0273-2289/03/105-108/0775/$20.00

CelF of *Orpinomyces* PC-2 Has an Intron and Encodes a Cellulase (CelF) Containing a Carbohydrate-Binding Module

Huizhong Chen,[1] Xin-Liang Li,[*,2] David L. Blum,[3] Eduardo A. Ximenes,[4] and Lars G. Ljungdahl[1]

[1]Center for Biological Resource Recovery, Department of Biochemistry and Molecular Biology, Life Science Building, The University of Georgia, Athens, GA 30602-7229;
[2]Fermentation Biotechnology Research Unit, NCAUR, USDA[†], ARS, 1815 N. University Street, Peoria, IL 61604, E-mail: lix@ncaur.usda.gov;
[3]Diversa Corporation, 4955 Directors Place, San Diego, CA 92121-1609; and
[4]Laboratorio de Enzimlogia, Departamento de Biologia Celular, Universidade de Brasilia, Asa Norte, Brasilia-DF, Brazil 70910-900

Abstract

A cDNA, designated *celF*, encoding a cellulase (CelF) was isolated from the anaerobic fungus *Orpinomyces* PC-2. The open reading frame contains regions coding for a signal peptide, a carbohydrate-binding module (CBM), a linker, and a catalytic domain. The catalytic domain was homologous to those of CelA and CelC of the same fungus and to that of the *Neocallimastix patriciarum* CELA, but CelF lacks a docking domain, characteristic for enzymes of cellulosomes. It was also homologous to the cellobiohydrolase IIs and endoglucanases of aerobic organisms. The gene has a 111-bp intron, located within the CBM-coding region. Some biochemical properties of the purified recombinant enzyme are described.

Index Entries: Cellulase; anaerobic fungi; *Orpinomyces*; cellulose-binding module; carbohydrate-binding module.

[†]Names are necessary to report factually on available data; however, the USDA neither guarantees nor warrants the standard of the product, and the use of the name by USDA implies no approval of the product to the exclusion of others that may also be suitable.

*Author to whom all correspondence and reprint requests should be addressed.

Introduction

Anaerobic fungi are part of the natural microflora of the alimentary tract of many herbivorous animals. Since the discovery of the first anaerobic fungus, *Neocallimastix frontalis*, by Orpin in 1975 from the rumen of sheep *(1)*, at least 17 different species of anaerobic fungi have been isolated from ruminant and nonruminant herbivores. Anaerobic fungi produce highly active hydrolytic enzymes *(2)*, such as endoglucanases, xylanases, lichenase, and esterases. The fungi physically associate with the lignocellulosic tissue of plant fragments. The hydrolytic enzymes exist as high molecular mass complexes and as individual molecules *(3,4)*. Several genes coding for the hydrolytic enzymes have been cloned and sequenced from the monocentric fungi *N. patriciarum (5–8)*, *N. frontalis (9)*, and *Piromyces* sp. *(10,11)* and from the polycentric fungi *Orpinomyces* PC-2 *(12–15)* and *Orpinomyces joyonii (16)*. Analyses of the primary structures of endoglucanases, xylanases, and lichenase from the anaerobic fungi have revealed that there are substantial sequence similarities between them from the rumen anaerobic fungi and from bacteria of the rumen, which suggests that these genes may have bacterial origin *(5,7,8,12–14)*. However, recently several family 6 cellulase genes have been isolated from *N. patriciarum (17)* and *Orpinomyces* PC-2 *(15)*, and their amino acid sequences are homologous to enzymes of aerobic fungi. An enolase gene from *N. frontalis* containing an intron *(18)* and a cyclophilin gene from *Orpinomyces* PC-2 that is heavily interrupted by introns *(3,19)* have been reported. So far no intron has been found in any of the sequenced genes coding for plant cell wall–degrading enzymes of the anaerobic fungi. Here we report the cloning and sequencing of a cDNA (*celF*) encoding a family 6 cellulase (CelF) from *Orpinomyces* PC-2 and characterization of the cellulase obtained by expression in *Escherichia coli*. Analysis of genomic DNA of the gene *celF* revealed that there is an intron, indicating that *celF* has a fungal origin. This appears as the first report for an intron-containing gene coding for an enzyme from an anaerobic fungus involved in plant cell wall degradation.

Materials and Methods

Strains, Vectors, Construction, and Screening of an Orpinomyces *cDNA Library*

Cultivation of the anaerobic fungus *Orpinomyces* PC-2, extraction and purification of mRNA, cDNA library construction, screening for cellulase clones, in vivo excision, and sequence determination and analysis have been described previously *(12,14)*. Cellulase-producing plaques were identified by incorporating 0.2% (w/v) remazol brilliant blue–carboxymethylcellulose (CMC) into top agar *(15)*.

Enzyme Purification and N-Terminal Sequencing

A single colony of *E. coli* XL-1 Blue harboring pCEL8 grown on Luria-Bertani (LB)-ampicillin medium was inoculated into a flask containing

500 mL of LB-ampicillin (50 µg/mL) broth. The culture was shaken (280 rpm) at 37°C and grown to an OD_{600} of approx 1.0. Isopropyl-1-thio-β-D-galactopyranoside (1 m*M*) was added, and the culture was incubated for another 4 h. Cells were harvested by centrifugation (5000*g*, 10 min), washed with 50 mL of buffer containing 25 m*M* Tris-HCl (pH 7.0), and resuspended in 30 mL of the same buffer. The cells were then disrupted by sonication, and cell debris was removed by centrifugation. The recombinant enzyme was purified using Q Sepharose, Mono S HR 10/10, and Superdex 75 10/30 columns on a fast performance liquid chromatography (FPLC) system. Purified CelF was subjected to N-terminal amino acid sequencing using an Applied Biosystems model 477A gas-phase sequencer equipped with an automatic on-line phenylthiohydantoin analyzer.

Enzyme Characterization

All enzyme assays were carried out in duplicate in 50 m*M* sodium citrate buffer at pH 6.0 and 40°C unless otherwise stated. Carboxymethylcellulase activity was assayed by mixing a 0.2-mL aliquot of appropriately diluted enzyme with 0.4 mL of buffer containing 1% (w/v) CMC (low viscosity; Sigma, St. Louis, MO). The reaction time was 30 min and was terminated by the addition 1.2 mL of 31 m*M* dinitrosalicylic acid reagent *(20)*. Reducing sugars were measured by reading the absorbance at 550 nm. Glucose was used as the standard. Hydrolysis of other polysaccharides was tested similar to that for CMC. Activity measurements of enzyme preparations toward *p*-nitrophenol (*p*NP)–linked substrates were performed in 0.3 mL of buffer containing 2 m*M* substrates. Reactions were terminated after 10 min by the addition of 0.8 mL of 1 *M* Na_2CO_3. The release of *p*NP was measured by reading the absorbance at 405 nm. *p*NP was used as the standard. One unit of enzyme activity was defined as the amount of enzyme required to release 1 µmol of glucose equivalent/min. Protein was quantified by the Bradford *(21)* method using bovine serum albumin as the standard.

The effect of pH on the activity and the pH optimum of the cellulase were determined at 40°C using the following buffers: 0.1 *M* sodium acetate (pH 4.2–5.4), 0.1 *M* sodium phosphate (pH 5.8–7.8), and 0.1 *M* Tris-HCl (pH 8.2). The effect of temperature on the cellulase was determined by assaying the enzyme at temperatures from 20 to 65°C. Thermostability was measured by incubating enzyme samples in 50 m*M* sodium citrate buffer, pH 6.0, for up to 24 h at temperatures from 40 to 65°C. The enzyme solution was chilled in an ice bath for 5 min and then assayed using the standard assay at 40°C. In all these assays, CMC was used as the substrate. For determination of K_m and V_{max} values, enzyme assays were performed at pH 6.0 and 40°C with CMC as the substrate at concentrations from 1.0 to 25 mg/mL. K_m and V_{max} values were obtained from Lineweaver-Burk plots. Zymogram analysis was done with lichenin (0.2% [w/v]) as the substrate according to Chen et al. *(12)*.

Sugars released from cellooligosaccharides and polysaccharides were analyzed with a Hewlett-Packard 1100 series high-performance liquid

chromatograph equipped with an autoinjector and a 1047A refractive index detector using a Bio-Rad Aminex HPX-42A carbohydrate column. Water was used as the mobile phase at a flow rate of 0.6 mL/min, and the column temperature was set at 80°C. Glucose, cellobiose, cellotriose, cellotetraose, and cellopentaose were used as the standards.

For assays of cellulose-binding capability, enzyme samples of 50 µg of protein were mixed with 30 mg of Avicel in a final volume of 1 mL of buffer (50 m*M* sodium citrate buffer, pH 6.0) at 4°C. After 1 h with continuous shaking, the Avicel was removed by centrifugation (15,000*g* at 4°C). Residual cellulase in the supernatant was assayed as already described.

Analysis of Genomic DNA

Oligonucleotides 5'ATGAAAATTTTACTTTTTGCCAG3' and 5'TTAGAATCCTGGTCTAGCATTTC3' corresponding to opposite strands of the end regions of *celF* (GenBank accession no. U97154) were used as primers with genomic DNA *(12)* and the cDNA libraries as templates for polymerase chain reactions (PCRs). Amplification was for 30 cycles with each cycle consisting of 90 s of melting at 95°C, 60 s of annealing at 55°C, and 90 s of extension at 72°C. Amplified DNA was separated on an agarose gel, and DNA bands were visualized by ethidium bromide staining. The PCR products amplified from the genomic DNA were cloned into PCRII vector (Invitrogen) and sequenced as described previously *(12,14)*.

Results and Discussion

Isolation and Sequencing of Orpinomyces CelF

Screening of the cDNA library constructed in λZAPII yielded 19 cellulase-producing plaques when 2×10^5 plaque-forming units were plated. The positive plaques were further enriched and purified. PCR, restriction enzyme digestion, and sequencing analyses revealed that 16 of these plaques represented cDNAs of *celA*, *celB*, and *celE*, which were reported previously *(13–15)*. The other three plaques represented three different new cellulase cDNAs (pCEL2, pCEL5, and pCEL8).

Sequencing of the inserts in the plasmids obtained from the plaques by in vivo excision revealed that pCEL8 contained a 1520-bp cDNA (*celF*) with a complete open reading frame (ORF) encoding a polypeptide (CelF) of 432 amino acids with a calculated mass of 46,736 Daltons. The detailed sequence data are deposited to GenBank and can be accessed as U97154. The putative translation start codon (ATG) for *celF* was assigned based on the fact that there were stop codons in all three frames preceding the ORF; there was no ATG codon upstream of the proposed ORF. After the ORF, a 3' untranslated AT-rich end of 127 bp was observed, but no typical long poly(A) stretch was found. The G+C content of the ORF of *celF* was 36.4% and that of the 5' and 3' noncoding regions was very low (11.8%). High A+T contents have also been found in other cDNAs from anaerobic fungi

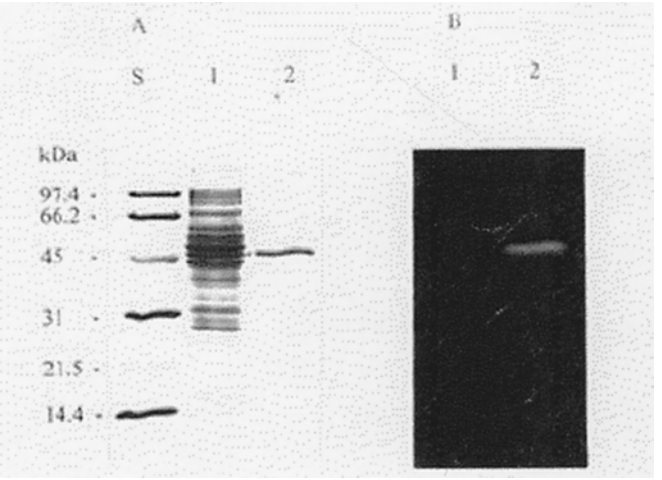


Fig. 1. SDS-PAGE and zymogram of the recombinant CelF. **(A)** Coomassie brilliant blue staining of SDS-PAG; **(B)** cellulase zymogram gel. Lane S, protein molecular mass standards; lane 1, crude *E. coli* proteins (50 μg); lane 2, the purified CelF (2 μg).

(12–15). The codon usage for *celF* is similar to that of other *Orpinomyces* PC-2 cellulase, xylanase, and lichenase genes.

Purification and N-Terminal Processing of Recombinant Enzyme from E. coli

Several steps (*see* Materials and Methods) were required to purify the recombinant enzyme to homogeneity. Starting from 650 mg of crude proteins, we were able to obtain 2.5 mg of purified CelF with 22-fold purification. Specific activity toward CMC of the purified enzyme was 120 U/mg. Based on the specific activity of the purified cellulase, the recombinant CelF protein constituted about 4.6% of the total *E. coli* proteins.

The first 10 N-terminal amino acid residues of the recombinant enzyme were AXGGAYAQXG. Except the two unidentified residues, the sequence matched the region corresponding to amino acid residues 21–30 of the deduced protein sequence, suggesting that the enzyme is processed in *E. coli* with a 20 amino acid signal sequence being removed to give a mature active enzyme of 412 amino acid residues.

Enzyme Characterization

The purified recombinant CelF appeared as a single band on sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 1). Zymogram analysis showed that the apparent molecular mass of CelF produced in *E. coli* was approx 45.5 kDa (Fig. 1), which appears to be consistent with the deduced molecular mass of the mature CelF lacking the proposed signal peptide of 20 amino acid residues. No other activity band with lower molecular mass was detected in the crude cell extracts,

indicating that the Asn-rich linker region was relatively stable. This is in contrast to several other anaerobic fungal hydrolytic enzymes containing the noncatalytic repeated peptide domains where the linker regions between the catalytic domains and the noncatalytic repeated peptide domains are susceptible to truncation *(7,13,14)*.

The activity of the cellulase was determined from pH 4.2 to 8.6 using CMC as the substrate. A typical pH profile was obtained with an optimum between pH 5.8 and 6.2, which was similar to values of most other hydrolytic enzymes from the anaerobic fungi *(12,15)* and was close to the physiologic pH in the rumen. The effect of temperature on the cellulase activity was determined in 50 m*M* sodium citrate at pH 6.0 from 20 to 65°C. Maximum activity was observed between 40 and 50°C, and the activity decreased rapidly above 55°C. Thermal stability was investigated by incubating the enzyme, up to 24 h, at different temperatures. Almost no activity loss was observed between 40 and 45°C in the above mentioned buffer at pH 6.0 for 24 h, and 61 and 12.9% of the enzyme activity were retained after 24 h of incubation at 50 and 55°C, respectively. Inactivation occurred at 60°C with only 15.5% of the enzyme activity remaining after 1 h of incubation. K_m and V_{max} values using CMC as the substrate at 40°C and pH 6.0 obtained from Lineweaver-Burk plots were 4.37 mg/mL and 173 U/mg of protein, respectively.

Activities of the purified CelF against various substrates are given in Table 1. The enzyme rapidly hydrolyzed acid-swollen cellulose, CMC, barley β-glucan, and lichenin. The CelF had a low but detectable activity toward microcrystalline cellulose (Avicel) with a specific activity of 0.039 U/mg of protein. No detectable hydrolysis was observed of *p*NP-β-D-glucoside, oat spelt xylan, or *p*NP-β-D-xyloside.

Hydrolysis products formed during the action of the purified recombinant CelF on cellooligosaccharides, Avicel, acid-swollen cellulose, and CMC were determined by hifh-performance liquid chromatography (HPLC) (Table 2). Cellobiose was not hydrolyzed, whereas cellotriose was slowly hydrolyzed to cellobiose and glucose. With cellotetraose as the substrate, cellobiose was the sole product detected. Thus, the second glucosidic linkage was mainly cleaved by the enzyme. Cellopentaose was largely converted into cellobiose, cellotriose, and some glucose, indicating that some cellotriose was further hydrolyzed to yield cellobiose and glucose. The hydrolysis patterns of cellooligosaccharides by CelF are somewhat similar to those by cellobiohydrolase II (CBHII) from *Trichoderma reesei (22)*. The end products of hydrolysis of Avicel, acid-swollen cellulose, and CMC yielded mainly cellobiose, cellotriose, and minor amounts of glucose. It is apparent that CelF had both endoglucanase and CBH activities, which is similar to that for CelA and CelC from the same fungus *(15)*.

Domain Structures of CelF and Its Homology with Other Cellulases

Analysis of the CelF sequence revealed that it contains a typical signal peptide sequence consisting of about 20 amino acid residues *(23)*.

Table 1
Substrate Specificity of *Orpinomyces* CelF Produced in *E. coli*[a]

Substrate[b]	Specific activity (U/mg protein)
Avicel	0.039
Acid-swollen cellulose	83.1
CMC	120.0
Barley β-glucan	1001.5
Lichenin	724.6

[a]Assays were performed at 40°C and pH 6.0 (50 m*M* sodium citrate) for 10 min (*p*NP-β-glycosides), 15 min (oat spelt xylan, barley β-glucan, and lichenin), 30 min (CMC), and 4 h (Avicel), respectively.
[b]The activities with *p*NP-β-D-glucoside, *p*NP-β-D-cellobioside, oat spelt xylan, or *p*NP-β-D-xyloside as substrate were < 1.0% of that against CMC.

Table 2
HPLC Analysis of Products of Cellooligosaccharide and Polysaccharide Hydrolysis by CelF[a]

	Products or residual substrates (μmol/mL)			
Substrate	G1	G2	G3	G4
Cellotriose (G3)	0.30	0.20	2.7	—
Cellotetraose (G4)	—	5.8	—	—
Cellopentaose	0.20	3.3	2.7	—
Avicel	—	0.20	0.11	—
Acid-swollen cellulose	0.037	0.72	0.34	—
CMC	0.017	0.31	0.12	—

[a]Reaction mixtures contained 0.25 U/mL of enzyme (CMC as substrate) and 3 m*M* cellooligosaccharides, 0.7% (w/v) CMC, or 0.5% (w/v) acid-swollen cellulose in 20 m*M* sodium citrate. Reactions were at pH 6.0 and 40°C for 4 h. In the case of Avicel, 1.0 U/mL of enzyme (CMC as substrate) and 1% (w/v) Avicel were combined and reactions were at pH 6.0 and 40°C for 16 h.

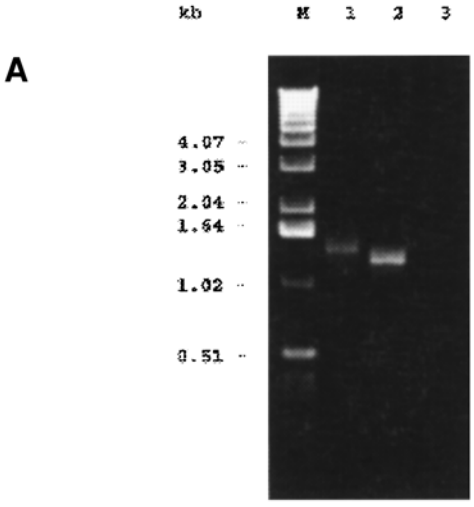
Immediately after the signal sequence, amino acid residues 22–57 constitute a typical fungal carbohydrate-binding module (CBM) *(17,24)*. The catalytic domain is located at the C-terminus (amino acid residues 106–432). It is separated from the CBM by an extremely Asn-rich linker (amino acid residues 67–105).According to classification of CBMs *(24)*, the CBM of CelF should be placed in family 1, which is exclusive to fungal hydrolases. It has a high degree of similarity with the CBM of CBHII from *T. reesei* (86% similarity and 53% identity). Six cysteine residues (nos. 22, 29, 39, 40, 46, and 56) forming three disulfide bridges that stabilize the polypeptide are conserved in CelF *(25,26)*. Three highly conserved aromatic residues (Tyr26, Trp52, and Tyr53), which are conspicuous building blocks of the flat-face binding to the cellulose surface *(23)*, are present

in the CBM. Moreover, three invariant amino acids (Gln28, Asn50, and Gln55) in suitable positions for hydrogen bonding with the cellulose surface are also present in the CBM *(23)*. The presence of a CBM in CelF is consistent with the fact that about 80% of the cellulase activity of the purified CelF adsorbed onto Avicel.

The deduced amino acid sequence of CelF of *Orpinomyces* PC-2 when compared with protein sequences in the SWISS PROT and GP data banks was homologous to several CBHIIs and endoglucanases belonging to family 6 glycosyl hydrolases *(27)*. The highest identity was with CELA of *N. patriciarum (17)*. The identity with it was 82.9% when the complete sequences including signal peptide, CBM, linker, and catalytic domain were compared. One deletion and/or insertion between these two enzymes was found in the linker region (after residue 86 in CelF), whereas five amino acids (residues 345–349) present at the carboxyl terminus of *Orpinomyces* PC-2 CelF were not found in *N. patriciarum* CELA. The catalytic domain of CelF of *Orpinomyces* PC-2 is homologous with the catalytic domains of CelA (75.2% similarity, 64.8% identity) and CelC (74.9% similarity, 63.6% identity) of the same organism *(15)*. It also has substantial similarity with the CBHIIs of *T. reesei* (52.6% similarity, 37.2% identity) *(26)* and *Fusarium oxysporum* (54.7% similarity, 38.9% identity) *(28)*, *Cellulomonas fimi* CENA (49.2% similarity, 32.0% identity) *(29)*, *Streptomyces halstedii* CELA1 (51.6% similarity, 31.2% identity) *(30)*, and *Thermomonospora fusca* CELE2 (50.3% similarity, 26.5% identity) *(31)*.

The three-dimensional structures of the catalytic domains of *T. reesei* CBHII *(32)* and *T. fusca* CELE2 have been determined *(33)*. Mutagenesis studies of *T. reesei* CBHII have shown that Asp245 is the likely proton donor in the catalytic event and that the neighboring Asp199 is charged, ensuring the protonation of Asp245. A function of the Tyr193 is to modulate the protonation states of the interacting carboxylates of Asp199 and 245 *(34)*. These three amino acid residues are conserved in family 6 glycosyl hydrolases and are present in the catalytic domain of CelF (Asp181, Asp223, and Tyr175), CelA, and CelC *(15)*. Family 6 glycosyl hydrolases contain both CBHIIs and endoglucanases. Endo- or exo-type activity depends on whether the enzyme contains loops that form a tunnel active site *(32,34)* or an open cleft (endoglucanase) *(33)*. Compared with *T. reesei* CBHII exoglucanase, there are deletions of two amino acids corresponding to the carboxyl terminal loop (from Ser391 to Gly409) and four amino acids corresponding to the second loop (from Pro178 to Ala191) for CelF of *Orpinomyces* PC-2, which suggests that the loops might only partially enclose the tunnel of the active site *(15)*. This indicates that the loops of CelF, like those of CelA and CelC, are distinct from those of CBHIIs and endoglucanases in the same family, which may result in CelF having both CBHII and endoglucanase activities *(15)*.

Although the catalytic domain of CelF is very similar to those of CelA and CelC, CelF has a CBM, whereas CelA and CelC contain a dockerin, which is involved in neither catalysis nor cellulose binding *(5,7,15)*. It has



A

B

GTATGGATATTTTATTTTTATAATTTAGGAAATAAATACTT
TTAAATAATTTAATTTAAGTATTGATTATTTTAAATTATTA
TATTACTTATAACAAATTAAGATATATAG

Fig. 2. Analysis of genomic sequence of *celF*. **(A)** Amplification of genomic DNA of *Orpinomyces* PC-2 *celF* coding region by PCR. Reaction solutions (20 µL) with genomic DNA (lane 1) and cDNA (lane 2) as the templates and without template (lane 3) were run on 1.5% agarose gel. DNA molecular standards were used in lane M. After electrophoresis, DNA bands were visualized by ethidium bromide staining. **(B)** Nucleotide sequence of an intron found in *Orpinomyces celF.*

been suggested that dockerins function as docking domains in a fashion similar to that of the dockerin domains of catalytic subunits of the cellulosome of *Clostridium thermocellum (4,10,35,36)*. The lack of a dockerin in CelF suggests that enzyme is not a part of cellulase/hemicellulase complexes found in *Orpinomyces* PC-2 *(37)* and other anaerobic fungi *(38)*. The presence of cellulases as both free and complex forms that are highly catalytically similar may provide the fungus with an advantageous strategy for the rapid decomposition of plant cell wall structures.

Coding Region of CelF *Contains an Intron*

The ORF region of *celF* was amplified by PCR, and its size is larger than that amplified from cDNA (Fig. 2A). Sequencing the DNA amplified from the genomic DNA template revealed that it produced a 1410-bp DNA fragment. Alignment of the sequences of genomic and cDNA of *celF* revealed an intron of 111 bp (Fig. 2B) located in the CBM-coding region. Splicing boundaries started with GT and ended with TA, which match the general consensus sequences found for introns in filamentous fungi *(39)*.

In a similar experiment, the ORF regions of *celA* and *celC* genomic DNA were amplified by PCR, and their sizes were the same as those amplified from their cDNAs (data not shown), indicating that *celA* and *celC* are devoid of introns. Other genes coding for cellulolytic enzymes of anaerobic fungi, which have been examined, are devoid of introns *(8,12–14,16)*. The fact that the ORF of *celF* contains an intron indicates that it has a eukaryotic origin. Several other genes isolated from anaerobic fungi are believed to have prokaryotic origins *(3)*.

Using the software package PHYLIP *(40)*, a phylogenetic tree based on cellulase sequences was created from neighbor-joining bootstrap analysis. It showed that CelF, CelA, CelC, and *N. patriciarum* CELA were clustered together. It has been suggested that a common ancestral precursor of cellulolytic aerobic fungi and rumen anaerobic fungi may have existed *(17)*. Our finding of an intron in *celF* provides solid evidence that the gene has a eukaroytic origin. Thus, two types of cellulase genes have been discovered in anaerobic fungi. One type includes those coding for endoglucanases possibly transferred from rumen bacteria *(7,13,14,16)*, and the other is those of *Orpinomyces* CelA, CelC, CelF and *Neocallimastix* CELA, which appear to have fungal origin (*[15,17]* this study). Moreover, current observations seem to support that *celF* CBM-coding region was replaced with those coding for a dockerin-coding sequence during evolution, resulting in the formation of *celA* and *celC*. Of course, the opposite could also be a possibility. In any event, the high sequence similarity among *celA*, *celC*, *celF*, and *Neocallimastix celA* strongly suggests the occurrence of gene transfer and duplication among the fungi.

Acknowledgments

We thank Dr. Mike Cotta for critical comments and Marsha Ebener for executive help. This work was supported by grants from the US Department of Energy (DE-FGO2-93ER2012) and from the Consortium for Plant Biotechnology Research, Georgia Research Alliance, and Aureozyme, Inc.

References

1. Orpin, G. C. (1975), *J. Gen. Microbiol.* **91,** 249–262.
2. Borneman, W. S., Akin, D. E., and Ljungdahl, L. G. (1989), *Appl. Environ. Microbiol.* **55,** 1066–1073.
3. Ljungdahl, L. G., Li, X.-L., and Chen, H. (1998), in *Thermophiles—The Keys to Molecular Evolution and the Origin of Life*, Wiegel, J. and Adams, M., eds., Taylor & Francis, London, UK, pp. 187–197.
4. Steenbakkers, P. J. M., Li, X.-L., Ximenes, E. A., Arts, J. G., Chen, H., Ljungdahl, L. G., and Op Den Camp, H. J. M. (2001), *J. Bacteriol.* **183,** 5325–5333.
5. Black, G. W., Hazlewood, G. P., Xue, G.-P., Orpin, C. G., and Gilbert, H. J. (1994), *Biochem. J.* **299,** 381–387.
6. Dalrymple, B. P., Cybinski, H. D., Layton, I., McSweeney, C. S., Xue, G.-P., Swadling, S. J., and Lowry, J. B. (1997), *Microbiology* **143,** 2605–2614.
7. Gilbert, H. J., Hazlewood, G. P., Laurie, J. I., Orpin, C. G., and Xue, G.-P. (1992), *Mol. Microbiol.* **6,** 2065–2072.

8. Zhou, L, Xue, G.-P., Orpin, C. G., Black, G. W., Gilbert, H. J., and Hazlewood, G. P. (1994), *Biochem. J.* **297,** 359–364.
9. Durand, R., Rascle, C., and Fèvre, M. (1996), *Curr. Genet.* **30,** 531–540.
10. Fanutti, C., Ponyi, T., Black, G. W., Hazlewood, G. P., and Gilbert, H. J. (1995), *J. Biol. Chem.* **270,** 29,314–29,322.
11. Millward-Sadler, S. J., Hall, J., Black, G. W., Hazlewood, G. P., and Gilbert, H. J. (1996), *FEMS Microbiol. Lett.* **141,** 183–188.
12. Chen, H., Li, X.-L., and Ljungdahl, L. G. (1997), *J. Bacteriol.* **179,** 6028–6034.
13. Chen, H., Li, X.-L., Blum, D. L., and Ljungdahl, L. G. (1998), *FEMS Microbiol. Lett.* **159,** 63–68.
14. Li, X.-L., Chen, H., and Ljungdahl, L. G. (1997), *Appl. Environ. Microbiol.* **63,** 628–635.
15. Li, X.-L., Chen, H., and Ljungdahl, L. G. (1997), *Appl. Environ. Microbiol.* **63,** 4721–4728.
16. Liu, J.-H., Selinger, L. B., Hu, Y.-J., Moloney, M. M., Cheng, K.-J., and Beauchemin, K. A. (1997), *Can. J. Microbiol.* **43,** 477–485.
17. Denman, S., Xue, G.-P., and Patel, B. (1996), *Appl. Environ. Microbiol.* **62,** 1889–1896.
18. Durand, R., Fischer, M., Rascle, C., and Fèvre, M. (1995), *Microbiology* **141,** 1301–1308.
19. Chen, H., Li, X.-L., and Ljungdahl, L. G. (1995), *Proc. Natl. Acad. Sci. USA* **92,** 2587–2591.
20. Miller, G. L. (1959), *Anal. Chem.* **31,** 426–428.
21. Bradford, M. M. (1976), *Anal. Biochem.* **72,** 248–254.
22. Harjunpää, V., Teleman, A., Koivula, A., Ruohonen, L., Teeri, T. T., Teleman, O., and Drakenberg, T. (1996), *Eur. J. Biochem.* **240,** 584–591.
23. Von Heijne, G. (1986), *Nucleic Acids Res.* **14,** 4683–4690.
24. Tomme, P., Warren, R. A. J., Miller, R. C. Jr., Kilburn, D. J., and Gilkes, N. R. (1995), in *Enzymatic Degradation of Insoluble Carbohydrates*, Saddler, J. N. and Penner, M. H., eds., ACS Symposium Series 618, American Chemical Society, Washington, DC.
25. Hoffrén, A.-M., Teeri, T. T, and Teleman, O. (1995), *Protein Eng.* **8,** 443–450.
26. Teeri, T. T., Lehtovaara, P., Kauppinen, S., Salovuori, I., and Knowles, J. (1987), *Gene* **51,** 43–52.
27. Henrissat, B. and Bairoch, A. (1993), *Biochem. J.* **293,** 781–788.
28. Sheppard, P. O., Grant, F. J., Oort, P. J., Sprecher, C. A., Foster, D. C., Hagen, F. S., Upshall, A., McKnight, G. L., and O'Hara, P. J. (1994), *Gene* **150,** 163–167.
29. Wong, W. K., Gerhard, D. C., Guo, Z. M., Kilburn, D. G., Warren, R. A. J., and Miller, R. C. Jr. (1986), *Gene* **44,** 315–324.
30. Fernández-Abalos, J. M., Sánchez, P., Coll, P. M., Villanueva, J. R., Peréz, P., and Santamaría, R. I. (1992), *J. Bacteriol.* **174,** 6368–6376.
31. Lao, G., Ghangas, G. S., Jung, E. D., and Wilson, D. B. (1991), *J. Bacteriol.* **173,** 3397–3407.
32. Rouvinen, J., Bergfors, T., Teeri, T. T., Knowles, J. K. C., and Jones, T. A. (1990), *Science* **249,** 380–386.
33. Spezio, M., Wilson, D. B., and Karplus, P. A. (1993), *Biochemistry* **32,** 9906–9916.
34. Koivula, A., Reinkainen, T., Rouhonen, L., Valkeajärvi, A., Claeyssens, M., Teleman, O., Kleywegt, G. J., Szardenings, M., Rouvinen, J., Jones, T. A., and Teeri, T. T. (1996), *Protein Eng.* **9,** 691–699.
35. Choi, S.-K. and Ljungdahl, L. G. (1996), *Biochemistry* **35,** 4906–4910.
36. Pages, S., Belaich, A., Belaich, J. P., Morag, E., Lamed, R., Shoham, Y., and Bayer, E. A. (1997), *Proteins* **29,** 517–527.
37. Li, X.-L., Chen, H., He, H., Blum, D. L., and Ljungdahl, L. G. (1997), in *Abstracts of the 97th General Meeting of the American Society for Microbiology*, American Society for Microbiology, Washington, DC, p. 424.
38. Dijkerman, R., Vervuren, M. B. F., Op Den Camp, H. J. M., and van der Drift, C. (1996), *Appl. Environ. Microbiol.* **62,** 20–25.
39. Gurr, S. J., Unkles, S. E., and Kinghorn, J. R. (1987), in *Gene Structure in Eukaryotic Microbes*, Kinghorn, J. R., ed., IRL, Oxford, England, UK, pp. 93–139.
40. Felsenstein, J. (1989), *Cladistics* **5,** 164–166.