JOURNAL OF BACTERIOLOGY, May 1995, p. 2543–2549
0021-9193/95/$04.00+0

Vol. 177, No. 9

Comparative Analyses Reveal a Highly Conserved Endoglucanase in the Cellulolytic Genus *Fibrobacter*

CHUZHAO LIN AND DAVID A. STAHL*

Departments of Veterinary Pathobiology and Microbiology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

Received 29 September 1994/Accepted 17 February 1995

An RNA probe complementary to the endoglucanase 3 gene (*cel-3*) of *Fibrobacter succinogenes* S85 hybridized to chromosomal DNAs from isolates representing the genetic diversity of the genus. The probe was subsequently used to identify putative *cel-3*-containing clones from genomic libraries of representative *Fibrobacter* isolates. Comparative sequence analyses of the cloned *cel-3* genes confirmed that *cel-3* is conserved among *Fibrobacter* isolates and that the ancestral *cel-3* gene appears to have coevolved with the genus, since the same genealogy was inferred from sequence comparisons of 16S rRNAs and *cel-3* genes. Hybridization comparisons using a xylanase gene probe suggested similar conservation of this gene. Together the data indicate that the cellulolytic apparatus is conserved among *Fibrobacter* isolates and that comparative analyses of homologous elements of the apparatus from different members, in relationship to the now established phylogeny of the genus, could serve to better define the enzymatic basis of fiber digestion in this genus.

Fibrobacter is one of the genera comprising the major cellulolytic bacteria of the rumen and cecum (20, 41). Comparative 16S rRNA sequencing studies initially established *Fibrobacter* as a genetically diverse yet phylogenetically coherent genus (31) composed of two well-resolved lines of descent (species). This finding was confirmed by additional comparative sequencing and genomic DNA similarity studies (1). Prior to these studies, all isolates were classified as strains of a single species (previously *Bacteroides succinogenes*). Although great genetic diversity is now apparent (ca. 92% 16S sequence similarity and less than 20% DNA homology between the two named *Fibrobacter* species), there are as yet no well-defined physiological traits that reflect this diversity.

For a variety of reasons, particularly the relatively recent recognition of the great genetic diversity within the genus, strain S85 of *Fibrobacter succinogenes* (the type species) has been most studied (9). The cellulolytic properties of that organism have been a research focus and demonstrated that fibrobacters, as is characteristic of other cellulolytic organisms, possess multiple polysaccharidases (e.g., cellulases and xylanases) with different substrate specificities and gene sequences (6, 9, 27, 28). To date, six different polysaccharidase gene sequences from *Fibrobacter* isolates are available in GenBank. At least nine additional cellulase and four xylanase genes have been identified in *F. succinogenes* S85 (27, 28).

The available data suggest that the overreaching and defining characteristic of the genus *Fibrobacter* is the capacity to release glucose and glucose polymers from plant cell walls (fiber digestion), the near-exclusive substrates for characterized members of the genus. However, two immediate questions arise. What is the subunit composition of the fiber-depolymerizing apparatus, and what unique characteristics define the distinct genealogical lineages within the genus? Given that these organisms are specialists in fiber digestion, we anticipate that the defining differences may be found within the enzyme systems participating in plant polymer degradation by individual organisms. If so, it is important to study the cellulolytic properties of *Fibrobacter* spp. comparatively. This was in part the motivation for the present study.

Since the genus is phylogenetically coherent, it is plausible that the immediate ancestor of the genus was cellulolytic. If so, then the cellulolytic apparatus of contemporary members may in part be derived from the ancestral type through speciation (orthologous) and in part from lateral transfer of genes associated with fiber digestion, originating from within the genus or from without. Since the cellulolytic apparatus is complex, the identification of enzymes present in the ancestor should provide a useful framework for identifying the core elements of the contemporary apparatus. For example, polysaccharidases not present in the ancestor may in part define lineage-specific differences, whereas conserved components would be anticipated to serve common roles in fiber digestion. The identification of individual enzyme origin can be approached through phylogenetic analyses of the enzymes in relationship to the phylogeny of the genus. If the enzymes are orthologous, then the phylogeny of the enzymes should parallel that of the organisms with which they are associated. Our past studies examined phylogenetic relationships within the genus by using 16S rRNA sequencing and DNA homology (1, 31). In this study we demonstrate that endoglucanase 3 gene (*cel-3*), first identified in *F. succinogenes* S85 (6), is conserved within the genus and that its sequence divergence parallels speciation within the genus. A similar parallel relationship was suggested by hybridization analysis of a conserved xylanase. Thus, the *cel-3* and possibly xylanase genes appear to be conserved components of the ancestral cellulolytic apparatus.

MATERIALS AND METHODS

Bacterial strains and plasmids. Cultures of *F. succinogenes* and *F. intestinalis* were grown at 37°C in 2XAB broth (26) with cellobiose as the sole carbon source. *Escherichia coli* LE392 and P2392 (Stratagene, La Jolla, Calif.) were used as hosts for lambda phage and grown in TB media (5 g of NaCl and 10 g of tryptone [Difco, Detroit, Mich.] per liter [pH 7.4]) supplemented with 10 mM $MgSO_4$ and 0.2% maltose. *E. coli* JM109 was used as the host for plasmid pGEM3Zf(−) (Promega, Madison, Wis.) and was maintained on 2XYT medium (16 g of tryptone, 10 g of yeast extract (Difco), and 10 g of NaCl per liter [pH 7.4]). Plasmids pRE3 and pBx7, containing 2.3- and 3-kb fragments encoding the endoglucanase 3 and xylanase genes, respectively, from *F. succinogenes* S85 were

* Corresponding author. Present address: Department of Civil Engineering, Northwestern University, 2145 Sheridan Ave., Evanston, IL 60208-3109. Phone: (708) 491-4997. Fax: (708) 491-4011. Electronic mail address: d-stahl@nwu.edu.

kindly provided by C. Forsberg (University of Guelph, Guelph, Ontario, Canada). The pRE3 clone was subsequently subcloned (during this research) in the *Bam*HI site of pGEM and designated pRE3-GEM; a 2.3-kb *Kpn*I-*Pst*I fragment of pBx7 was subcloned in pGEM and designated pKP/P-GEM.

Heterologous hybridization. Genomic DNA was isolated by using a modified Kirby-Marmur procedure (1, 29). Digestion with restriction enzymes (*Eco*RI, *Bam*HI, *Hin*dIII, and *Pst*I) was carried out as specified by the manufacturer (BRL, Gaithersburg, Md.). Digestion products were separated by electrophoresis on 0.6% agarose gels in Tris-acetate-EDTA (TAE) buffer and visualized by ethidium bromide staining (35). DNA fragments were prepared by depurination in 0.2 N HCl (two times, 15 min each) followed by denaturation in 0.4 N NaOH (two times, 15 min each) and subsequently transferred to a nylon membrane (Bio-Rad, Richmond, Calif.) under alkaline conditions (0.4 N NaOH) with a vacuum transfer apparatus (35). The membranes were baked for 2 h at 80°C prior to hybridization.

RNA transcripts radiolabeled with ^{32}P (riboprobes) were synthesized from plasmids pRE3-GEM and pKP/P-GEM as specified by the manufacturer (Promega). Plasmids pRE3-GEM and pKP/P-GEM were first linearized with *Hin*dIII and *Pst*I, respectively. The linearized plasmids were extracted with an equal volume of phenol-chloroform equilibrated with TE (10 mM Tris-HCl, 1 mM EDTA [pH 8.0]) and recovered by ethanol precipitation. The riboprobe reaction mix contained 0.6 μg of DNA, 100 μCi of [α-^{32}P]CTP (ICN Biomedicals, Irvine, Calif.), and 10 U of bacteriophage T7 RNA polymerase (BRL) in 1× T7 buffer (40 mM Tris-HCl, 6 mM $MgCl_2$, 10 mM dithiothreitol, 50 μg of bovine serum albumin per ml, 4 mM spermidine, 0.4 mM each ATP, GTP, and UTP, 0.08 mM CTP [pH 7.5]) at 37°C for 30 min. The labeled transcript (probe) was separated from unincorporated nucleotides by using a Nensorb 20 column (DuPont Corporation, Wilmington, Del.) according to manufacturer's specifications, using 50% methanol for elution.

Hybridization and all wash steps were carried out in rotating glass tubes, using a convective hybridization incubator (Robbins Scientific, Sunnyvale, Calif.). The membranes were first wetted in hybridization buffer [6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 0.5% (wt/vol) sodium dodecyl sulfate (SDS), 10× Denhardt's solution (35), 40% (vol/vol) formamide, 100 μg of poly(A) (Sigma, St. Louis, Mo.)] and incubated (prehybridized) in 5 ml of hybridization buffer at 40°C (low stringency) or 50°C (high stringency) for at least 1 h with hybridization buffer. Labeled transcript was then added to approximately 10^6 cpm/ml, and hybridization continued overnight (12 to 16 h). Membranes were washed with 2× SSC–0.5% SDS at 40°C (low stringency) or 50°C (high stringency), air dried, and exposed to Kodak XRP film (Eastman Kodak Co., Rochester, N.Y.).

Genomic DNA (100 μg) was partially digested with *Sau*3A or *Mbo*I (BRL). The digested products were fractionated on 10 to 40% linear sucrose gradients (35) by centrifugation at 29,000 rpm (140,000 × *g*, 20 h at 20°C) using a Beckman SW41 rotor. Fractions were collected through a needle inserted in the bottom of the centrifuge tube. DNA fragments of approximately 10 kb were used for ligation with *Bam*HI-digested lambda-Dash (Stratagene, La Jolla, Calif.) according to manufacturer's instructions. The ligation reaction product was packaged in phage lambda by using a Gigapack Gold kit (Stratagene) and titered on *E. coli* LE392. Subsequent manipulations of the phage used *E. coli* P2392 as the host to limit recombination.

Clones containing *cel-3* were identified by using the riboprobe transcribed from pRE3-GEM, and isolation was verified by four serial plaque hybridizations (35). Hybridization conditions were the same as used for the low-stringency hybridization specified above. Restriction maps were constructed by both complete and partial digestion techniques (35). Two synthetic oligonucleotide probes (5′-TACGACTCACTATAGGG-3′ and 5′-CTAAAGGGAACGAATTCG-3′), complementary to the T7 and T3 promoter regions, respectively, were used for partial digestion mapping. The probes were 5′ end labeled with [γ-^{32}P]ATP, using polynucleotide kinase (40). The recombinant lambda DNA was partially digested with selected restriction enzymes, and the partial digestion products were separated on agarose gels. The DNA was then transferred to nylon membranes by the vacuum transfer method previously described (35) and hybridized with either ^{32}P-labeled oligonucleotide probe. The sizes of the fragments hybridizing to either probe were determined by comparison with molecular weight standards following exposure to X-ray film. Completely digested lambda DNA was hybridized with the *cel-3* riboprobe, and the appropriate fragments were subcloned into the pGEM plasmid specified above for sequence determination.

Sequence determination and analysis. Sequences of *Fibrobacter cel-3* genes were determined by using the double-strand dideoxy sequencing method (35). The nested deletion method using an exonuclease III-mung bean nuclease deletion kit (Stratagene) was employed to reduce subcloning of internal restriction fragments. Certain regions were sequenced by using specific synthetic primers designed on the basis of flanking sequence information. Sequences were determined from both strands for *F. succinogenes* A3c and REH9-1 and *F. intestinalis* DR7.

Sequences obtained were analyzed with the GDE sequence editor (24) and PHYLIP (7) software packages. Phylogenetic trees were constructed from evolutionary distances estimated for each inferred amino acid sequence pairs by using the Fitch-Margoliash and least-squares methods (FITCH) with evolutionary clock (KITSCH) as well as neighbor-joining and UPGMA (NEIGHBOR) program in the PHYLIP package. Pairwise nucleic acid sequence similarities were estimated with an alignment editor (AE2) (24), using the Jukes and Cantor (22) correction and considering only the homologous positions.

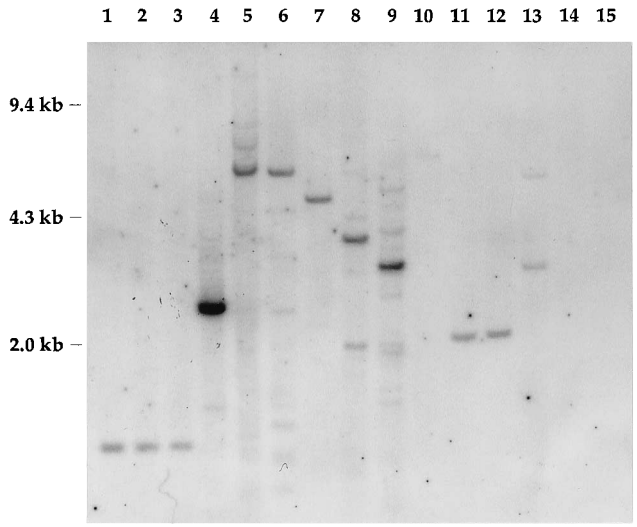

FIG. 1. Southern analysis using an RNA probe transcribed from the *cel-3* gene from *F. succinogenes* S85. The genomic DNA was digested with *Hin*dIII. Lanes: 1 to 10, *F. succinogenes* MM4, MB4, HM2, S85, A3c, B1, BL2, REH9-1, GC5, and MC1; 11 to 15, *F. intestinalis* JG1, LH1, DR7, C1a, and NR9.

Nucleotide sequence accession numbers. *cel-3* gene sequencer for *F. succinogenes* A3c, *F. succinogenes* REH9-1, and *F. intestinalis* DR7 can be obtained from GenBank under accession numbers L39838, L39839, and L39840.

RESULTS

Riboprobes transcribed from the *cel-3* and xylanase genes of *F. succinogenes* S85 hybridized to restriction fragments of most strains characterized (Fig. 1 and 2). Generally, strong hybridization was observed for only one or two restriction fragments from each, although numerous bands of lesser intensity were evident in some. All strains, with the apparent exception of strains C1a and NR9 (lanes 14 and 15), demonstrated hybridization to the *cel-3* probe. All strains hybridized to the xylanase gene probe under the specified conditions of hybridization and film exposure. Further hybridization studies of *F. succinogenes* A3c and REH9-1 and *F. intestinalis* DR7, using a higher hy-

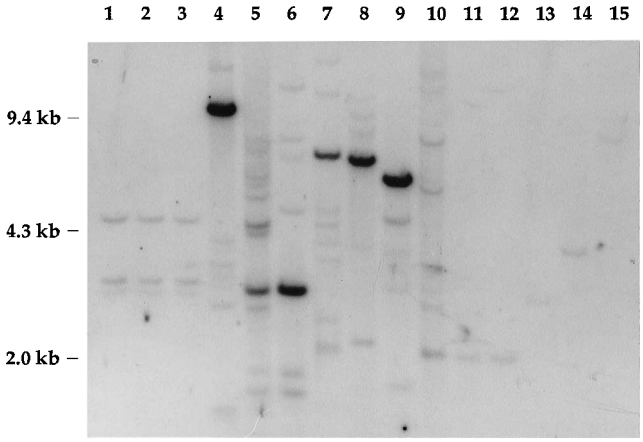

FIG. 2. Southern analysis using an RNA probe transcribed from a 2.3-kb *Kpn*I-*Pst*I fragment of the xylanase gene from *F. succinogenes* S85. The genomic DNA was digested with *Hin*dIII. The order of strains applied to each lane is the same as in Fig. 1.

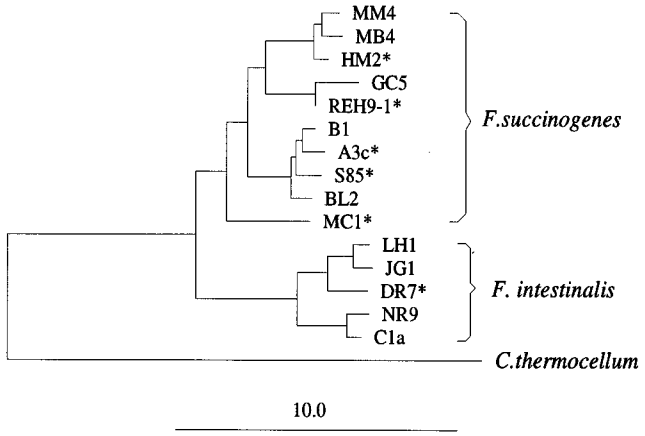

FIG. 3. *Fibrobacter* phylogeny based on comparative 16S rRNA sequence analysis and using the 16S rRNA sequence of *C. thermocellum* as an outgroup. An asterisk indicates that a genomic library has been constructed for the strain.

bridization and washing stringency, demonstrated hybridization of the *cel-3* probe to only one restriction fragment, suggesting the presence of one gene copy.

The intensity of hybridization generally reflected the relatedness of individual strains to *F. succinogenes* S85, the source of the *cel-3* and xylanase gene probes. Phylogenetic relationships among members of the genus, inferred from 16S rRNA sequence divergence, are represented in Fig. 3. In general, the hybridization signal decreased with decreasing phylogenetic relationship. Although hybridization was not evident for two strains of *F. intestinalis* (C1a and NR9), the presence of a homologous gene in these strains was supported by identification of a single antigenically related band on total protein polyacrylamide gels in strain NR9 (25). Strains NR9 and C1a are related at approximately 95% DNA homology (1).

The restriction digestion analysis also revealed a restriction site modification common to representatives of *F. succinogenes* lineage composed of strains HM2, MM4, and MB4. Modification of the GATC sequence common to *Sau*3A and *Bam*HI restriction enzymes was suggested by failure of these enzymes to digest genomic DNA isolated from these strains. Modification of the cytosine nucleotide within the common recognition sequence was indicated by the ability of an isoschizomer of *Sau*3A, *Mbo*I, to restrict the DNA. Although of little determinative value, the presence of a restriction site modification common to, and restricted to, this group is consistent with the phylogenetically based classification.

Genomic libraries were constructed in a phage lambda cloning vector for selected strains representing the diversity of the genus *Fibrobacter* now recognized in pure culture (Fig. 3). Clones hybridizing to a labeled RNA transcript of the *F. succinogenes* S85 *cel-3* gene were isolated from phage libraries of *F. succinogenes* REH9-1, A3c, MC1, and HM2 and *F. intestinalis* DR7 and partially mapped prior to sequence analysis. More extensive mapping of five overlapping *cel-3*-positive clones from *F. succinogenes* A3c was completed as shown in Fig. 4. A 2.5-kb region containing the putative *cel-3* gene from this strain, flanked by *Sph*I-*Hin*dIII restriction sites, was subcloned in pGEM for sequence determination. Riboprobes generated from each of the cloned genes hybridized specifically with the restriction fragment corresponding to the dominant hybridization signal in Fig. 1. We have not examined hybridization of these riboprobes to restriction digests of all representative strains, and the presence of a homologous gene in *F. intestinalis* C1a and NR9 has not yet been demonstrated by hybridization or sequence analyses. However, as noted earlier, a polyclonal antibody raised against the *F. succinogenes* enzyme recognizes a single protein band resolved by one-dimensional polyacrylamide gel electrophoresis of *F. intestinalis* NR9 total protein (25). Although *F. intestinalis* C1a has not been examined for the presence of an antigenically related protein, it has 95% DNA homology with strain NR9 (1) and is anticipated to possess it. Thus, these data suggest the presence of a homologous gene in strains C1a and NR9 that was not identified by the heterologous probe at the stringency of hybridization used in this study.

Nucleotide sequences of the putative *cel-3* genes from *Fibrobacter* strains and the homologous sequences of the genes for cellodextrinase A from *F. succinogenes* S85 and endoglucanase C from *Clostridium thermocellum* (36) were manually aligned, using conserved nucleic acid and amino acid sequence motifs. The aligned amino acid sequences is shown in Fig. 5. The inferred amino acid sequences of cellodextrinase A from *F. succinogenes* S85 and endoglucanase C from *C. thermocellum* lack the amino-terminal sequence common to *cel-3* (about 330 amino acids). As discussed below, the C-terminal region corresponds to the catalytic domain of these enzymes.

Phylogenetic trees were constructed from the inferred amino acid sequences for the C-terminal region as well as the presumed N-terminal cellulose binding domain (CBD) (Fig. 6). All methods of inference generated identical tree topologies. Identical relationships were inferred with either the N- or C-terminal region or the entire sequence. DNA similarity values for each pair of sequences corresponding to the N-terminal domain and the C-terminal domain are shown in Table 1. Strains with the greatest 16S rRNA similarity (*F. succinogenes* S85 and A3c) also exhibited the most similar *cel-3* gene sequences. *C. thermocellum* and *Fibrobacter* spp. have about 80% 16S rRNA similarity. The clostridial endoglucanase C gene (*clocelca*) is about 42% similar to the C-terminal domain of the *Fibrobacter cel-3* genes. The cellodextrinase gene from *F. suc-*

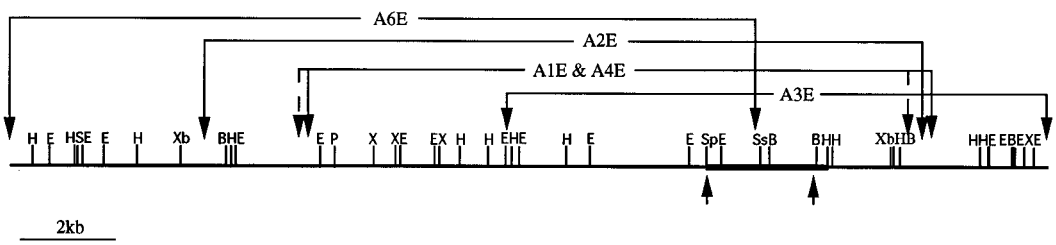

FIG. 4. Restriction map of the *cel-3* gene and its flanking regions of *F. succinogenes* A3c. The region between *Sph*I and *Hin*dIII (thick line) was subcloned in pGEM as pGA3c. The *cel-3* coding region is within the up-pointing arrows. Abbreviations: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; S, *Sal*I; Sp, *Sph*I; Ss, *Sst*I; Xb, *Xba*I; X, *Xho*I.

```
A3c     0     MQLKKFYPQISILGIATVMAL--SACGDDNTQIPLPSDPGSEIRDS------IPNNNPQPGTDSIPAIDTTSTDTSKTLPPTELPAEGPITLPQGLGVLV
S85     0     MQLKNFYPKMSVLGIATVMAL--TACGDENTQALFANNPVPGAENQ------VPVSSSDMSPTSSDAVIDPTSSSAAVVDPSTLPAEGPITMPEGLGTLV
9-1     0     MNFKT--KSSLALSAAALLFL--SLCGDDSSKSAAPDEPAVNPTDSTTIPTPLPTDSTTVPADSTANPTDSSAATNPTVDPSTLPAEGPITLPEGKGLLV
DR7     0     MRFRNL----ALISAASAFALWTSACSDDKS---------SAPEEDAPLVSSVVEESSSSVADKNDPESSSSDSTTASSSSIAAPTN---------GLLL
cedA    0     MK-------------------------------------------------------------------------------------------------
celC    0     MV-------------------------------------------------------------------------------------------------

A3c     92    DDFEDANHES-KIGDFWYTYDDKDNGGASSIVTPLNDDGDIMPGRVNNGSNYALQVNYTLDRGEYEYDPYVGWGLQITEDSANGRFGGITYWYKGGAHEV
S85     92    DDFEDGDNLS-KIGDYWYTYNDNDNGGASIITTPLNEEENIIPGRVNNGSNYALQVNYTLDRGDYEYDPYVGWGVQVAPDEANGHFGGLTYWYKGGAHEV
9-1     96    DDFEDGDNHSATIDNYWYTYSDNDNYCASVITTPVNEDGDIIRGAVNNGSKYALQVNYTLDKGEYAYDPYVGWGIQVAEDDANGRFGGITYWYKGGAHEV
DR7     78    DDMEDGDG-GTSIGSGWYTYNDNDNKAASVITTPVDESGNPIASPTDNGSSYAFKVEFTLDKGDYAYDPYVGWGFEVRARR-RQQVSGIRYSYKGAAHYI
cedA    2     -------------------------------------------------------------------------------------------------
celC    2     -------------------------------------------------------------------------------------------------

A3c     191   HIEVSDVKDYDVHLAKVPASRSWKQAVIRFKDLVQGGWGEEVVFDAKHLKAISFQAKGSKSKVMTDSLFIDNIYLQDTSEVEKDQPDMEIKDPVIPTVTF
S85     191   HIEITDVEDYDVHLAKFPASRTWKQAVVRFKDLVQGGWGKEIPFDAKHIMAISFQAKGNKSKLVTDSLFIDNIYLQDSSEVEKDQPDMEIKDPVIPVVEF
9-1     196   HIEISDVTDYDVHLAKVKASRTWTQAVIRFKDLVQGGWGAEVPFDAKHIRAISFQAKGN-AKVTSDSLFIDNIYLQDTSEVEADKPDMTINDPVIPKIEF
DR7     176   HVETSDVIDYDVHLSAVKKSDTWTTVTIDFNNLVQGGWGERVAFDPAHVTNISFQVKGN-GKV--DSVMIDDVYFIMEEDLPPKTADMTIHAPQVISPNI
cedA    2     ------------------------KERLRGVNLGGWFSQV-----------------------------------DCIEEKDPVGFPGVVGHIKTF
celC    2     -------------------------SFKA-GINLGGWISQY-----------------------------------QVFSK----------EHFDTF

A3c     291   TEAEITVTNPLQEKAMKYLNKGVNFTNWLENADGKFKS-FELGEKDVQILAENGFKSLRLPIDLDLYATNRDAFVKGTDTELK--FDDDTLFLVLDSFVE
S85     291   TEAEITVTNPLQEKAMKYLNKGVNFTNWLENADGKFKS-FELGESDVKILADNGFKSLRLPIDLDLYATNRDAFIAGTDTELK--FDDDTLFLVLDSFVE
9-1     295   TEAEITVTSPLQAKAMKYLNKGINFTNWLENADGKFKE-LLFDATDIKLLADNGIKSLRLPIDLDLYATNRDAFVKDTTGTVKLAFDDSTLFTVLDSFVE
DR7     273   --GDITISTPLQEKAMKYLDKGVNFTNWLEEADGKFTGKFELGENDVKILSENGFKALRLPIDLDLYVDNRAEYLADTTGTVKLQMNDS-IYIVLDSFVN
cedA    39    ----------------------------------------LGTSDFKRIHDAGFNHVRLPVD---Y------FNLFKGDDLK-P-DEEI-FALLDKALK
celC    28    ----------------------------------------ITEKDIETIAEAGFDHVRLPFD---Y------PIIESDDNVG-EYKEDG-LSYIDRCLE

A3c     388   WTGKYNMSLVIDYHEY-DNSYNTTSVKDTNYI------KMMAEVWKHVAAHYAESPREDLFFELLNEPDMSDGKVTAAQWTVAAQAMIDAIRTVDTKHTI
S85     388   WTAKYNMSFVIDYHEY-DNSYNTTSAKDPNYI------KMMAETWKHVAAHYAESPREDLFFELLNEPDMSDGKVTAATWTTAAQAMIDAIRTVDTKHTI
9-1     394   WTGNYNMSFVIDYHEY-DNSYNATSAKDNKYI------QMMAETWKHVAAHYAENTREDIFFELLNEPDMSAGKVTAAQWTVAAQAMIDSIRSVDKKHTI
DR7     370   WTKRYGMSLTIDYHEY-DNSYNVTTSTDAKYL------AMMANVWKAVAAHYASNEREDLFFELLNEPDMKDGKVTQANWTIAAQGMIDSIRTVDTKHTI
cedA    86    DIQNADLDVILDLHKCPGHDFHLASNHEQAFFADANARKDTRKIWAFMAERY-SS-MPRVMMELLNEPAASDSKV----WDKVKDEIFWEIRKHAPKNTI
celC    76    WCKKYNLGLVLDMHHAPGYRFQ--DFKTSTLFEDPNQQKRFVDIWRFLAKRYINE-REHIAFELLNEVVEPDSTR----WNKLMLEYIKAIREIDSTMWL

A3c     481   LFGDAQWYSISLLAKRTPFTDDNIIYVIHTYEPFAFTHQGGSWTDYATIHDIPFPYDPAKWSTVSGDFGVNKSTKAYVKTNIKNYYKTG------SKEAI
S85     481   LFGDAQWYSITLLAKRTPFTDDNIIYVIHTYEPFAFTHQGGSWTDYATIHDIPFPYDPAKWSTVSGDFGVNKSTKSYVKTNIKNYYKTG------SKEAI
9-1     487   LFGDAQWYSISLLAKRTPFTDDNIVYVIHTYEPFAFTHQGGSWTDYATIKGIPFPYDPAKWSTVSGDFGVTKSTQSYVKSNIKNYYKTG------SKEAI
DR7     463   IFGDAQWYSIALLAKRTPFTDDNIIYAVHSYEPYIFTHQGASWTEASSIKNLPFPYDTTKWSVYSSDFGVNASTASWIKSAVKNYYKTG------SKGYI
cedA    180   VVGANKWNSAREFEFLTPLDDDNAIYSFHTYTPVTFTHQGAAWID------DPF-FKIER--PWPGDYA---APE--VGGTTR----LNVEYGRWDKAQL
celC    169   YIGGNNYNSPDELKNLADIDDDYIVYNFHFYNPFFFTHQKAHWSE------SAMAYNRTV--KYPGQYE---GIEEFVKNNPKYSFMMELNNLKLNKELL

A3c     575   LEQILKAKKWAATNNVPVIINEFGALNLRSTAESRINYLTAMREICDTLQIPWTHWGYTG-NFSVIENGKLIEGLDKALGVG---------------K*
S85     575   LEQILKAKKWAATNNVPVIINEFGALNLRSTAESRLNYLTAMREICDTLQIPWTHWGYTG-NFSVIENGKLIEGLDKALGVGS--------------K*
9-1     581   MVEILKAKKWAATNNVPVIINEFGALNLRSTAQDRLNYLTAMREICDTLQIPWTHWGYTG-NFSLFEGGSQGNQAHRRYRQGPWLGAA---------E*
DR7     557   LNQIYTAKKWAVENQVPVIINEFGAYNVASTAEDRLNYLTAVRESCDSLQVPWQHWGYTG-GFEVIRDGKLIDGMDKALGL-----------------*
cedA    262   QASIQNALDFRAKYDLPVSCNEFGVY-VQVPRKYQLAWMRDFLDILRDADVGYSYWNYKNLDFGLVSKGESLHNNLEQYNNPERLDRELMEMIA--KG*
celC    258   RKDLKPAIEFREKKKCKLYCGEFGVI-AIADLESRIKWHEDYISLLEEYDIGGAVWNYKKMDFEIYN-----------EDRKPVSQELVNILARRKT*
```

FIG. 5. Aligned amino acid sequences of endoglucanase 3 of *Fibrobacter* spp., cellodextrinase (*cedA*) of *F. succinogenes* S85, and endoglucanase C (*celC*) of *C. thermocellum*.

cinogenes S85 is related to the C-terminal domain of *cel-3* at about the 48% similarity level.

DISCUSSION

Functional domains of endoglucanase 3 of *Fibrobacter* spp. On the basis of sequence similarities, endoglucanase 3 and endoglucanase C, encoded by *cel-3* of *Fibrobacter* spp. and *clocelca* of *C. thermocellum*, respectively, and the cellodextrinase (encoded by *celca*) from *F. succinogenes* S85 are members of the A3 cellulase family (11). However, the *celca* and *clocela* genes are approximately half the size of *cel-3*. The derived amino acid sequences align only with the C-terminal region of endoglucanase 3. The C-terminal region appears to be the catalytic domain (CD). The *celca* gene product, endoglucanase C, is enzymatically active (34, 37) against carboxymethyl cellulose and *p*-nitrophenolcellobioside. A truncated *cel-3* gene product (containing only the C-terminal region from *F. succinogenes* S85 [30] and A3c [25]) expressed in *E. coli* also retained catalytic activity on carboxymethyl cellulose and *p*-nitrophenolcellobioside. Amino acid residues important to catalytic activities (2, 3) were found to be conserved in the C-terminal domain of the *cel-3.*

The N-terminal region of endoglucanase 3 is also conserved among *Fibrobacter* spp. and has been suggested to be a CBD (30) even though this function has yet to be directly demon-

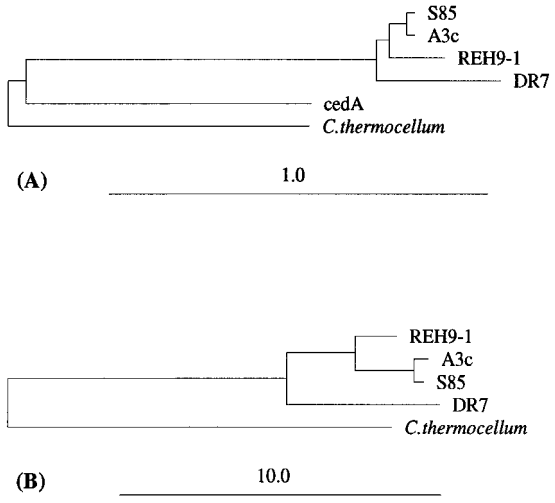

FIG. 6. (A) Phylogenetic tree based on comparative analysis of the amino acid sequences deduced from the *cel-3* gene of *Fibrobacter* spp., the *cedA* gene from *F. succinogenes* S85, and the *celC* gene of *C. thermocellum*, using the FITCH program in the PHYLIP package (7). Only the C-terminal 270 amino acids were considered. A phylogenetic tree based on comparative 16S rRNA sequence analysis (1) is presented for comparison (B).

TABLE 1. Nucleic acid sequence similarity data[a]

	cel-3				*cedA*	*celC*
	A3c	S85	REH9-1	DR7		
cel-3						
A3c		91.1	82.1	69.7	49.1	43.2
S85	81.6		82.7	70.5	48.8	42.7
REH9-1	78.2	78.8		68.9	47.7	40.2
DR7	60.6	60.2	62.5		44.8	40.5
cedA	NA	NA	NA			44.9
celC	NA	NA	NA	NA		

[a] The lower left half is the similarity value (percentage) between pairs of nucleic acid sequences corresponding to the N-terminal region (amino acids 89 to 269 of *F. succinogenes* S85). The upper right half corresponds to the C-terminal domain (270 to stop codon). NA, not applicable.

strated. Additionally, it contains a putative lipoprotein signal sequence for cleavage (13, 30), with two basic amino acids preceding the consensus sequence L-X-Y-C. Interestingly, there are other carbohydrate hydrolases suggested to be lipoproteins. These include an endoglucanase from *Pseudomonas solanacearum* (18), two pullulanases from *Klebsiella pneumoniae* and *K. aerogenes* (5, 23), and a chitobiase from *Vibrio harveyi* (39). The significance of this covalent modification is unknown, but it has been suggested that the lipid serves to retain hydrophilic proteins in the vicinity of the cellular membrane (45) and thus ensure the availability of enzymatic digestion products to the bacteria. A phase separation study using Triton X-114 provided indirect physical evidence for the *cel-3* gene product of *F. intestinalis* DR7 being a lipoprotein, as the *cel-3* gene product from *F. intestinalis* DR7 was demonstrated to partition in the Triton phase (25).

A stretch of linker sequence (11) rich in serine, threonine, and proline immediately follows this putative signal sequence and precedes a more conserved part of the N-terminal region. Thus, endoglucanase 3 appears to be organized as linker-CBD-CD. However, the putative linker sequences are highly variable. *F. intestinalis* DR7 has a much shorter linker sequence. Sequences following the linker (about 180 amino acids in length) are more conserved. It is apparent from Table 1 that the 180-amino-acid sequence of the N-terminal region (positions 89 to 269, strain S85 amino acid numbering) of endoglucanase 3 is less conserved than the C-terminal region (270 to stop codon 659). Thus, the CD is more conserved than the putative CBD.

Regulatory features of *Fibrobacter* cellulase genes. Putative promoter sequences were identified for *F. succinogenes* A3c and *F. intestinalis* DR7 *cel-3* genes. These are very similar to those previously identified for *F. succinogenes* S85 (30). However, no obvious promoter sequence could be identified in the sequence of *F. succinogenes* REH9-1, as shown in Fig. 7. It is apparent from this figure that the putative *cel-3* gene promoter is very different in organization from *E. coli* σ^{70} and σ^{70}-like promoters (12). The spacing between the putative −35 and −10 regions appear to vary from 21 to 23 bp, in contrast to the 17 ± 1-bp spacing characteristic of *E. coli* (12, 32). No consistent promoter organization can be found for *Fibrobacter* spp. when the putative *cel-3* promoter sequences are compared with putative promoter sequences from the other glycosyl hydrolases included in Fig. 7.

Origin and conservation of cellulase genes. The results of these comparative studies must be considered in the context of the extensive collection of sequences for glycosyl hydrolases now available. This collection, which includes representatives of all three domains (44), has been classified into various families based on sequence relationships (11, 15–17). This classification revealed that there has been frequent transfer of these genes between the three domains. Additionally, it was apparent that enzymes having similar substrate specificities could belong to different glycosyl hydrolase families, and conversely, enzymes associated with a single family could have different substrate specificities. The acquisition of new substrate specificities by glycosyl hydrolases has been suggested to be a common evolutionary event (15), and cellulase genes have been shown to evolve by gene duplication (10) or region shuffling (43). In an attempt to reconcile the difference between functional and phylogenetically based classification schemes, a new classification system based on amino acid sequence similarities of glycosyl hydrolases has been proposed to complement the existing Internation Union of Biochemists enzyme nomenclature (15, 16). The revised classification emphasizes common structural features of the enzyme rather than common substrate specificity. This classification scheme is of particular interest because it implicitly embodies the evolutionary history of the genes.

Given the now recognized diversity of depolymerizing enzymes encoded by individual cellulolytic organisms and their multiple evolutionary origins, full appreciation of the depolymerizing apparatus must include evolutionary considerations. In this regard, it is essential to distinguish between paralogous (divergence following gene duplication), orthologous (divergence through speciation), and xenologous (obtained via lateral transfer) genes. This discrimination among glycosyl hydrolases is not straightforward, since the specific functions of many or most of these enzymes have yet to be clearly defined. The present study suggested that resolution can be assisted by consideration of the evolutionary history of the organism(s) as well

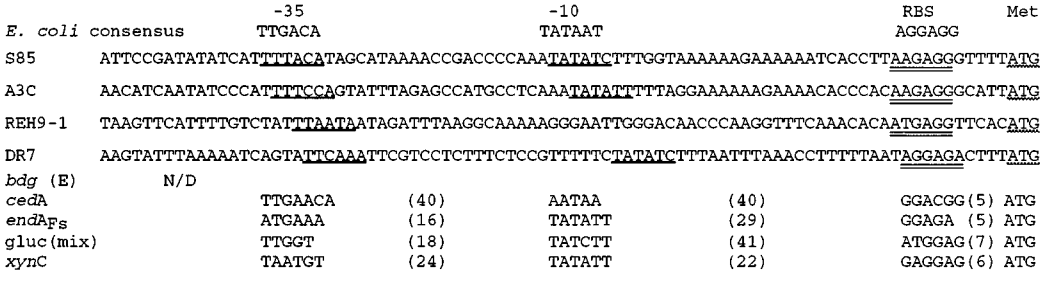

FIG. 7. Putative promoter sequences for *Fibrobacter cel-3* genes. Both −35 and −10 regions are underlined. The ribosomal binding site (RBS) is double underlined. The ATG start codon is marked by a wavy underline. The regulatory features from other *Fibrobacter* glycosyl hydrolase genes retrieved from GenBank are included for comparison (4, 21, 30, 33, 42), including an unpublished family E β-1,4-endoglucanase gene, *bdg*. Numbers in parentheses indicate the number of nucleotides between the two adjacent features. N/D, not identified. *endA*$_{Fs}$, *endA* from *F. succinogenes*; gluc(mix), mixed-linkage β-glucanase from *F. succinogenes*.

as the enzyme systems. This comparison supported the orthologous nature of endoglucanase 3 gene (*cel-3*) of *Fibrobacter* spp., as indicated by congruent genealogies inferred from sequence divergence of *cel-3* genes and 16S rRNAs. Thus, the *cel-3* enzyme is apparently a conserved element within the depolymerizing enzyme complex of this group of organisms. This finding implies that a *Fibrobacter*-like organism, ancestral to contemporary strains and species, possessed the progenitor of endoglucanase 3. As yet, we have no understanding of the ecological reflection of speciation within this genus (1). Of some interest in this regard is speculation that the origin and diversification of *Fibrobacter* spp. coincided with the invasion of land in the Silurian (38) by plants with lignin or lignin-like constituents in their cell walls. If this is so, diversification of the group in part might coincide with the evolution of different plant varieties (and associated cell wall types).

Flint et al. also reported gene conservation among fibrobacters isolated by using another cloned cellulase as a hybridization probe (8). Huang et al. demonstrated the presence of cellobiosidase of the same antigenicity among a collection of fresh *Fibrobacter* isolates (19). The present study also suggested that a xylanase gene is conserved among *Fibrobacter* isolates.

Both endoglucanase C from *C. thermocellum* and cellodextrinase A from *F. succinogenes* S85 are homologous to the C-terminal region of endoglucanase 3 of *Fibrobacter* spp. However, they are presumably not orthologous, since both lack the amino-terminal binding domain common to *cel-3* (see below). It is of interest that the sequence of the cellodextrinase A gene from *F. succinogenes* S85 is related, but peripheral, to the *cel-3* gene sequences. This finding suggests that it is derived from a gene ancestral to *cel-3* but is of more recent origin than that encoded by the clostridium. However, the available data cannot distinguish between origin within the *Fibrobacter* progenitor or via acquisition (xenologous). Additional comparative sequencing is necessary to provide evidence of origin.

The major implication of these results is that the characterized cellulases and xylanases are integral components of a conserved fiber-digesting apparatus. A similar observation of coevolution of a complex enzyme system is congruent divergence of the nitrogenase genes and 16S rRNAs of *Rhizobium* species (14). However, this observation still begs the question of what unique attributes define the individual lineages of *Fibrobacter* spp. Such attributes might include differential regulation of expression of conserved elements of the fiber-digesting enzyme array or variation in other enzymes peripheral to the conserved array.

ACKNOWLEDGMENTS

This research was supported by grants from the U.S. Environmental Protection Agency and the U.S. Department of Agriculture to D.A.S.

We thank Cecil Forsberg at the University of Guelph, Guelph, Ontario, Canada, for providing the two plasmids used in this study. We also thank Paula Spicer, who participated in the very early stages of this study.

REFERENCES

1. **Amann, R. I., C. Lin, R. Key, L. Montgomery, and D. A. Stahl.** 1992. Diversity among *Fibrobacter* isolates: towards a phylogenetic classification. Syst. Appl. Microbiol. **15:**23–31.
2. **Beguin, P., and J.-P. Aubert.** 1994. The biological degradation of cellulose. FEMS Microbiol. Rev. **13:**25–58.
3. **Belaich, A., H.-P. Fierobe, D. Baty, B. Bausetta, C. Bagnara-Tardif, C. Gaudin, and J.-P. Belaich.** 1992. The catalytic domain of endoglucanase A from *Clostridium cellulolyticum*: effects of arginine 79 and histidine 122 mutations on catalysis. J. Bacteriol. **174:**4677–4682.
4. **Cavicchioli, R., P. D. East, and K. Watson.** 1991. *endAFs*, a novel family E endoglucanase gene from *Fibrobacter succinogenes* AR1. J. Bacteriol. **173:** 3265–3268.
5. **Chapon, C., and O. Raibaud.** 1985. Structure of two divergent promoters located in front of the gene encoding pullulanase in *Klebsiella pneumoniae* and positively regulated by the *malT* product. J. Bacteriol. **164:**639–645.
6. **Crosby, B., B. Collier, D. Y. Thomas, R. M. Teather, and J. D. Erfle.** 1984. Cloning and expression in *Escherichia coli* of cellulase genes from *Bacteroides succinogenes*, p. 573–576. *In* Proceedings of the 5th Canadian Bioenergy R & D Seminar. Elsevier Applied Science Publishers Ltd., New York.
7. **Felsenstein, J.** 1993. PHYLIP (phylogeny inference package), version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.
8. **Flint, H. J., C. A. McPherson, G. Avgustin, and C. S. Stewart.** 1990. Use of a cellulase-encoding gene probe to reveal restriction fragment length polymorphisms among ruminal strains of *Bacteroides succinogenes*. Curr. Microbiol. **20:**63–67.
9. **Forsberg, C. W., and K. Cheng.** 1990. Integration of rumen microorganisms and their hydrolytic enzymes during the digestion of lignocellulosic materials, p. 411–423. *In* D. E. Akin et al. (ed.), Microbial and plant opportunities to improve lignocellulose utilization by ruminants. Elsevier Science Publishing Co. Inc., New York.
10. **Fukumori, F., N. Sashihara, T. Kudo, and K. Horikoshi.** 1986. Nucleotide sequences of two cellulase genes from alkalophilic *Bacillus* sp. strain N-4 and their strong homology. J. Bacteriol. **168:**479–485.
11. **Gilkes, N. R., B. Henrissat, D. G. Kilburn, R. C. Miller Jr., and R. A. J. Warren.** 1991. Domains in microbial β-1,4-glycanases: sequence conservation, function, and enzyme families. Microbiol. Rev. **55:**303–315.
12. **Harley, C. B., and R. P. Reynolds.** 1987. Analysis of *E. coli* promoter sequences. Nucleic Acids Res. **15:**2343–2361.
13. **Hayashi, S., and H. C. Wu.** 1990. Lipoproteins in bacteria. J. Bioenerg. Biomembr. **22:**451–471.
14. **Hennecke, H., K. Kaluza, B. Thony, M. Fuhrmann, W. Ludwig, and E. Stackebrandt.** 1985. Concurrent evolution of nitrogenase genes and 16S rRNA in *Rhizobium* species and other nitrogen fixing bacteria. Arch. Microbiol. **142:**342–348.
15. **Henrissat, B.** 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. **280:**309–316.
16. **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.
17. **Henrissat, B., M. Claeyssens, P. Tomme, L. Lemesle, and J.-P. Mornon.** 1989. Cellulase families revealed by hydrophobic cluster analysis. Gene **81:** 83–95.
18. **Huang, J., M. Sukordhaman, and M. A. Schell.** 1989. Excretion of the *egl* gene product of *Pseudomonas solanacearum*. J. Bacteriol. **171:**3767–3774.
19. **Huang, L., M. McGavin, C. W. Forsberg, J. S. Lam, and K. J. Cheng.** 1990. Antigenic nature of the chloride-stimulated cellobiosidase and other cellulases of *Fibrobacter succinogenes* subsp. *succinogenes* S85 and related fresh isolates. Appl. Environ. Microbiol. **56:**1229–1234.
20. **Hungate, R. E.** 1966. The rumen and its microbes. Academic Press, Inc., New York.
21. **Iyo, A. H., and C. W. Forsberg.** 1994. Features of the cellodextrinase gene (cedA) from *Fibrobacter succinogenes* S85. Can. J. Microbiol. **40:**592–596.
22. **Jukes, T. H., and C. R. Cantor.** 1969. Evolution of protein molecules, p. 21–132. *In* H. N. Munro (ed.), Mammalian protein metabolism. Academic Press, Inc., New York. New York.
23. **Katsuragi, N., N. Takizawa, and Y. Murooka.** 1987. Entire nucleotide sequence of the pullulanase gene of *Klebsiella aerogenes* W70. J. Bacteriol. **169:**2301–2306.
24. **Larsen, N., G. J. Olsen, B. L. Maidak, M. J. McCaughey, R. Overbeek, T. J. Macke, T. L. Marsh, and C. R. Woese.** 1993. The Ribosomal Database Project. Nucleic Acids Res. **21:**3021–3023.
25. **Lin, C., and D. A. Stahl.** Unpublished data.
26. **Macy, J. M., J. R. Farrand, and L. Montgomery.** 1982. Cellulolytic and noncellulolytic bacteria in rat gastrointestinal tracts. Appl. Environ. Microbiol. **44:**1428–1434.
27. **Malburg, L. M., and C. W. Forsberg.** 1993. *Fibrobacter succinogenes* S85 possesses at least 9 distinct glucanase genes. Can. J. Microbiol. **39:**882–891.
28. **Malburg, L. M., D. C. Smith, H. E. Schellhorn, and C. W. Forsberg.** 1993. *Fibrobacter succinogenes* S85 has multiple xylanase genes. J. Appl. Bacteriol. **75:**564–573.
29. **Marmur, J.** 1961. A procedure for the isolation of deoxyribonucleic acid from microorganisms. J. Mol. Biol. **3:**208–218.
30. **McGavin, M. J., C. W. Forsberg, B. Crosby, A. W. Bell, D. Dignard, and D. Y. Thomas.** 1989. Structure of the *cel-3* gene from *Fibrobacter succinogenes* S85 and characteristics of the encoded gene product, endoglucanase 3. J. Bacteriol. **171:**5587–5595.
31. **Montgomery, L., B. Flesher, and D. A. Stahl.** 1988. Transfer of *Bacteroides succinogenes* (Hungate) to *Fibrobacter* gen. nov. as *Fibrobacter succinogenes* comb. nov. and description of *Fibrobacter intestinalis* sp. nov. Int. J. Syst. Bacteriol. **38:**430–435.
32. **O'Neill, M. C.** 1989. *Escherichia coli* promoters. 1. Consensus as it relates to spacing class, specificity, repeat substructure, and three-dimensional organization. J. Biol. Chem. **264:**5522–5530.

33. **Paradis, F. W., H. Zhu, P. J. Krell, J. P. Phillips, and C. W. Forsberg.** 1993. The *xynC* gene from *Fibrobacter succinogenes* S85 codes for a xylanase with two similar catalytic domains. J. Bacteriol. **175:**7666–7672.
34. **Petre, D., J. Millet, R. Longin, P. Beguin, H. Girard, and J.-P. Aubert.** 1986. Purification and properties of the endoglucanase C of *Clostridium thermocellum* produced in *Escherichia coli*. Biochimie **68:**687–695.
35. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
36. **Schwarz, W. H., S. Schimming, K. P. Ruechnagel, S. Burgschwaiger, G. Kreil, and W. L. Staudenbauer.** 1988. Nucleotide sequence of the celC gene encoding endoglucanase C of *Clostridium thermocellum*. Gene **63:**23–30.
37. **Schwarz, W. H., S. Schimming, and W. L. Staudenbauer.** 1987. Degradation of barley β-glucan by endoglucanase C of *Clostridium thermocellum*. Appl. Microbiol. Biotechnol. **29:**25–31.
38. **Shear, W. A.** 1991. The early development of terrestrial ecosystems. Nature (London) **351:**283–289.
39. **Soto-Gil, R. W., and J. W. Zyskind.** 1989. N,N′-diacetylchitobiase of *Vibrio harveyi*. J. Biol. Chem. **264:**14778–14783.
40. **Stahl, D. A., B. Flesher, H. Mansfield, and L. Montgomery.** 1988. Use of phylogenetically based hybridization probes for studies of ruminal microbial ecology. Appl. Environ. Microbiol. **54:**1079–1084.
41. **Stewart, C. S., and M. P. Bryant.** 1988. The rumen bacteria, p. 21–75. *In* P. N. Hobson (ed.), The rumen microbial ecosystem. Elsevier Applied Science Publisher Ltd., New York.
42. **Teather, R. M., and J. D. Erfle.** 1990. DNA sequence of a *Fibrobacter succinogenes* mixed-linkage β-glucanase (1,3-1,4-β-D-glucan 4-glucanohydrolase) gene. J. Bacteriol. **172:**3837–3841.
43. **Warren, R. A. J., C. F. Beck, N. R. Gilkes, D. G. Killburn, M. L. Langsford, R. C. Miller, Jr., G. P. O'Neill, M. Scheufens, and W. K. R. Wong.** 1986. Sequence conservation and region shuffling in an endoglucanase and an exoglucanase from *Cellulomonas fimi*. Proteins **1:**335–341.
44. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51:**221–271.
45. **Wu, H. C., and M. Tokunaga.** 1986. Biogenesis of lipoproteins in bacteria. Curr. Top. Microbiol. Immunol. **125:**127–158.