

Genome Sequence of the Cellulolytic Gliding Bacterium *Cytophaga hutchinsonii*^{∇†}

Gary Xie,^{1,2} David C. Bruce,^{1,2} Jean F. Challacombe,^{1,2} Olga Chertkov,^{1,2} John C. Detter,^{1,2}
Paul Gilna,^{1,2} Cliff S. Han,^{1,2} Susan Lucas,^{2,4} Monica Misra,¹ Gerald L. Myers,¹
Paul Richardson,^{2,5} Roxanne Tapia,^{1,2} Nina Thayer,¹ Linda S. Thompson,^{1,2}
Thomas S. Brettin,^{1,2} Bernard Henrissat,³ David B. Wilson,⁶
and Mark J. McBride^{7*}

Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico¹; Department of Energy Joint Genome Institute, Walnut Creek, California²; Architecture et Fonction des Macromolécules Biologiques, UMR 6098, Centre National de la Recherche Scientifique, and Universities of Aix-Marseille I and II, Marseille, France³; Lawrence Livermore National Laboratory, Livermore, California⁴; Lawrence Berkeley National Laboratory, Berkeley, California⁵; Cornell University, Ithaca, New York⁶; and University of Wisconsin—Milwaukee, Milwaukee, Wisconsin⁷

Received 29 January 2007/Accepted 23 March 2007

The complete DNA sequence of the aerobic cellulolytic soil bacterium *Cytophaga hutchinsonii*, which belongs to the phylum *Bacteroidetes*, is presented. The genome consists of a single, circular, 4.43-Mb chromosome containing 3,790 open reading frames, 1,986 of which have been assigned a tentative function. Two of the most striking characteristics of *C. hutchinsonii* are its rapid gliding motility over surfaces and its contact-dependent digestion of crystalline cellulose. The mechanism of *C. hutchinsonii* motility is not known, but its genome contains homologs for each of the *gld* genes that are required for gliding of the distantly related bacteroidete *Flavobacterium johnsoniae*. *Cytophaga-Flavobacterium* gliding appears to be novel and does not involve well-studied motility organelles such as flagella or type IV pili. Many genes thought to encode proteins involved in cellulose utilization were identified. These include candidate endo- β -1,4-glucanases and β -glucosidases. Surprisingly, obvious homologs of known cellobiohydrolases were not detected. Since such enzymes are needed for efficient cellulose digestion by well-studied cellulolytic bacteria, *C. hutchinsonii* either has novel cellobiohydrolases or has an unusual method of cellulose utilization. Genes encoding proteins with cohesin domains, which are characteristic of cellulosomes, were absent, but many proteins predicted to be involved in polysaccharide utilization had putative D5 domains, which are thought to be involved in anchoring proteins to the cell surface.

Cellulose is probably the most abundant biopolymer on earth (5). The β -1,4-linked glucose chains of cellulose form highly ordered crystalline fibrils that are insoluble and relatively recalcitrant to degradation. Cellulolytic bacteria produce a suite of enzymes that act synergistically to convert crystalline cellulose to glucose (38, 62). The cellulolytic bacteria that have been studied most use either of two strategies to digest cellulose. Some, such as aerobic actinomycetes belonging to the genera *Streptomyces* and *Thermobifida*, secrete soluble extracellular cellulolytic enzymes that usually contain carbohydrate binding modules (CBMs) that anchor them to the substrate (71). These enzymes attack the insoluble substrate and release glucose, cellobiose, and short oligomers that are taken up and metabolized by the cells. In contrast, anaerobic bacteria of the genera *Clostridium* and *Ruminococcus* produce multiprotein complexes, called cellulosomes, that often remain attached to cells (7, 9, 19, 50, 62). Cellulosomes contain cellulose-binding proteins, enzymes involved in the hydrolysis of cellulose and of

other polysaccharides, and scaffolding proteins that hold the multiprotein complex together. Cellulosomes often anchor the bacteria to their substrate and result in localized release of sugars that are taken up by the cells.

Cytophaga hutchinsonii (Fig. 1), the type species of the genus *Cytophaga*, is an abundant aerobic cellulolytic soil bacterium (37, 55, 64, 72). *C. hutchinsonii* is a gram-negative bacterium that belongs to the phylum *Bacteroidetes* (also known as the *Cytophaga-Flavobacterium-Bacteroides* group). *C. hutchinsonii* utilizes very few substrates as sole carbon and energy sources. Besides cellulose, its only known substrates are cellobiose and glucose (37). Wild strains generally use these soluble sugars poorly and exhibit a preference for crystalline cellulose. *C. hutchinsonii* requires direct contact with cellulose for efficient digestion, and most of the cellulolytic enzymes appear to be cell associated (16, 47, 64). Another unusual feature of cellulose degradation by *C. hutchinsonii* is that reducing sugars such as glucose and cellobiose do not accumulate in the medium when it digests cellulose (64). This is probably not simply the result of efficient uptake and metabolism of these sugars, since incubation of cells under anaerobic conditions, which should interfere with these processes, does not result in accumulation of reducing sugars.

C. hutchinsonii was selected for genome sequencing for several reasons. First, biochemical and physiological studies

* Corresponding author. Mailing address: Department of Biological Sciences, University of Wisconsin—Milwaukee, 3209 N. Maryland Ave., 181 Lapham Hall, Milwaukee, WI 53211. Phone: (414) 229-5844. Fax: (414) 229-3926. E-mail: mcbride@uwm.edu.

† Supplemental material for this article may be found at <http://aem.asm.org/>.

[∇] Published ahead of print on 30 March 2007.

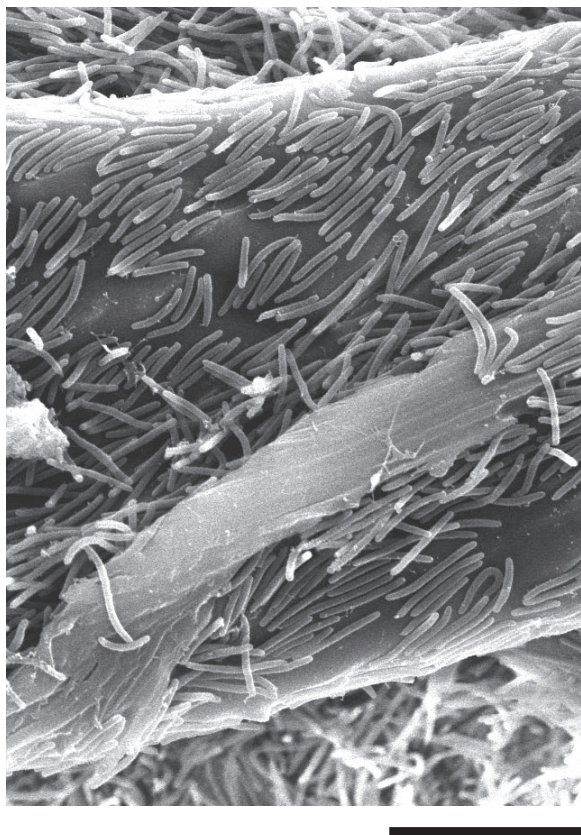


FIG. 1. Scanning electron micrograph of *C. hutchinsonii* cells digesting cellulose filter paper. Bar, 10 μ m.

suggested that it might use a novel strategy for cellulose utilization. Second, techniques to genetically manipulate *C. hutchinsonii* are available (44). Most cellulolytic bacteria have resisted genetic analysis, so the development of these techniques promises new insights into bacterial cellulose utilization. Third, *C. hutchinsonii* exhibits a form of rapid gliding motility whose mechanism remains unexplained despite decades of study (43, 64, 72). Gliding may help to facilitate cellulose digestion, since gliding cells align themselves with and move along cellulose fibers as they digest them (64, 72). Finally, few genome sequences have been reported for members of the phylum *Bacteroidetes*, and none are closely related to *C. hutchinsonii*. This paper highlights novel features of the *C. hutchinsonii* genome, with particular emphasis on genes and proteins likely to be involved in cellulose utilization and gliding motility.

MATERIALS AND METHODS

Sequencing of the *C. hutchinsonii* genome. The random shotgun method was used to sequence the genome of *C. hutchinsonii* ATCC 33406. Large (40 kb)-, medium (8 kb)-, and small (3 kb)-insert random libraries were partially sequenced, with an average success rate of 90% and an average high-quality read length of 634 nucleotides. Sequences were assembled with parallel phrap (High Performance Software, LLC). Possible misassemblies were corrected with Dupfinisher (C. F. Han and P. Chain, presented at the 2006 International Conference on Bioinformatics and Computational Biology, Las Vegas, NV, 26 to 29 June 2006) or by analysis of transposon insertions in bridge clones. Gaps between contigs were closed by editing, custom primer walking, or PCR amplification. The completed genome sequence of *C. hutchinsonii* contains 105,896 reads,

achieving an average of 15-fold sequence coverage per base, with an error rate of <1 in 100,000.

Annotation. Gene predictions were obtained using Glimmer, and tRNAs were identified using tRNAscan-SE. Basic analyses of the gene predictions were performed by comparing coding sequences against the Pfam, BLOCKS, and Prodom databases. Protein localizations were predicted with PSORTb (21), and lipoproteins were identified using LipoP (33). A team of annotators added gene definitions and functional classes, using BLAST results and information from the Pfam (63; <http://pfam.janelia.org/index.html>), BLOCKS (26), Prodom (59), and SMART (58) databases. Metabolic pathways were constructed using MetaCyc as a reference data set (14). Genes encoding candidate glycoside hydrolases were detected with routines used for updates of the Carbohydrate Active Enzymes (CAZY) database (18) at <http://www.cazy.org/CAZY/>. Detailed information about the genome properties and genome annotation can be obtained from the JGI Integrated Microbial Genomes website (40) at <http://img.jgi.doe.gov/pub/main.cgi>.

Comparative analysis. To obtain a list of orthologs from bacteroidete genomes, a perl script that determines bidirectional best hits was written. Genes *g* and *h* are considered orthologs if *h* is the best BLASTP hit for *g* and vice versa, with E values of $\leq 10^{-15}$. A gene is considered strain specific if it has no hits with an E value of 10^{-5} or less.

Phylogenetic trees. Preadigned 16S rRNA sequences were derived from the Ribosomal Database Project site (17; <http://rdp.cme.msu.edu/index.jsp>). Manual alignment adjustments were made as needed, with the assistance of the BioEdit multiple alignment tool of Hall (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>). The refined multiple alignment was used as input for the generation of a phylogenetic tree, using the program package PHYLP (20). The fastDNAMl (52) and nucleic acid sequence maximum likelihood (51) methods with default settings were employed to obtain character-based trees. Both the fastDNAMl and maximum likelihood trees yielded similar clusters and arrangements of taxa within them.

Electron microscopy. Cells for scanning electron microscopy were grown on Whatman filter paper on DMA agar as previously described (44). Cells were fixed with 2.5% glutaraldehyde, 0.1% OsO₄ in 10 mM HEPES buffer, pH 7.2, for 30 min at 25°C. Fixed cells were washed twice in 10 mM HEPES buffer, postfixed in 1% OsO₄, and washed twice with distilled water. Cells were dehydrated with ethanol and critical point dried with CO₂. Samples were mounted, sputter coated with 60% gold–40% palladium, and viewed with a Hitachi S-570 scanning electron microscope.

Nucleotide sequence accession number. The sequences of *C. hutchinsonii* can be accessed using GenBank accession number CP000383.

RESULTS AND DISCUSSION

General genome features. The genome of *C. hutchinsonii* consists of a single, circular, 4,433,218-bp chromosome that has a GC content of 38.85% (see Fig. S1 in the supplemental material). A total of 3,790 protein-encoding genes are predicted, 1,986 of which have been assigned a tentative function. The GC skew allowed prediction of the site of the origin of replication to near nucleotide position 4,038,464. Surprisingly, ribosomal genes, tRNA genes, and typical replication-related genes, such as *dnaA*, *dnaN*, *recF*, and *gyrA*, which are often located near the origin of replication in bacteria, were not located near this site. Three rRNA operons were identified in the genome. Analysis of the 16S rRNA sequences confirmed that *C. hutchinsonii* belongs to the phylum *Bacteroidetes* (Fig. 2). It is a member of the class *Sphingobacteria* and is only distantly related to well-studied bacteroidetes, such as members of the genera *Bacteroides*, *Porphyromonas*, *Prevotella* (all from class *Bacteroidetes*), and *Flavobacterium* (class *Flavobacteriia*).

Cellulose utilization. One of the most distinctive features of *C. hutchinsonii* is its ability to rapidly digest crystalline cellulose. Cellulose utilization by most well-studied microorganisms involves the activities of multiple enzymes, including endo- β -1,4-glucanases, cellobiohydrolases (also called exo- β -1,4-glu-

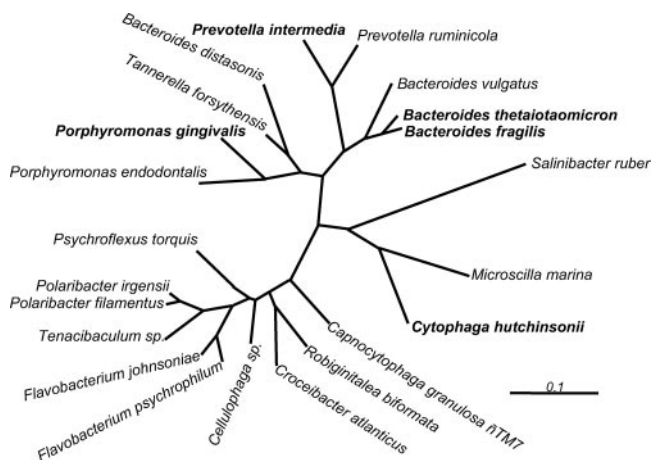


FIG. 2. 16S rRNA-based phylogenetic tree of the phylum *Bacteroidetes* derived from 16S rRNA gene sequence data. The tree was constructed using the fastDNAMl method (20, 51, 52). Species with complete genome sequences that were used in this study are indicated in bold. Bar, 0.1 nucleotide substitution per site.

canases), and β -glucosidases, that act synergistically to convert crystalline cellulose to glucose (39, 70) (Fig. 3). Endo- β -1,4-glucanases attack the cellulose chain at exposed internal glycosidic bonds. They nick the long polymeric substrate and provide sites of attack for cellobiohydrolases, which release cellobiose from the exposed ends. Members of one class of cellobiohydrolases attack the nonreducing end of the polymer, whereas members of the second class attack the reducing end (4). β -Glucosidases hydrolyze cellobiose and short cellodextrin oligomers to glucose. Analysis of the amino acid sequences of cellulolytic enzymes allows them to be grouped into a limited number of glycoside hydrolase families (27–29). Examination of the *C. hutchinsonii* genome identified many genes encoding possible cellulolytic enzymes. These include possible endo- β -1,4-glucanases belonging to glycoside hydrolase families GH5 and GH9 and β -glucosidases belonging to GH3 (Table 1). Genes encoding proteins with similarity to known cellobiohydrolases, which are typically members of family GH6 or GH48 in bacteria (23, 71), were not detected. Since cellobiohydrolases are needed for efficient cellulose utilization by other organisms, *C. hutchinsonii* either has novel cellobiohydrolases or uses an unusual method of cellulose utilization that does not require such enzymes. Some GH9 endo- β -1,4-glucanases are processive (22). The presence of processive endoglucanases could account for the ability of *C. hutchinsonii* to digest crystalline cellulose in the apparent absence of cellobiohydrolases. However, none of the *C. hutchinsonii* enzymes display high levels of similarity to experimentally characterized processive endoglucanases, so further experimental evidence is needed to determine this possibility. The apparent absence of cellobiohydrolases may partially explain why reducing sugars do not accumulate in the medium when cells of *C. hutchinsonii* digest cellulose (64). All four of the candidate β -glucosidases are predicted to reside in the periplasm, and three of them are predicted lipoproteins (Table 1). The periplasmic location of these enzymes may also limit the amount of reducing sugars released beyond the confines of the cell. In addition to enzymes involved in cellulose utilization, *C. hutchinsonii* has

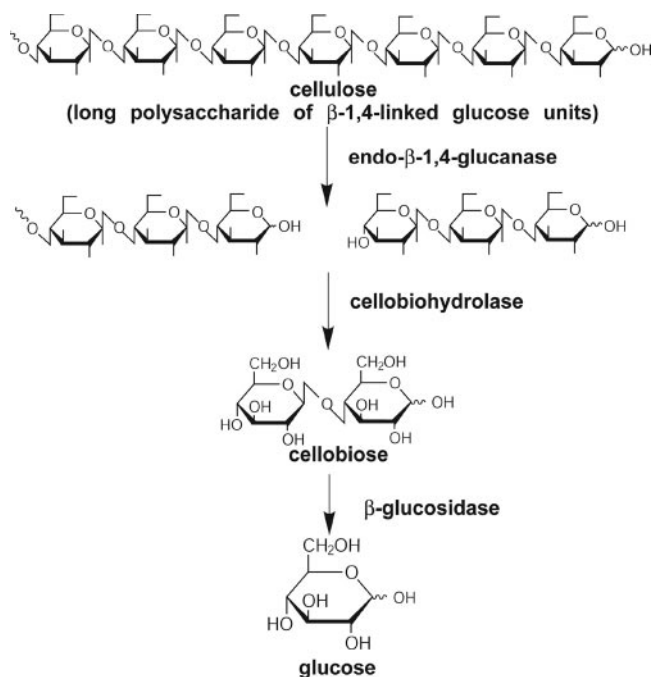


FIG. 3. Enzymes involved in cellulose breakdown in well-studied cellulolytic bacteria. *C. hutchinsonii* lacks obvious homologs of known cellobiohydrolases.

genes predicted to encode glycohydrolases involved in the digestion of hemicelluloses, such as xylan (Table 2), and genes encoding β -glycosidases of uncertain specificity (Table 3). The remaining predicted glycohydrolases (Table 4) are likely to be involved in peptidoglycan turnover and utilization of intracellular storage compounds. Genes encoding predicted pectate lyases (Table 5) and genes encoding polysaccharide esterases, some of which are predicted to be involved in xylan utilization (Table 6), were also identified. *C. hutchinsonii* does not grow on xylan or pectin as a sole carbon and energy source but may use xylanases, pectate lyases, and polysaccharide esterases to gain access to cellulose in plant material. A similar situation has been observed for the cellulolytic anaerobes *Clostridium thermocellum* and *Clostridium cellulovorans* (6).

Recognizable CBMs are found on many of the glycohydrolases (Tables 2 and 3), but surprisingly, none of the proteins related to endo- β -1,4-glucanases contain such a module (Table 1). The predicted endo- β -1,4-glucanases may have novel CBMs or may interact with other CBM-containing proteins to form multiprotein complexes that function in cellulose digestion. Equally surprising is that no genes encoding proteins with type A CBMs, the only ones known to bind to crystalline cellulose (11), were detected in the entire genome. Cells of *C. hutchinsonii* do bind to crystalline cellulose and actively glide along the fibers (Fig. 1). Presumably, components of the motility machinery are involved in these interactions. The attachment of cells and movement along the cellulose fibers may make dedicated CBM domains on the cellulolytic enzymes unnecessary. If type A CBMs were present, they might even prevent gliding and thus decrease the ability of the cells to access and digest cellulose.

Many of the *C. hutchinsonii* proteins with possible roles in

TABLE 1. Predicted *C. hutchinsonii* glycohydrolases that may be involved in cellulose digestion

Gene	Predicted function of gene product ^a	Predicted localization ^b	Mol mass ^c (kDa)	Modular structure ^d
CHU_0778	Related to endo- β -1,4-glucanase	Unknown	81.7	CeID_N-GH9
CHU_1107	Candidate endo- β -1,4-glucanase	Extracellular	135.2	GH5-X1-PKD-PKD-FN3
CHU_1280	Candidate endo- β -1,4-glucanase	Unknown	66.2	CeID_N-GH9
CHU_1335	Related to endo- β -1,4-glucanases	Extracellular/outer membrane	207.3	GH9-PKD-PKD-PKD-BIG-BIG-BIG-BIG-BIG-BIG-BIG-BIG-BIG-BIG-D5
CHU_1336	Related to endo- β -1,4-glucanases	Extracellular	105.3	GH9-PKD-PKD-PKD
CHU_1655	Candidate endo- β -1,4-glucanase	Extracellular/outer membrane	92.6	CeID_N-GH9
CHU_1727	Related to endo- β -1,4-glucanases	Unknown	67.9	GH5
CHU_2103	Candidate endo- β -1,4-glucanase	Extracellular	38.8	GH5
CHU_2235	Distantly related to endo- β -1,4-glucanases	Unknown	64.2	GH9
CHU_2268	Candidate β -glucosidase	Periplasmic lipoprotein	83.7	GH3
CHU_2273	Candidate β -glucosidase	Periplasmic lipoprotein	89.7	GH3
CHU_3577	Candidate β -glucosidase	Periplasmic lipoprotein	83.2	GH3
CHU_3784	Candidate β -glucosidase	Periplasmic	81.8	GH3

^a Assigned by routines used for updating the CAZY database (<http://www.cazy.org/CAZY/>), using the following criteria. Typically, $\geq 70\%$ amino acid identity to a protein domain with a biochemically determined function at the time of analysis resulted in "candidate" status. A 30% to 70% amino acid identity to a protein domain with known function resulted in "related to" status. Less than 30% amino acid identity to a protein domain with known function resulted in "distantly related to" status. Because the threshold of similarity that correlates with a change of substrate specificity is variable from one glycoside hydrolase family to another, the criteria were tightened or loosened appropriately for several families. All analyses were conducted domain by domain because of the modular structure of many of the proteins.

^b Predicted using the default settings of PSORTb (21). Predicted lipoproteins were identified using LipoP (33).

^c Molecular mass of primary product of translation, including any predicted signal peptide.

^d Modular structure is indicated by the following abbreviations: BIG, bacterial immunoglobulin-like domain group 2 (Pfam number PF02368); CeID_N, N-terminal immunoglobulin-like domain of cellulase (Pfam number PF02927); D5, carboxy-terminal domain of *Rhodothermus marinus* xylanases predicted to be involved in attachment to the cell surface (34); FN3, fibronectin type 3 domain (Pfam number PF00041); GH, glycoside hydrolase (number indicates family), as assigned by CAZY; PKD, polycystic kidney disease protein PKD1 (Pfam number PF00801); and X1, conserved domain of unknown function.

polysaccharide utilization are large and have complex multidomain structures (Tables 1, 2, 3, 5, and 7). In addition to the glycohydrolase and CBM domains discussed above, other domains present on some of the predicted proteins include domains with immunoglobulin-like folds, including group 2 bacterial immunoglobulin-like domains (BIG_2; Pfam number

PF02368), N-terminal immunoglobulin-like domains of cellulase (CeID_N; Pfam number PF02927), immunoglobulin I-set domains (IG; Pfam number PF00047), fibronectin type 3-like domains (FN3; Pfam number PF00041), and polycystic kidney disease protein-like domains (PKD; Pfam number PF00801). Parallel β helix repeat domains (PbH1; SMART database

TABLE 2. Predicted *C. hutchinsonii* glycohydrolases that may be involved in hemicellulose digestion

Gene	Predicted function of gene product ^a	Predicted localization ^b	Mol mass ^c (kDa)	Modular structure ^d
CHU_1155	Candidate xyloglucanase	Extracellular	132.2	GH74-PKD-PKD-PKD-PKD-PKD-D5
CHU_1239	Candidate acetylxyylan esterase/xylanase	Extracellular/outer membrane	148.1	CE6-X1-GH10-X1-CBM4-PKD-PKD-PKD-PKD-D5
CHU_1240	Candidate xylanase/esterase	Extracellular/outer membrane	183.0	CE4-GH8-CBM9-PKD-PKD-PKD-PKD-PKD-X1-D5
CHU_2041	Candidate acetylxyylan esterase/ β -xylosidase/ α -L-arabinofuranosidase	Extracellular/outer membrane	166.6	CE6-GH43-X1-X1-CBM9-PKD-PKD-PKD-PKD-X1-D5
CHU_2042	Candidate xylanase	Extracellular/outer membrane	127.4	GH5-X1-CBM9-PKD-PKD-PKD-PKD-X1-D5
CHU_2043	Candidate xylanase	Extracellular/outer membrane	128.5	GH10-X1-CBM9-PKD-PKD-PKD-PKD-PKD-X1-D5
CHU_2044	Related to β -xylosidases/ α -L-arabinofuranosidases	Extracellular/outer membrane	154.4	GH43-CBM6-X1-CBM9-PKD-PKD-PKD-PKD-X1-D5
CHU_2105	Candidate xylanase	Unknown	76.3	CBM4-GH10
CHU_2379	Candidate xylanase	Extracellular/outer membrane	119.5	GH11-CBM9-PKD-PKD-PKD-PKD-PKD-X1-D5
CHU_0353	Candidate β -mannanase	Extracellular	86.6	CBM6-GH26-PKD-PKD-D5

^a Assigned by routines used for updating the CAZY database (<http://www.cazy.org/CAZY/>), using the criteria indicated in Table 1.

^b Predicted using the default settings of PSORTb (21). Predicted lipoproteins were identified using LipoP (33).

^c Molecular mass of primary product of translation, including any predicted signal peptide.

^d Modular structure is indicated by the following abbreviations: CE, carbohydrate esterase (number indicates family), as assigned by CAZY; CBM, carbohydrate binding module (number indicates family), as assigned by CAZY; D5, carboxy-terminal domain of *Rhodothermus marinus* xylanases predicted to be involved in attachment to the cell surface (34); GH, glycoside hydrolase (number indicates family), as assigned by CAZY; PKD, polycystic kidney disease protein PKD1 (Pfam number PF00801); and X1, conserved domain of unknown function.

TABLE 3. Predicted β -glycosidases of uncertain specificity

Gene	Predicted function of gene product ^a	Predicted localization ^b	Mol mass ^c (kDa)	Modular structure ^d
CHU_0013	Candidate β -glycosidase	Periplasmic	110.0	GH3-BL
CHU_0961	Candidate β -glycosidase	Extracellular	151.0	CBM4-CeID_N-GH9-X3-Y98-Y98-Y98-Y98-Y98
CHU_0981	Candidate β -glycosidase	Unknown	37.5	GH16
CHU_1075	Candidate β -glycosidase	Extracellular/outer membrane	274.2	GH8-repeat-repeat
CHU_1842	Candidate β -glycosidase	Lipoprotein	62.3	GH5
CHU_2149	Candidate β -glycosidase	Extracellular/outer membrane	249.7	GH5-PKD-PKD-PKD-PKD-PKD-PKD-PKD-D5
CHU_2802	Candidate β -glycosidase	Extracellular	40.1	GH16
CHU_2852	Candidate β -glycosidase	Extracellular/outer membrane	119.6	GH8-X1-X1-X1-X1-X1
CHU_3441	Related to β -glycosidases	Extracellular/outer membrane	294.1	GH8-repeat
CHU_3727	Candidate β -glycosidase	Extracellular	126.8	GH8-X1-CBM9-Y98-Y98-Y98-D5
CHU_3811	Candidate β -glycosidase	Cytoplasmic	53.5	GH1

^a Assigned by routines used for updating the CAZY database (<http://www.cazy.org/CAZY/>), using the criteria indicated in Table 1.

^b Predicted using the default settings of PSORTb (21). Predicted lipoproteins were identified using LipoP (33).

^c Molecular mass of primary product of translation, including any predicted signal peptide.

^d Modular structure is indicated by the following abbreviations: BL, β -lactamase (Pfam number PF00144); CeID_N, N-terminal immunoglobulin-like domain of cellulase; CBM, carbohydrate binding module (number indicates family), as assigned by CAZY; D5, carboxy-terminal domain of *Rhodothermus marinus* xylanases predicted to be involved in attachment to the cell surface (34); GH, glycoside hydrolase (number indicates family), as assigned by CAZY; PKD, polycystic kidney disease protein PKD1 (Pfam number PF00801); and X1, X3, and Y98, conserved domains of unknown function.

number SM00710) and novel conserved domains (X1, X3, X100, and Y98) were also identified. The functions of these domains are not known, but possibilities include binding to cellulose, binding to other components of the cellulolytic machinery, or interaction with the cell surface. Eighteen of the proteins listed in Tables 1, 2, 3, 5, 6, and 7 have carboxy-terminal domains that are similar to the D5 domains of *Rhodothermus marinus* cell-associated xylanolytic enzymes (34). The *R. marinus* D5 domains have been postulated to be involved in attachment to the cell surface. Carboxy-terminal D5 domains are also found on 39 additional *C. hutchinsonii* proteins (Fig. 4). These may also be part of the cell-bound polysaccharide utilization machinery of *C. hutchinsonii* or may have other functions. Experimental evidence indicates that *C. hutchinsonii* requires direct contact with cellulose for efficient digestion and that its cellulolytic enzymes are primarily cell associated (47, 64), and

some of the domains described above may be involved in this localization. Cellulolytic members of the gram-positive genera *Clostridium* and *Ruminococcus* also produce cell-associated cellulases. These bacteria have multiprotein cellulolytic complexes, called cellulosomes, composed of scaffoldin proteins with multiple cohesin domains that bind to complementary dockerin domains on the cellulolytic enzymes (6). The high-affinity cohesin-dockerin interactions result in the formation of the large cell surface cellulosomal structures. *C. hutchinsonii* does not encode obvious scaffoldin proteins or proteins with recognizable cohesin or dockerin domains but may have cellulosome-like structures composed of different cellulolytic multienzyme complexes.

Other predicted cell-associated proteins that may be involved in cellulose utilization include homologs of the *Bacteroides thetaiotaomicron* outer membrane starch utilization pro-

TABLE 4. *C. hutchinsonii* glycohydrolases with predicted functions unrelated to cellulose or hemicellulose digestion

Gene	Predicted function of gene product ^a	Predicted localization ^b	Mol mass ^c (kDa)	Modular structure ^d
CHU_0026	Related to peptidoglycan lytic transglycosylases (10)	Unknown	55.1	GH23-LysM-LysM
CHU_0157	Related to peptidoglycan lytic transglycosylases (10)	Unknown	36.5	GH23
CHU_2792	Related to peptidoglycan lytic transglycosylases (10)	Periplasmic	56.3	SBP-GH23
CHU_3030	Related to peptidoglycan lytic transglycosylases (10)	Lipoprotein	65.9	GH23
CHU_0470	Candidate β -N-acetylglucosaminidase	Lipoprotein	42.9	GH3
CHU_0399	Distantly related to glucoamylases	Unknown	68.5	GH15
CHU_0801	Candidate α -amylase	Cytoplasmic	46.4	GH57
CHU_0803	Candidate α -glucosidase	Cytoplasmic	92.8	GH31
CHU_0959	Candidate α -glycosidase	Cytoplasmic	60.3	GH13
CHU_1345	Candidate glycogen branching enzyme	Unknown	79.1	GH13
CHU_2409	Candidate α -amylase	Cytoplasmic	56.7	GH13
CHU_2602	Candidate α -glycosidase; possible maltooligosyl trehalose synthase	Unknown	105.6	GH13
CHU_2290	Candidate 4- α -glucanotransferase	Cytoplasmic	58.0	GH77

^a Assigned by routines used for updating the CAZY database (<http://www.cazy.org/CAZY/>), using the criteria indicated in Table 1.

^b Predicted using the default settings of PSORTb (21). Predicted lipoproteins were identified using LipoP (33).

^c Molecular mass of primary product of translation, including any predicted signal peptide.

^d Modular structure is indicated by the following abbreviations: GH, glycoside hydrolase (number indicates family), as assigned by CAZY; LysM, lysin motif (SMART database number SM00257); and SBP, solute-binding protein Bac 3 (Pfam number PF00497).

TABLE 5. Predicted *C. hutchinsonii* polysaccharide lyases

Gene	Predicted function of gene product ^a	Predicted localization ^b	Mol mass ^c (kDa)	Modular structure ^d
CHU_0455	Distantly related to pectate lyases	Unknown	62.0	PL-PbH1-PbH1-PbH1-PbH1-PbH1-PbH1-PbH1-PbH1
CHU_1157	Candidate rhamnogalacturonan lyase	Extracellular/outer membrane	132.1	PL11-PKD-PKD-PKD-PKD-PKD-PKD-D5
CHU_1162	Candidate pectate lyase	Extracellular	100.6	PL1-X1-PKD-PKD-PKD-PKD-PKD-D5
CHU_2148	Candidate polysaccharide lyase	Unknown	40.0	PL14

^a Assigned by routines used for updating the CAZY database (<http://www.cazy.org/CAZY/>), using the criteria indicated in Table 1.

^b Predicted using the default settings of PSORTb (21). Predicted lipoproteins were identified using LipoP (33).

^c Molecular mass of primary product of translation, including any predicted signal peptide.

^d Modular structure is indicated by the following abbreviations: D5, carboxy-terminal domain of *Rhodothermus marinus* xylanases predicted to be involved in attachment to the cell surface (34); PbH1, parallel beta-helix repeats (SMART database number SM00710); PKD, polycystic kidney disease protein PKD1 (Pfam number PF00801); and PL, polysaccharide lyase (number indicates family), as assigned by CAZY.

teins SusC and SusD. *B. thetaiotaomicron* is an anaerobic inhabitant of the human large intestine and is a distant relative of *C. hutchinsonii*. It does not digest cellulose but does utilize starch. Like those of *C. hutchinsonii*, the *B. thetaiotaomicron* polysaccharide-degrading enzymes are primarily cell associated (2, 57). Several genes that are required for starch utilization have been identified, and the encoded proteins have been characterized (54, 57, 60, 61). SusC and SusD are involved in binding starch on the cell surface and in transport of oligomers across the outer membrane. Homologs of SusC and SusD are common in the phylum *Bacteroidetes* and include the cell surface proteins RagA and RagB, respectively, of *Porphyromonas gingivalis* (24). Analysis of the *C. hutchinsonii* genome revealed the presence of two genes related to *susC* (CHU_0546 and CHU_0553) and two genes related to *susD* (CHU_0547 and CHU_0554). Since *C. hutchinsonii* does not utilize extracellular starch and the only polysaccharide that it is known to use is cellulose, we speculate that the proteins encoded by these genes may be involved in binding and utilization of cellulose or products of cellulose digestion.

Transport. *C. hutchinsonii* has 184 genes predicted to be involved in transport. Of particular interest are genes potentially involved in transport of the products of cellulose hydrolysis. Transporters of the phosphoenol-pyruvate:sugar phosphotransferase system are absent from *C. hutchinsonii*,

but six genes encoding putative sugar transport permeases of the major facilitator superfamily are present (CHU_0773, CHU_0325, CHU_0960, CHU_1068, CHU_1656, and CHU_3606). The predicted transport proteins are members of the fucose:H⁺ symporter family, which has well-characterized members that transport fucose, glucose, and galactose, among other sugars. These proteins may be involved in transport of the products released during cellulose digestion. Another transporter that may play a role in cellulose utilization is the putative gluconate transporter encoded by CHU_1403. The remaining small-molecule transporters are predicted to function in transport of amino acids and inorganic ions across the cytoplasmic membrane and in cell envelope biogenesis. Genes encoding ExbB, ExbD, 4 TonB-like proteins, 15 TonB-dependent outer membrane receptors, and a TolC-like protein were also identified and may play roles in transport across the outer membrane.

Translocation of proteins across the cytoplasmic membrane is apparently mediated by SecA, SecE, SecY, SecDF, SecG, YidC, and YajC of the Sec system and by TatA and TatC of the twin-arginine transport system. The major components involved in type II secretion (GspD, GspE, GspF, and GspG) are also present and, presumably, mediate protein translocation across the outer membrane. Possible ABC transporters that may transport specific proteins (type I transport) and auto-

TABLE 6. Predicted *C. hutchinsonii* carbohydrate esterases

Gene	Predicted function of gene product ^a	Predicted localization ^b	Mol mass ^c (kDa)	Modular structure ^d
CHU_0603	Related to polysaccharide deacetylases	Cytoplasmic membrane	43.1	CE4
CHU_0697	Related to polysaccharide deacetylases	Unknown	36.4	CE4
CHU_0721	Related to acetylxyylan esterases	Lipoprotein	44.2	CE2
CHU_1037	Candidate UDP-3-0-acyl N-acetylglucosamine deacetylase	Unknown	51.3	CE11
CHU_2040	Related to esterases	Extracellular	107.2	CE1-X1-CBM9-PKD-PKD-PKD-X1-D5
CHU_2113	Related to GlcNAc deacetylases	Cytoplasmic membrane	29.3	CE4
CHU_2175	Related to GlcNAc deacetylases	Cytoplasmic	26.0	CE4
CHU_2408	Related to esterases	Unknown	35.6	CE1
CHU_3337	Related to acetylxyylan esterases	Lipoprotein	44.4	CE2

^a Assigned by routines used for updating the CAZY database (<http://www.cazy.org/CAZY/>), using the criteria indicated in Table 1.

^b Predicted using the default settings of PSORTb (21). Predicted lipoproteins were identified using LipoP (33).

^c Molecular mass of primary product of translation, including any predicted signal peptide.

^d Modular structure is indicated by the following abbreviations: CE, carbohydrate esterase (number indicates family), as assigned by CAZY; CBM, carbohydrate binding module (number indicates family), as assigned by CAZY; D5, carboxy-terminal domain of *Rhodothermus marinus* xylanases predicted to be involved in attachment to the cell surface (34); PKD, polycystic kidney disease protein PKD1 (Pfam number PF00801); and X1, conserved domain of unknown function.

TABLE 7. Predicted *C. hutchinsonii* polysaccharide binding proteins lacking predicted catalytic domains

Gene	Predicted function of gene product ^a	Predicted localization ^b	Mol mass ^c (kDa)	Modular structure ^d
CHU_1051	Candidate polysaccharide binding protein	Extracellular/outer membrane	138.2	FN3-X1-CBM6-X1-PKD-X1-D5
CHU_1221	Candidate polysaccharide binding protein	Extracellular/outer membrane	175.4	CBM9-IG-PKD-PKD-PKD-PKD-PKD-D5
CHU_1238	Candidate polysaccharide binding protein	Extracellular/outer membrane	132.7	X100-X1-CBM9-PKD-PKD-PKD-PKD-X1-D5

^a Assigned by routines used for updating the CAZY database (<http://www.cazy.org/CAZY/>), using the criteria indicated in Table 1.

^b Predicted using the default settings of PSORTb (21). Predicted lipoproteins were identified using LipoP (33).

^c Molecular mass of primary product of translation, including any predicted signal peptide.

^d Modular structure is indicated by the following abbreviations: CBM, carbohydrate binding module (number indicates family), as assigned by CAZY; D5, carboxy-terminal domain of *Rhodothermus marinus* xylanases predicted to be involved in attachment to the cell surface (34); FN3, fibronectin type 3 domain (Pfam number PF00041); IG, immunoglobulin I-set domain (Pfam number PF00047); PKD, polycystic kidney disease protein PKD1 (Pfam number PF00801); and X1 and X100, conserved domains of unknown function.

transporters that may facilitate transport of themselves or other proteins across the outer membrane (type V transport) were also identified. The *C. hutchinsonii* genome encodes 194 predicted lipoproteins, and genes encoding machinery for their processing (*lgt*, *lspA*, and *lnt*) and translocation (*lolA*, *lolD*, and *lolE*) are also present.

Metabolism of glucose. *C. hutchinsonii* carries out aerobic respiration of glucose to CO₂ (47, 64), and genome analysis indicated the presence of a complete Embden-Meyerhof-Parnas pathway and tricarboxylic acid cycle. Genes encoding each of the NADH dehydrogenase subunits, cytochrome *c*, cytochrome *c* oxidase, and components of ATP synthase were also present, accounting for the ability of *C. hutchinsonii* to carry out aerobic respiration of glucose. *C. hutchinsonii* also may have the ability to import and metabolize gluconate. In addition to the apparent gluconate transporter mentioned above, the protein encoded by CHU_1404 is predicted to exhibit gluconokinase and 6-phosphogluconate dehydrogenase activities and is likely to be involved in gluconate metabolism (see Fig. S2 in the supplemental material). As indicated above, *C. hutchinsonii* appears to carry enzymes involved in digestion of xylan but does not use xylan as a sole source of carbon and energy. *C. hutchinsonii* also fails to grow on common components of xylan, such as xylose and arabinose (37, 64), and analysis of the genome revealed the apparent absence of transporters for these sugars and of key enzymes needed for their metabolism, including arabinose isomerase, xylose isomerase, ribulokinase, and xylulokinase. Transporters and enzymes involved in the utilization of galacturonic acid were also apparently lacking, explaining the inability to grow on pectin despite the presence of genes encoding candidate pectate lyases.

Biosynthetic capabilities. *C. hutchinsonii* grows on minimal media with filter paper cellulose or glucose as the sole carbon and energy source and thus has the ability to synthesize all of its organic components from these substrates. As expected, analysis of the genome sequence revealed genes encoding biosynthetic enzymes needed to synthesize the 20 common amino acids and to synthesize nucleotides, fatty acids, and vitamins and coenzymes, such as biotin, folic acid, lipoic acid, nicotinic acid, pantothenic acid, pyridoxine, riboflavin, and thiamine. Genes encoding enzymes for the synthesis of heme and menaquinone were also identified.

Signal transduction and regulation of gene expression. *C. hutchinsonii* carries a variety of proteins that are predicted to regulate gene expression in response to external or internal stimuli. Sigma factors include an RpoD (σ^{70}) homolog, a σ^{54} homolog, and 19 sigma factors belonging to the extracytoplasmic function (ECF) subfamily. *C. hutchinsonii* RpoD is similar in size (32.7 kDa) and sequence to other bacteroidete RpoD proteins and is much smaller than *Escherichia coli* σ^{70} . Like the other bacteroidete RpoD proteins, *C. hutchinsonii* RpoD lacks regions found in most nonbacteroidete RpoD proteins, such as the N-terminal region 1.1 and the segment between regions 1.2 and 2.1 (68). The novel structure of bacteroidete RpoD sigma factors may account for some of the unusual features of bacteroidete promoters that have been studied (8). The presence of large numbers of ECF sigma factors is a common property of members of the phylum *Bacteroidetes* (35). ECF sigma factors in other bacteria regulate the expression of genes with extracytoplasmic function. *C. hutchinsonii* has 53 predicted histidine kinase two-component signal transduction proteins, 34 predicted response regulatory proteins, and 7 proteins that contain both histidine kinase and response regulatory domains. *C. hutchinsonii* is predicted to carry 17 cyclic nucleotide-binding proteins and probably uses cyclic AMP to regulate gene expression and enzymatic activities. In contrast, it does not appear to use the common bacterial signaling molecule cyclic di-GMP. Most bacteria carry proteins with GGDEF domains, involved in the synthesis of cyclic di-GMP, and proteins with EAL domains, thought to function as phosphodiesterases for cyclic di-GMP turnover (56). Analysis of the *C. hutchinsonii* genome indicated the complete absence of genes encoding proteins with these domains. GGDEF and EAL domain proteins were also absent from the other members of the phylum *Bacteroidetes* with complete genome sequence data that we analyzed, including *B. thetaiotaomicron*, *Bacteroides fragilis* YCH46, *Porphyromonas gingivalis*, *Prevotella intermedia*, and *Tannerella forsythensis*.

Gliding motility. *C. hutchinsonii* cells cannot swim in liquid, but they attach to and move along cellulose fibers as they digest them (64) (Fig. 1). The mechanism of this gliding motility is not known. Gliding cells of *C. hutchinsonii* move at speeds of up to 5 $\mu\text{m}/\text{second}$ over glass surfaces and occasionally reverse their direction of movement. Cells also attach to the glass by



FIG. 4. Multiple sequence alignment of C-terminal domains in *C. hutchinsonii* proteins, found by BLASTP similarity searches using D5 of *Rhodothermus marinus* xylanase Xyn10A and *R. marinus* mannanase Man26A as queries. Black shading indicates amino acid identity, and gray shading indicates amino acid similarity. A threshold for shading was set at 60% of the consensus.

one pole and rotate in place at frequencies of about 2 revolutions per second. The ability to glide over surfaces may provide *C. hutchinsonii* cells with a selective advantage for utilization of insoluble cellulose. Gliding motility may allow cells to migrate along cellulose fibers in search of regions that are most amenable to enzymatic attack and to penetrate deep into the cellulose matrix as the substrate is digested. In some other bacteria, flagella and type IV pili allow cells to move over surfaces (25, 41). Electron microscopic analyses have failed to identify these organelles on cells of *C. hutchinsonii*, and analysis of the genome also failed to identify genes encoding critical compo-

nents of flagella and type IV pili, suggesting that *C. hutchinsonii* gliding motility relies on other machinery.

Gliding motility is characteristic of many members of the phylum *Bacteroidetes* (42). The mechanism of gliding motility has been studied extensively for one bacteroidete, *Flavobacterium johnsoniae*, which is distantly related to *C. hutchinsonii* (Fig. 2). Sixteen genes (*gldA*, *gldB*, *gldD*, *gldF*, *gldG*, *gldH*, *gldI*, *gldJ*, *gldK*, *gldL*, *gldM*, *gldN*, *sprA*, *sprB*, *sprC*, and *sprD*) involved in *F. johnsoniae* gliding motility have been identified (1, 12, 13, 30–32, 45, 46). *gldA*, *gldF*, and *gldG* are thought to encode components of an ATP-binding cassette transporter

that is required for gliding (1, 30). The functions of the other motility proteins are less certain. In a current model of *F. johnsoniae* gliding, the Gld proteins comprise the gliding motor, which propels cell surface adhesins comprised of Spr proteins. *C. hutchinsonii* has homologues of each of the *gld* and *spr* genes, and these are likely to be involved in cell movement. Disruption of *C. hutchinsonii gldG* by homologous recombination with a plasmid containing an internal fragment of the gene resulted in a loss of motility, implicating *C. hutchinsonii gldG* in gliding (C. Guntur and M. J. McBride, unpublished data).

Gliding motility is not confined to the bacteroidetes, and recent evidence indicates that there are multiple, genetically unrelated mechanisms for bacterial gliding (42, 43). *Myxococcus xanthus* has two gliding motility systems. The S motility system relies on extension and retraction of type IV pili to pull cells, whereas the mechanism of the A motility system is not yet certain (48, 65, 69, 73). *Mycoplasma mobile* appears to use a third gliding motility system composed of cytoskeletal proteins and large cell surface proteins (49). Most *F. johnsoniae* and *C. hutchinsonii gld* genes have no obvious homologs in the genomes of *M. xanthus* and *M. mobile* (12), suggesting that *Cytophaga-Flavobacterium* gliding motility is distinct from myxobacterial or mycoplasma gliding.

F. johnsoniae does not digest cellulose, but it does digest another insoluble polysaccharide, chitin. Surprisingly, all mutations that eliminate gliding also eliminate the ability of *F. johnsoniae* to digest chitin (12, 13, 15, 45, 46). The connection between motility and chitin digestion is not understood. Perhaps cells glide along the insoluble polymer until they reach regions amenable to digestion. Alternatively, the motility machinery, which can move large particles over the cell surface (53), may be needed to move oligomers to sites where they can be digested further or taken into the cell. The role, if any, of motility in *C. hutchinsonii* cellulose digestion remains to be explored. The regular arrangement of gliding cells on cellulose fibers (Fig. 1) suggests that motility may position the cell-associated cellulolytic machinery so that digestion of the insoluble polymer is optimized.

All motile bacteria that have been studied control their motility apparatus to move in a favorable direction. This typically involves a complex signal transduction system, such as the *E. coli* chemotaxis system, that senses the environment and affects the functioning of the motility apparatus. *C. hutchinsonii* presumably controls its motility machinery in response to external stimuli, and it has genes encoding proteins that are similar to *E. coli* CheY (CHU_2082), CheB (CHU_2078), and CheR (CHU_2079). These genes are clustered together on the genome, supporting the idea that their products may function together. *C. hutchinsonii* CheB is unusual in that it lacks the N-terminal response regulatory domain that is found in most CheB proteins. Homologs of *cheB* and *cheR* are also found in other bacteroidetes that exhibit gliding motility, such as *F. johnsoniae* and *Tenacibaculum* sp. strain MED152, but are not found in the nonmotile bacteroidetes *B. thetaiotaomicron* BPI5482, *B. fragilis* YCH46 or NCTC9343, *P. gingivalis* W83, *P. intermedia* 17, and *T. forsythensis*. *C. hutchinsonii* has another gene, CHU_1237, that encodes a protein with similarity to both CheB and CheR. This protein has a CheB-like domain at the amino terminus followed by a CheR-like domain, a PAS domain, and a histidine kinase domain at the carboxy terminus.

The CheB-like domain lacks the response regulator domain that is present in most CheB proteins. Genes encoding other expected components of a bacterial chemotaxis system, such as homologs of methyl-accepting chemotaxis proteins (MCPs) and of CheA, CheW, and CheZ, are apparently lacking. The absence of obvious MCPs is surprising since CheB (methyltransferase) and CheR (methyl-esterase) are expected to modify MCPs. The *C. hutchinsonii* CheB and CheR homologs may modify novel chemotaxis proteins or may have roles unrelated to motility and chemotaxis. The region surrounding *C. hutchinsonii cheB*, *cheR*, and *cheY* contains several genes whose products are likely to be involved in signal transduction. These include an apparent histidine kinase with a PAS domain (CHU_2084), two proteins that each contain histidine kinase and response regulatory domains (CHU_2077 and CHU_2081), and a large protein (CHU_2080) that contains a CHASE domain, a GAF domain, a histidine kinase domain, and a response regulatory domain. PAS, CHASE, and GAF domains are commonly found in sensory or signal transduction proteins (3, 66, 74) and could have roles in *C. hutchinsonii* tactic responses.

Comparative analysis of bacteroidete genomes. The complete genome sequences of five members of the phylum *Bacteroidetes*, namely, *C. hutchinsonii*, *B. thetaiotaomicron* BPI5482, *B. fragilis*, *P. gingivalis* W83, and *P. intermedia* 17, were analyzed. *B. thetaiotaomicron*, *B. fragilis*, *P. gingivalis*, and *P. intermedia* are nonmotile anaerobic members of the class *Bacteroidetes*, and each typically inhabits animal digestive tracts. *C. hutchinsonii* is an aerobic bacterium of the class *Sphingobacteria* that is commonly found in soil. Given these differences in phylogeny and lifestyle, it is not surprising that there are many features of the *C. hutchinsonii* genome that set it apart from the others. Direct comparison of the complete nucleotide sequences, using NUCmer and dotplot (36) analyses, revealed no recognizable synteny between *C. hutchinsonii* and the other bacteroidetes. Five hundred fifty-five core bacteroidete genes (requiring bidirectional best hits as a minimum criterion) were shared between the five organisms tested, i.e., *C. hutchinsonii*, *B. thetaiotaomicron*, *P. gingivalis*, *P. intermedia*, and *T. forsythensis*. *C. hutchinsonii*, unlike the other bacteria, appears to have a complete tricarboxylic acid cycle and a complete respiratory electron transport chain, as expected from its demonstrated ability to carry out aerobic respiration. Other differences are noted in Table S1 and Fig. S2 in the supplemental material.

The complete genome sequence of *C. hutchinsonii* provides a window on the inner workings of this common cellulolytic soil bacterium. The mechanism of cellulose utilization employed by cells of *C. hutchinsonii* appears to differ from that employed by cellulosome-containing anaerobic clostridia (7, 9) or by aerobic cellulolytic bacteria, such as *Thermobifida fusca* and *Saccharophagus degradans*, which employ soluble cellulases (67, 71). The apparent absence of cellobiohydrolases and type A CBMs also sets *C. hutchinsonii* apart from most well-studied cellulolytic microorganisms. We hypothesize that *C. hutchinsonii* uses cell surface endoglucanases to attack the insoluble cellulose fibers. Soluble oligomers may be transported actively across the outer membrane with the assistance of SusC-like and SusD-like proteins, with further digestion occurring in the periplasm and cytoplasm. An efficient surface motility system allows the bacterium to attach to and migrate along its insoluble substrate as it digests it. The genome sequence data and

the availability of techniques to genetically manipulate *C. hutchinsonii* will allow experiments to probe the cellulolytic and motility systems of this common but poorly understood bacterium.

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy under contract no. W-7405-ENG-36 and by a UWM-RGI award to M.M.

We thank M. Burmeister and H. Owen for assistance with scanning electron microscopy and E. Leadbetter for constructive comments on the manuscript.

REFERENCES

- Agarwal, S., D. W. Hunnicutt, and M. J. McBride. 1997. Cloning and characterization of the *Flavobacterium johnsoniae* (*Cytophaga johnsonae*) gliding motility gene, *gldA*. Proc. Natl. Acad. Sci. USA **94**:12139–12144.
- Anderson, K. L., and A. A. Salyers. 1989. Biochemical evidence that starch breakdown by *Bacteroides thetaiotaomicron* involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes. J. Bacteriol. **171**:3192–3198.
- Aravind, L., and C. P. Ponting. 1997. The GAF domain: an evolutionary link between diverse phototransducing proteins. Trends Biochem. Sci. **22**:458–459.
- Barr, B. K., Y. L. Hsieh, B. Ganem, and D. B. Wilson. 1996. Identification of two functionally different classes of exocellulases. Biochemistry **35**:586–592.
- Bayer, E., and R. Lamed. 1992. The cellulose paradox: pollutant *par excellence* and/or a reclaimable natural resource? Biodegradation **3**:171–188.
- Bayer, E. A., J. P. Belaich, Y. Shoham, and R. Lamed. 2004. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu. Rev. Microbiol. **58**:521–554.
- Bayer, E. A., L. J. W. Shimon, Y. Shoham, and R. Lamed. 1998. Cellulosomes—structure and ultrastructure. J. Struct. Biol. **124**:221–234.
- Bayley, D. P., E. R. Rocha, and C. J. Smith. 2000. Analysis of *cepA* and other *Bacteroides fragilis* genes reveals a unique promoter structure. FEMS Microbiol. Lett. **193**:149–154.
- Beguín, P., and M. Lemaire. 1996. The cellulosome: an extracellular, multi-protein complex specialized in cellulose degradation. Crit. Rev. Biochem. Mol. Biol. **31**:201–236.
- Blackburn, N. T., and A. J. Clarke. 2001. Identification of four families of peptidoglycan lytic transglycosylases. J. Mol. Evol. **52**:78–84.
- Boraston, A. B., D. N. Bolam, H. J. Gilbert, and G. J. Davies. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. **382**:769–781.
- Braun, T. F., M. K. Khubbar, D. A. Saffarini, and M. J. McBride. 2005. *Flavobacterium johnsoniae* gliding motility genes identified by *mariner* mutagenesis. J. Bacteriol. **187**:6943–6952.
- Braun, T. F., and M. J. McBride. 2005. *Flavobacterium johnsoniae* GldJ is a lipoprotein that is required for gliding motility. J. Bacteriol. **187**:2628–2637.
- Caspi, R., H. Foerster, C. A. Fulcher, R. Hopkinson, J. Ingraham, P. Kaipa, M. Krummenacker, S. Paley, J. Pick, S. Y. Rhee, C. Tissier, P. Zhang, and P. D. Karp. 2006. MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res. **34**:D511–D516.
- Chang, L. Y. E., J. L. Pate, and R. J. Betzig. 1984. Isolation and characterization of nonspreading mutants of the gliding bacterium *Cytophaga johnsonae*. J. Bacteriol. **159**:26–35.
- Chang, W. T. H., and D. W. Thayer. 1977. The cellulase system of a *Cytophaga* species. Can. J. Microbiol. **23**:1285–1292.
- Cole, J. R., B. Chai, R. J. Farris, Q. Wang, S. A. Kulam, D. M. McGarrell, G. M. Garrity, and J. M. Tiedje. 2005. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Res. **33**:D294–D296.
- Coutinho, P. M., and B. Henrissat. 1999. Carbohydrate-active enzymes: an integrated database approach, p. 3–12. In H. J. Gilbert, G. Davies, B. Henrissat, and B. Svensson (ed.), Recent advances in carbohydrate bioengineering. The Royal Society of Chemistry, Cambridge, United Kingdom.
- Doi, R. H., M. Goldstein, S. Hashida, J. S. Park, and M. Takagi. 1994. The *Clostridium cellulovorans* cellulosome. Crit. Rev. Microbiol. **20**:87–93.
- Felsenstein, J. 1997. An alternating least squares approach to inferring phylogenies from pairwise distances. Syst. Biol. **46**:101–111.
- Gardy, J. L., M. R. Laird, F. Chen, S. Rey, C. J. Walsh, M. Ester, and F. S. L. Brinkman. 2005. PSORTb v. 2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. Bioinformatics **21**:617–623.
- Gilad, R., L. Rabinovich, S. Yaron, E. A. Bayer, R. Lamed, H. J. Gilbert, and Y. Shoham. 2003. CelI, a noncellulosomal family 9 enzyme from *Clostridium thermocellum*, is a processive endoglucanase that degrades crystalline cellulose. J. Bacteriol. **185**:391–398.
- Gilkes, N. R., D. Kilburn, E. Kwan, S. Mansfield, J. Saddler, H. Shen, H. Stalbrand, and A. Warren. 1999. Recent insights into the activities of cello-hydrolases from *Cellulomonas fimi* and other cellulolytic microorganisms, p. 24–34. In K. Ohmiya, K. Hayashi, K. Sakka, Y. Kobayashi, S. Karita, and T. Kimura (ed.), Genetics, biochemistry and ecology of cellulose degradation. Uni Publishers Co., Tokyo, Japan.
- Hanley, S. A., J. Aduse-Opoku, and M. A. Curtis. 1999. A 55-kilodalton immunodominant antigen of *Porphyromonas gingivalis* W50 has arisen via horizontal gene transfer. Infect. Immun. **67**:1157–1171.
- Harshey, R. M. 1994. Bees aren't the only ones: swarming in gram-negative bacteria. Mol. Microbiol. **13**:389–394.
- Henikoff, S., and J. G. Henikoff. 1994. Protein family classification based on searching a database of blocks. Genomics **19**:97–107.
- Henrissat, B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. **280**:309–316.
- Henrissat, B., and A. Bairoch. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. **293**:781–788.
- Henrissat, B., and A. Bairoch. 1996. Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. **316**:695–696.
- Hunnicutt, D. W., M. J. Kempf, and M. J. McBride. 2002. Mutations in *Flavobacterium johnsoniae* *gldF* and *gldG* disrupt gliding motility and interfere with membrane localization of GldA. J. Bacteriol. **184**:2370–2378.
- Hunnicutt, D. W., and M. J. McBride. 2001. Cloning and characterization of the *Flavobacterium johnsoniae* gliding motility genes *gldD* and *gldE*. J. Bacteriol. **183**:4167–4175.
- Hunnicutt, D. W., and M. J. McBride. 2000. Cloning and characterization of the *Flavobacterium johnsoniae* gliding motility genes *gldB* and *gldC*. J. Bacteriol. **182**:911–918.
- Juncker, A. S., H. Willenbrock, G. von Heijne, H. Nielsen, S. Brunak, and A. Krogh. 2003. Prediction of lipoprotein signal peptides in gram-negative bacteria. Protein Sci. **12**:1652–1662.
- Karlsson, E. N., M. A. Hachem, S. Ramchuran, H. Costa, O. Holst, S. F. Svenningsen, and G. O. Hreggvidsson. 2004. The modular xylanase Xyn10A from *Rhodothermus marinus* is cell-attached, and its C-terminal domain has several putative homologues among cell-attached proteins within the phylum *Bacteroidetes*. FEMS Microbiol. Lett. **241**:233–242.
- Kill, K., T. T. Binnewies, T. Sicheritz-Ponten, H. Willenbrock, P. F. Hallin, T. M. Wassenaar, and D. W. Ussery. 2005. Genome update: sigma factors in 240 bacterial genomes. Microbiology **151**:3147–3150.
- Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, and S. L. Salzberg. 2004. Versatile and open software for comparing large genomes. Genome Biol. **5**:R12.
- Larkin, J. M. 1989. Nonphotosynthetic, nonfruiting gliding bacteria, p. 2010–2138. In J. T. Staley, M. P. Bryant, N. Pfennig, and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 3. Williams and Wilkins, Baltimore, MD.
- Leschine, S. B. 1995. Cellulose degradation in anaerobic environments. Annu. Rev. Microbiol. **49**:399–426.
- 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.
- Markowitz, V. M., F. Korzeniewski, K. Palaniappan, E. Szeto, G. Werner, A. Padki, X. Zhao, I. Dubchak, P. Hugenholz, I. Anderson, A. Lykidis, K. Mavromatis, N. Ivanova, and N. C. Kyrpides. 2006. The integrated microbial genomes (IMG) system. Nucleic Acids Res. **34**:D344–D348.
- Mattick, J. S. 2002. Type IV pili and twitching motility. Annu. Rev. Microbiol. **56**:289–314.
- McBride, M. J. 2001. Bacterial gliding motility: multiple mechanisms for cell movement over surfaces. Annu. Rev. Microbiol. **55**:49–75.
- McBride, M. J. 2004. *Cytophaga-Flavobacterium* gliding motility. J. Mol. Microbiol. Biotechnol. **7**:63–71.
- McBride, M. J., and S. A. Baker. 1996. Development of techniques to genetically manipulate members of the genera *Cytophaga*, *Flavobacterium*, *Flexibacter*, and *Sporocytophaga*. Appl. Environ. Microbiol. **62**:3017–3022.
- McBride, M. J., and T. F. Braun. 2004. GldI is a lipoprotein that is required for *Flavobacterium johnsoniae* gliding motility and chitin utilization. J. Bacteriol. **186**:2295–2302.
- McBride, M. J., T. F. Braun, and J. L. Brust. 2003. *Flavobacterium johnsoniae* GldH is a lipoprotein that is required for gliding motility and chitin utilization. J. Bacteriol. **185**:6648–6657.
- Middleton, K. S. 1963. Studies on the cellulose-decomposing soil cytophagas. M.A. thesis. University of California-Berkeley, Berkeley.
- Mignot, T., J. W. Shaevez, P. L. Hartzell, and D. R. Zusman. 2007. Evidence that focal adhesion complexes power bacterial gliding motility. Science **315**:853–856.
- Miyata, M. 2005. Gliding motility of mycoplasmas: the mechanism cannot be explained by current biology, p. 137–163. In A. Blanchard and G. Browning (ed.), Mycoplasmas: pathogenesis, molecular biology and emerging strategies for control. Horizon Scientific Press, Norwich, United Kingdom.
- Morrison, M., and J. Miron. 2000. Adhesion to cellulose by *Ruminococcus albus*: a combination of cellulosomes and Pil-proteins? FEMS Microbiol. Lett. **185**:109–115.

51. Navidi, W. C., G. A. Churchill, and A. von Haeseler. 1991. Methods for inferring phylogenies from nucleic acid sequence data by using maximum likelihood and linear invariants. *Mol. Biol. Evol.* **8**:128–143.
52. Olsen, G. J., H. Matsuda, R. Hagstrom, and R. Overbeek. 1994. fastDNAmL: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Comput. Appl. Biosci.* **10**:41–48.
53. Pate, J. L., and L.-Y. E. Chang. 1979. Evidence that gliding motility in prokaryotic cells is driven by rotary assemblies in the cell envelopes. *Curr. Microbiol.* **2**:59–64.
54. Reeves, A. R., J. N. D'Elia, J. Frias, and A. A. Salyers. 1996. A *Bacteroides thetaiotaomicron* outer membrane protein that is essential for utilization of maltooligosaccharides and starch. *J. Bacteriol.* **178**:823–830.
55. Reichenbach, H. 1992. The order Cytophagales, p. 3631–3675. In A. Balows, H. G. Truper, M. Dworkin, W. Harder, and K. M. Schleifer (ed.), *The prokaryotes*. Springer-Verlag, New York, NY.
56. Romling, U., M. Gomelsky, and M. Y. Galperin. 2005. C-di-GMP: the dawning of a novel bacterial signaling system. *Mol. Microbiol.* **57**:629–639.
57. Salyers, A. A., A. Reeves, and J. D'Elia. 1996. Solving the problem of how to eat something as big as yourself: diverse bacterial strategies for degrading polysaccharides. *J. Ind. Microbiol.* **17**:470–476.
58. Schultz, J., F. Milpetz, P. Bork, and C. P. Ponting. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. USA* **95**:5857–5864.
59. Servant, F., C. Bru, S. Carrere, E. Courcelle, J. Gouzy, D. Peyruc, and D. Kahn. 2002. ProDom: automated clustering of homologous domains. *Brief. Bioinform.* **3**:246–251.
60. Shipman, J. A., J. E. Berleman, and A. A. Salyers. 2000. Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*. *J. Bacteriol.* **182**:5365–5372.
61. Shipman, J. A., K. H. Cho, H. A. Siegel, and A. A. Salyers. 1999. Physiological characterization of *susG*, an outer membrane protein essential for starch utilization by *Bacteroides thetaiotaomicron*. *J. Bacteriol.* **181**:7206–7211.
62. Singh, A., and K. Hayashi. 1995. Microbial cellulases: protein architecture, molecular properties, and biosynthesis. *Adv. Appl. Microbiol.* **40**:1–44.
63. Sonnhammer, E. L., S. R. Eddy, and R. Durbin. 1997. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* **28**:405–420.
64. Stanier, R. Y. 1942. The cytophaga group: a contribution to the biology of myxobacteria. *Bacteriol. Rev.* **6**:143–196.
65. Sun, H., D. R. Zusman, and W. Shi. 2000. Type IV pilus of *Myxococcus xanthus* is a motility apparatus controlled by the *frz* chemosensory system. *Curr. Biol.* **10**:1143–1146.
66. Taylor, B. L., and I. B. Zhulin. 1999. PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol. Mol. Biol. Rev.* **63**:479–506.
67. Taylor, L. E. N., B. Henrissat, P. M. Coutinho, N. A. Ekborg, S. W. Hutcheson, and R. M. Weiner. 2006. Complete cellulase system in the marine bacterium *Saccharophagus degradans* strain 2-40T. *J. Bacteriol.* **188**:3849–3861.
68. Vingadassalom, D., A. Kolb, C. Mayer, T. Rybkine, E. Collatz, and I. Podglajen. 2005. An unusual primary sigma factor in the Bacteroidetes phylum. *Mol. Microbiol.* **56**:888–902.
69. Wall, D., and D. Kaiser. 1999. Type IV pili and cell motility. *Mol. Microbiol.* **32**:1–10.
70. Warren, R. A. 1996. Microbial hydrolysis of polysaccharides. *Annu. Rev. Microbiol.* **50**:183–212.
71. Wilson, D. B. 1992. Biochemistry and genetics of actinomycete cellulases. *Crit. Rev. Biotechnol.* **12**:45–63.
72. Winogradsky, S. 1929. Etudes sur la microbiologie du sol. *Ann. Inst. Pasteur* **43**:549–633.
73. Wolgemuth, C., E. Hoiczky, D. Kaiser, and G. Oster. 2002. How myxobacteria glide. *Curr. Biol.* **12**:369–377.
74. Zhulin, I. B., A. N. Nikolskaya, and M. Y. Galperin. 2003. Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in bacteria and archaea. *J. Bacteriol.* **185**:285–294.