although it is possible that their presence in mostly single copies per genome has contributed to the lack of detection in metagenomic studies. Given their prevalence in transcriptomic and proteomic analyses of model organisms (23, 32), it should be of interest to observe what metatranscriptomic analyses of both complex and simple cellulolytic communities reveal. The development of molecular tools for the detection of GHF48 genes in enrichments from environmental samples, as described here, offers an initial step toward this goal.

ACKNOWLEDGMENTS

This research was supported by a grant from the BioEnergy Science Center (BESC), Oak Ridge National Laboratory, a U.S. Department of Energy (DOE) Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

We thank Melissa Beckwith and Norm Cushman for providing us with access to their composting facility at Middlebury College, Vermont. We also thank Mascoma Corporation for kindly providing type cultures of *Clostridium thermocellum* DSM 1313, *Clostridium cellulolyticum*, and *Clostridium straminisolvens*.

REFERENCES

1. **Atlas, R. M.** 1996. Handbook of microbiological media, 2nd ed. CRC Press LLC, Boca Raton, FL.
2. **Berger, E., D. Zhang, V. V. Zverlov, and W. H. Schwarz.** 2007. Two noncellulosomal cellulases of *Clostridium thermocellum*, Cel9I and Cel48Y, hydrolyse crystalline cellulose synergistically. FEMS Microbiol. Lett. **268:**194–201.
3. **Brulc, J. M., D. A. Antonopoulos, M. E. B. Miller, M. K. Wilson, A. C. Yannarell, E. A. Dinsdale, R. E. Edwards, E. D. Frank, J. B. Emerson, P. Wacklin, P. M. Coutinho, B. Henrissat, K. E. Nelson, and B. A. White.** 2009. Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. Proc. Natl. Acad. Sci. U. S. A. **106:**1948–1953.
4. **Burrell, P. C., C. O'Sullivan, H. Song, W. P. Clarke, and L. L. Blackall.** 2004. Identification, detection, and spatial resolution of *Clostridium* populations responsible for cellulose degradation in a methanogenic landfill leachate bioreactor. Appl. Environ. Microbiol. **70:**2414–2419.
5. **Devillard, E., D. B. Goodheart, S. K. Karnati, E. A. Bayer, R. Lamed, J. Miron, K. E. Nelson, and M. Morrison.** 2004. *Ruminococcus albus* 8 mutants defective in cellulose degradation are deficient in two processive endocellulases, Cel48A and Cel9B, both of which possess a novel modular architecture. J. Bacteriol. **186:**136–145.
6. **Ding, S. Y., W. S. Adney, T. B. Vinzant, S. R. Decker, J. O. Baker, S. R. Thomas, and M. E. Himmel.** 2003. Glycoside hydrolase gene cluster of *Acidothermus cellulolyticus*, p. 332–360. *In* S. D. Mansfield and J. N. Saddler (ed.), Applications of enzymes to lignocellulosics. ACS symposium series, vol. 855. American Chemical Society, Washington, DC.
7. **Dror, T. W., E. Morag, A. Rolider, E. A. Bayer, R. Lamed, and Y. Shoham.** 2003. Regulation of the cellulosomal *celS* (*cel48A*) gene of *Clostridium thermocellum* is growth rate dependent. J. Bacteriol. **185:**3042–3048.
8. **Elifantz, H., L. A. Waidner, V. K. Michelou, M. T. Cottrell, and D. L. Kirchman.** 2008. Diversity and abundance of glycosyl hydrolase family 5 in the North Atlantic Ocean. FEMS Microbiol. Ecol. **63:**316–327.
9. **Ferrer, M., O. V. Golyshina, T. N. Chernikova, A. N. Khachane, D. Reyes-Duarte, V. A. P. M. Dos Santos, C. Strompl, K. Elborough, G. Jarvis, A. Neef, M. M. Yakimov, K. N. Timmis, and P. N. Golyshin.** 2005. Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. Environ. Microbiol. **7:**1996–2010.
10. **Guimarães, B. G., H. Souchon, B. L. Lytle, J. H. D. Wu, and P. M. Alzari.** 2002. The crystal structure and catalytic mechanism of cellobiohydrolase CelS, the major enzymatic component of the *Clostridium thermocellum* cellulosome. J. Mol. Biol. **320:**587–596.
11. **Hall, T. A.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. **41:**95–98.
12. **Haruta, S., Z. Cui, Z. Huang, M. Li, M. Ishii, and Y. Igarashi.** 2002. Construction of a stable microbial community with high cellulose-degradation ability. Appl. Microbiol. Biotechnol. **59:**529–534.
13. **Irwin, D. C., S. Zhang, and D. B. Wilson.** 2000. Cloning, expression and characterization of a family 48 exocellulase, Cel48A, from *Thermobifida fusca*. Eur. J. Biochem. **267:**4988–4997.
14. **Izquierdo, J. A., and K. Nüsslein.** 2006. Distribution of extensive *nifH* gene diversity across physical soil microenvironments. Microb. Ecol. **51:**441–452.
15. **Kakiuchi, M., A. Isui, K. Suzuki, T. Fujino, E. Fujino, T. Kimura, S. Karita, K. Sakka, and K. Ohmiya.** 1998. Cloning and DNA sequencing of the genes encoding *Clostridium josui* scaffolding protein CipA and cellulase CelD and identification of their gene products as major components of the cellulosome. J. Bacteriol. **180:**4303–4308.
16. **Kato, S., S. Haruta, Z. J. Cui, M. Ishii, and Y. Igarashi.** 2004. Effective cellulose degradation by a mixed-culture system composed of a cellulolytic *Clostridium* and aerobic non-cellulolytic bacteria. FEMS Microbiol. Ecol. **51:**133–142.
17. **Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief. Bioinform. **5:**150–163.
18. **Lane, D. J.** 1991. 16S/23S rRNA sequencing, p. 115–175. *In* E. Stackebrandt and M. Goodfellow (ed.), Nucleic acid techniques in bacterial systematics. Wiley & Sons, Chichester, United Kingdom.
19. **Lynd, L. R., P. J. Weimer, W. H. van Zyl, and I. S. Pretorius.** 2002. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol. Mol. Biol. Rev. **66:**506–577.
20. **Ng, T. K., A. Ben-Bassat, and J. G. Zeikus.** 1981. Ethanol production by thermophilic bacteria: fermentation of cellulosic substrates by cocultures of *Clostridium thermocellum* and *Clostridium thermohydrosulfuricum*. Appl. Environ. Microbiol. **41:**1337–1343.
21. **Parsiegla, G., M. Juy, C. Reverbel-Leroy, C. Tardif, A. Belaıch, H. Driguez, and R. Haser.** 1998. The crystal structure of the processive endocellulase CelF of *Clostridium cellulolyticum* in complex with a thiooligosaccharide inhibitor at 2.0 angstrom resolution. EMBO J. **17:**5551–5562.
22. **Pedraza-Reyes, M., and F. Gutiérrez-Corona.** 1997. The bifunctional enzyme chitosanase-cellulase produced by the gram-negative microorganism *Myxobacter* sp. AL-1 is highly similar to *Bacillus subtilis* endoglucanases. Arch. Microbiol. **168:**321–327.
23. **Reverbel-Leroy, C., S. Pages, A. Belaıch, J. P. Belaıch, and C. Tardif.** 1997. The processive endocellulase CelF, a major component of the *Clostridium cellulolyticum* cellulosome: purification and characterization of the recombinant form. J. Bacteriol. **179:**46–52.
24. **Sabathe, F., A. Belaich, and P. Soucaille.** 2002. Characterization of the cellulolytic complex (cellulosome) of *Clostridium acetobutylicum*. FEMS Microbiol. Lett. **217:**15–22.
25. **Sánchez, M. M., F. I. J. Pastor, and P. Diaz.** 2003. Exo-mode of action of cellobiohydrolase Cel48C from *Paenibacillus* sp. BP-23. A unique type of cellulase among *Bacillales*. Eur. J. Biochem. **270:**2913–2919.
26. **Schloss, P. D., and J. Handelsman.** 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl. Environ. Microbiol. **71:**1501–1506.
27. **Schlüter, A., T. Bekel, N. N. Diaz, M. Dondrup, R. Eichenlaub, K. H. Gartemann, I. Krahn, L. Krause, H. Kromeke, O. Kruse, J. H. Mussgnug, H. Neuweger, K. Niehaus, A. Puhler, K. J. Runte, R. Szczepanowski, A. Tauch, A. Tilker, P. Viehover, and A. Goesmann.** 2008. The metagenome of a biogas-producing microbial community of a production-scale biogas plant

fermenter analysed by the 454-pyrosequencing technology. J. Biotechnol. **136:**77–90.

28. **Shiratori, H., H. Ikeno, S. Ayame, N. Kataoka, A. Miya, K. Hosono, T. Beppu, and K. Ueda.** 2006. Isolation and characterization of a new *Clostridium* sp. that performs effective cellulosic waste digestion in a thermophilic methanogenic bioreactor. Appl. Environ. Microbiol. **72:**3702–3709.
29. **Shiratori, H., H. Ohiwa, H. Ikeno, S. Ayame, N. Kataoka, A. Miya, T. Beppu, and K. Ueda.** 2008. *Lutispora thermophila* gen. nov., sp. nov., a thermophilic, spore-forming bacterium isolated from a thermophilic methanogenic bioreactor digesting municipal solid wastes. Int. J. Syst. Evol. Microbiol. **58:**964–969.
30. **Shiratori, H., K. Sasaya, H. Ohiwa, H. Ikeno, S. Ayame, N. Kataoka, A. Miya, T. Beppu, and K. Ueda.** 2009. *Clostridium clariflavum* sp. nov. and *Clostridium caenicola* sp. nov., moderately thermophilic, cellulose-/cellobiose-digesting bacteria isolated from methanogenic sludge. Int. J. Syst. Evol. Microbiol. **59:**1764–1770.
31. **Steenbakkers, P. J. M., A. Freelove, B. van Cranenbroek, B. M. C. Sweegers, H. R. Harhangi, G. D. Vogels, G. P. Hazlewood, H. J. Gilbert, and H. J. M. Op den Camp.** 2002. The major component of the cellulosomes of the anaerobic fungi from the genus *Piromyces* is a family 48 glycosyl hydrolase. DNA Sequence **13:**313–320.
32. **Stevenson, D. M., and P. J. Weimer.** 2005. Expression of 17 genes in *Clostridium thermocellum* ATCC 27405 during fermentation of cellulose or cellobiose in continuous culture. Appl. Environ. Microbiol. **71:**4672–4678.
33. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.
34. **Weimer, P. J., and J. G. Zeikus.** 1977. Fermentation of cellulose and cellobiose by *Clostridium thermocellum* in the absence and presence of *Methanobacterium thermoautotrophicum*. Appl. Environ. Microbiol. **33:**289–297.