International Journal of Biochemistry Research & Review
6(2): 62-74, 2015, Article no.IJBcRR.2015.038
ISSN: 2231-086X

SCIENCEDOMAIN *international*
www.sciencedomain.org

Clostridium thermocellum Cel5L – Cloning and Characterization of a New, Thermostable GH5 Cellulase

Phillip J. Brumm[1*], Spencer Hermanson[1], Krishne Gowda[2], Dan Xie[1], and David A. Mead[1,2]

[1]C5•6 Technologies, 2905 Parmenter St., Middleton, WI, United States of America.
[2]Lucigen Corporation, 2905 Parmenter St., Middleton, WI, United States of America.

Authors' contributions

This work was carried out in collaboration between all authors. Author PJB designed the study, wrote the protocol and supervised the work. Authors SH, KG and DX carried out all laboratories work and performed the statistical analysis. Author DAM managed the analyses of the study. All authors read and approved the final manuscript.

.

Article Information

DOI: 10.9734/IJBcRR/2015/15199
Editor(s):
(1) Rosario Gomez Garcia, Department of Biochemistry, Loyola University, USA.
Reviewers:
(1) Anonymous, India.
(2) Anonymous, India.
(3) Ummirul Mukminin Kahar, Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, Malaysia.
Complete Peer review History: http://www.sciencedomain.org/review-history.php?iid=847&id=3&aid=7625

Original Research Article

Received 12th November 2014
Accepted 19th December 2014
Published 5th January 2015

ABSTRACT

Aims: *Clostridium thermocellum* is a thermophilic, anaerobic bacterium that ferments cellulose and produces ethanol. While many of the predicted cellulases have been characterized from *C. thermocellum*, characterization of additional members is still needed to unravel the function of these enzymes.
Study Design: This report describes the first cloning, expression, characterization, and evaluation in cellulose degradation of *Clostridium thermocellum* Cel5L as well as its relationship to other cellulases.
Results: Cel5L, a 526 amino acid protein that shows low homology to other cellulases of *C. thermocellum*, possesses both *endo* and *exo* activity on β-glucan, carboxymethylcellulose, and cellooligosaccharides; the enzyme does not hydrolyze xylan, xyloglucan or glucomannan. When

**Corresponding author: E-mail: pbrumm@c56technologies.com;*

combined with exo-cellulases and β-glucosidase, Cel5L generates more glucose from cellulose than other tested *C. thermocellum* GH5 cellulases, which is surprising due to the lack of an annotated carbohydrate binding module. The Cel5L phylogenetic tree shows orthologs in both mesophilic and thermophilic cellulose degraders.

Conclusion: The presence of these orthologs in a variety of molecular constructs demonstrates the importance of Cel5L and its orthologs in cellulose degradation, both in *C. thermocellum* as well as other gram-positive cellulose degraders.

Keywords: C. thermocellum; cellulase Cel5L; Cthe_0405; endoglucanase; biomass.

1. INTRODUCTION

Cellulases are endoglucanases that degrade native (crystalline) cellulose either from the chain ends (*exo*-acting cellulases) or the interior of the chain (*endo*-acting cellulases). This definition is sometimes, but not always, used to distinguish cellulases from other endoglucanases such as *beta*-glucanases (which are active on *beta*-glucan but not native cellulose) [1]. Finding novel thermostable bacterial cellulases has been difficult because growth on cellulosic substrates and production of thermostable cellulases is rare among thermophilic microorganisms [2]. In the 1980s, initial work showed that *Clostridium thermocellum* (Cthe), a thermophilic anaerobic cellulose-degrading ethanologen [3,4], utilizes complex, high molecular weight structures termed the cellulosome [5,6] to rapidly degrade cellulose *in vivo*. Progress on understanding the structure and function of the cellulosome has been recently reviewed [1,7] and many questions on the exact mechanism of cellulose degradation by cellulosomal cellulases remains under study. Glycoside hydrolases (GH) are enzymes that catalyze the hydrolysis of the glycosidic linkage of sugar polymers, increasing the number of reducing ends in the products. Based on CAZy analysis [8-10], enzymes and proteins are divided into GH families, or groups of proteins related by sequence and predicted structure. Cellulases are only found within a limited number of these GH families, including GH5, GH6, GH7, GH8, GH9, GH12, GH45, GH48, GH51 and GH61 (discussed in detail in [1]). Carbohydrate binding modules are protein domains that can bind to carbohydrate chains, but which have no intrinsic enzymatic activity of their own. Only CBM modules from families CBM1 through CBM4 appear able to bind to cellulose. The Cthe genome has been sequenced multiple times, initially in 2007 (*Clostridium thermocellum* ATCC 27405™, complete genome, GenBank: CP000568) and other Cthe strains have also been sequenced recently [10-12]. Based on CAZy analysis of the genome sequence, Cthe ATCC 27405™is predicted to have genes for twenty-eight potential cellulases, of which twenty-five have dockerin domains, indicating a cellulosomal location. A significant number of these twenty-five cellulosomal cellulases have been characterized including CelA [13], CelD [14], CelE [15], CelF [16], CelJ [17], CelO [18], CelQ [19], CelR [20], and CelS [21]. Only two cellulases, Cel9I [22] and Cel48Y [23] are secreted as soluble enzymes and Cel5C [24] is intracellular. Of the twenty-five potential cellulosomal cellulases, eight have been annotated as being members of the glycosyl hydrolase family 5 (GH5) (Table 1) using the CAZy naming convention [8].

The Cthe genome codes for significantly more predicted cellulases than gram-positive organisms producing soluble cellulolytic enzymes, which typically have seven to twelve predicted cellulases [30]. Efficient cellulosomal function may require the larger number of potential cellulases and GH5 cellulases in *C. thermocellum*. It is unclear if this higher number of cellulases and GH5 family members is a result of gene duplication, lateral gene transfer, or some other mechanism. While seventeen of the predicted cellulases have been characterized from *C. thermocellum*, characterization of additional cellulases is still needed to unravel the genetics and function of these enzymes. The cloning, characterization, and bioinformatic analysis of *C. thermocellum* Cel5L is described here for the first time.

Table 1. *C. thermocellum* ATCC 27405™ GH5 family members

Gene number	GH family	CBM family	Name	Ref.
Cthe_0405	GH5		CelI5L	This work
Cthe_0536	GH5		CelB	[25]
Cthe_0797	GH5		CelE	[15,26]
Cthe_0821	GH5	CBM32		
Cthe_1472	GH5 GH26	CBM11	CelH	[27]
Cthe_2147	GH5	CBM3	CelO	[18]
Cthe_2193	GH5	CBM6 CBM13		
Cthe_2807	GH5		CelC	[25]
Cthe_2872	GH5		CelG	[28,29]

Legend: Cthe GH5 members identified using CAZy website [8]

2. MATERIALS AND METHODS

2.1 Materials

10G electro-competent *E. coli* cells, BL21(DE3), BL21(C43) chemically competent *E. coli* cells, and pEZSeq (a lac promoter vector) were obtained from Lucigen, Middleton, WI. *Clostridium thermocellum* CelE, CelG, CelH, CelI, CelK, CelO, and CAZyme β-Glucosidase 1 (*Alicyclobacillus acidocaldarius* gene AaLAA1DRAFT_0397) are produced and sold by C5-6 Technologies, Middleton, WI. pET28a vector was obtained from Merck Chemicals, San Diego, CA. Pure oligosaccharides, azurine cross-linked-labeled substrates, and D-Glucose (GOPOD Format) Assay Kits were obtained from Megazyme International (Wicklow, Ireland). 4-methylumbelliferyl-β-D-cellobioside (MUC), 4-methylumbelliferyl-β-D-xylopyranoside (MUX), and 4-methylumbelliferyl-β-D-glucoyranoside (MUG) were obtained from Research Products International Corp. (Mt. Prospect, IL). 4-methylumbelliferyl-β-D-lactoside (MUG) was obtained from Marker Gene Technologies, Inc. (Eugene, OR). Q Sepharose Fast Flow column media was purchased from GE Healthcare Life Sciences (Piscataway, NJ). Avicel® PH-101, pNP-β-glucoside, pNP-β-cellobioside and carboxymethyl cellulose (CMC) were purchased from Sigma-Aldrich (St. Louis, MO). Pierce Coomassie Protein Assay was obtained from Thermo Scientific (Fitchburg, WI). Mini-PROTEAN® polyacrylamide gels for electrophoresis were obtained from Bio-Rad (Hercules, CA). All other chemicals were of analytical grade. Frozen cell pellets of *Clostridium thermocellum* strain 27405™ cultures were a kind gift of Dr. Paul Weimer, USDA Agricultural Research Service, U.S. Dairy Forage Research Center, University of Wisconsin, Madison, WI and were used for library preparation. *Clostridium thermocellum* strain 27405™ is on deposit at the American Type Culture Collection, Manassas, VA.

YT plate media (16 g/l tryptone, 10 g/l yeast extract, 5 g/l NaCl, 30 mg/l kanomycin sulfate and 16 g/l agar) was used in all molecular biology screening experiments. Terrific Broth (12 g/l tryptone, 24 g/l yeast extract, 9.4 g/l K_2HPO_4, 2.2 g/l KH_2PO_4, 30 mg/l kanomycin sulfate and 4.0 g/l glycerol added after autoclaving) was used for protein expression in *E. coli* liquid cultures.

2.2 Library Construction, Screening and Cloning

A cell concentrate of *Clostridium thermocellum* was lysed using a combination of SDS and proteinase K, and genomic DNA was purified using phenol/chloroform extraction [31]. The genomic DNA was precipitated, treated with RNase to remove residual contaminating RNA, and fragmented by hydrodynamic shearing (HydroShear apparatus, GeneMachines, San Carlos, CA) to generate fragments of 2-4 kb. The fragments were purified on an agarose gel, end-repaired, and ligated into pEZSeq. To express putative cellulases, the *C. thermocellum* library was transformed into electrocompetent *E. coli*BL21(DE3) cells. Individual colonies were picked and grown in 96 well blocks. Aliquots of the cultures were collected by centrifugation and lysed using CelLytic IIB reagent (Sigma). The lysates were assayed for cellulase activity at 70ºC in 0.2 ml of 50 mM acetate buffer, pH 5.8 containing 0.2% AZCL-HE-Cellulose.

2.3 Enzyme Production and Purification

For enzyme production, cultures were grown from a single colony in 1000 ml of Terrific Broth at 37°C, 120 rpm. When the culture reached A_{600}=0.8, IPTG was added to a final concentration of 1 mM, and incubation was continued overnight. Cells were harvested by centrifugation, resuspended in Tris-HCl, pH 8.0, lysed by sonication and the centrifuged lysate was incubated at 70°C for 30 minutes. Precipitated proteins were removed by centrifugation, and Cel5L was further purified by anion exchange chromatography on Q Sepharose Fast Flow equilibrated with 50 mM Tris-HCl, pH 8.0 using a gradient of 0 to 250 mM NaCl in 50 mM Tris-HCl, pH 8.0.

2.4 Enzyme Assays

The *endo*-specificity of cellulases was determined in 0.50 ml of 50 mM acetate buffer, pH 5.8, containing 0.2% azurine cross-linked-labeled (AZCL) insoluble substrates and 20 µg of enzyme protein. Assays were performed at 70ºC, with shaking at 1400 rpm, for 60 minutes in a Thermomixer R (Eppendorf, Hamburg, Germany). Tubes were clarified by centrifugation and absorbance values determined using a Bio-Tek EL_x800 plate reader at 600 nm. The *exo*-specificity of cellulases was determined in 0.10 ml of 50 mM acetate buffer, pH 5.8, containing 1.0 µg of 4-methylumbelliferyl substrate and 10 µg of enzyme protein. Tubes were incubated at 70ºC, with shaking at 500 rpm, for 30 minutes in a Thermomixer R; after incubation, the tubes were examined using a hand-held UV lamp and compared to negative and positive controls.

Enzyme *endo*-cellulase activity was measured using a micro version of the Modified Somogyi Method for reducing sugars [32]. The reaction mixtures containing 200 µl of substrate (2%β-glucan or other carbohydrate in 50 mM acetate buffer, pH 5.8) and 5 µl enzyme sample were incubated at 70°C for 10 minutes [33]. Micromoles of sugars formed were determined using a glucose standard curve, and unit activity calculated as micromoles of reducing sugar per minute per milligram of protein at 70°C.

2.5 Effects of Temperature and pH on Enzyme Activity

The temperature optimum of Cel5L was determined using the reducing sugar assay, with 2% β-glucan as substrate at pH 5.8 between 30°C and 90°C. The pH optimum of Cel5L was determined using the reducing sugar assay with 2% β-glucan as substrate at 70°C between pH values of 3.0 and 11.0. All experiments were done in triplicate.

Glucose production from celloligosaccharides was determined using a coupled reaction. The reaction mixtures containing 200 µl of substrate in 50 mM acetate buffer, pH 5.8 and 5 µl enzyme sample were incubated at 70°C for 30 minutes with shaking at 1000 rpm. The reaction was terminated by incubation at 95°C for 10 minutes, and then the glucose formed was measured on aliquots of the reaction mixtures using the Megazyme D-Glucose (GOPOD Format) Assay Kit and expressed as nanomoles glucose formed in 30 minutes.

Cellulose hydrolysis experiments were conducted at 60°C in a final volume of 1.0 ml of 50 mM acetate buffer, pH 5.8, containing 5 mM $CaCl_2$ and either 2.0 mg Avicel microcrystalline cellulose or 3.4 mg of Whatman 1 filter paper and 100 micrograms of each purified cellulase. Thermostable *beta*-glucosidase (100 micrograms was added to all reactions to convert cellodextrin products to glucose. Glucose formed was determined using the Megazyme D-Glucose (GOPOD Format) Assay Kit according to the manufacturer's directions in at least triplicate.

Protein concentrations were measured using the Coomassie Plus reagent using bovine serum albumin as standard according to the manufacturer's instructions. Gel electrophoresis was performed using 4-20% acrylamide Mini-PROTEAN® gels according to the manufacturer's instructions and stained with Coomassie R-250.

2.6 Bioinformatics

The Gen Bank accession number of the sequence reported in this paper is Gene ID: 4808408. The complete *Clostridium thermocellum* strain 27405^T genome is available:http://www.ncbi.nlm.nih.gov/nuccore/NC_009012. Inter Pro Scan Family analysis available:(http://www.ebi.ac.uk/Tools/InterProScan/) was used to identify protein domains within Cel5L and retrieve sequences of *C. thermocellum* cellulases. BLASTp (Basic Local Alignment Search Tool [34]) available: (http://blast.ncbi.nlm.nih.gov/Blast.cgi) analysis tools were used to identify closest orthologs of Cel5L. Multiple sequence alignments were done

using Clustal Omega [35] available:http://www.ebi.ac.uk/Tools/msa/clustalo/help/. Glycosyl hydrolase predictions were obtained available from http://www.cazy.org/geno/acc_geno.html. Signal sequence predictions were determined using available:(http://www.cbs.dtu.dk/services/SignalP/)[36]. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model [37]. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. Evolutionary analyses were conducted in MEGA5 [38].

3. RESULTS AND DISCUSSION

3.1 Screening

During screening of the *Clostridium thermocellum* DNA library for cellulytic activity, an endoglucanase was discovered that hydrolyzed AZCL-barley-*beta*-glucan, AZCL-HE-cellulose, and 4-methylumbelliferyl-β-D-cellobioside, indicating an enzyme with both endo-cellulase and exo-cellulase activities. Sequencing of the ends of the DNA insert identified Cthe_0405, Cel5L, a putative 526 amino acid, and 59.9 kDa protein, having the predicted sequence shown in Fig. 1. The insert did not contain the full size gene, beginning at amino acid residue 35 in the predicted sequence. Sequence analysis predicted that the enzyme was a member of the glycoside hydrolase family 5, subfamily 1 (GH5_1) [39,40]. The gene encodes a signal sequence and two dockerin domains, indicating it is secreted from the cell and targeted to the cellulosome. Cel5L possesses no carbohydrate binding modules (CBM), suggesting a limited ability to degrade crystalline substrates.

3.2 Cel5L Cloning, Purification and Characterization

Using 50 nanograms of genomic DNA, the Cel5L cellulase gene was amplified using the following primers (*C. thermocellum* strain 27405^{T}corresponding genomic sequence):

Cel5L F: TACGGTAATCAGGTCTGGCTGACCGGCT-3' (506023-506050)
Cel5L R: TTACCCCTTCTCTATTTCGAAAGAATGCGA-3' (507244-507270)

The N-terminal primer was designed to begin at the consensus signal-sequence cleavage site, between amino acids 32 and 33 and have a stop codon at the end. The amplicon was ligated into pET28a vector and transformed into *E. coli* BL21 (DE3) cells.

Cloned Cel5L was prepared as described in Materials and Methods. The final enzyme purity was greater than 95% for Cel5L (Fig. 2), as well as for all other enzymes used in the comparisons.

```
        10         20         30         40         50         60
MRKVKALLLG LIVLAVALLP TVSFKSPTVA ADPNNDDWLH VEGNKIVDMY GNQVWLTGCN
        70         80         90        100        110        120
WFGFNTGTNV FDGVWSCNMR EALKGMADRG INFLRIPIST ELLYQWSQGI YPKANVNDFV
       130        140        150        160        170        180
NPELKGKNSL ELFDFAVQCC KEFGIKIMVD IHSPATDAMG HMYPLWYDGQ FTTEIWISTL
       190        200        210        220        230        240
EWLTERYKND DTILALDLKN EPHGTPGSEL MAKWDGSTDL NNWKHAAETC AKRILAINPN
       250        260        270        280        290        300
ILIVVEGVEV YPKPGYDYTA VDEWGKESKY FYNWWGGNLR GVRDYPIDLG KHQKQLVYSP
       310        320        330        340        350        360
HDYGPLVHKQ PWFYEGFNKE TLYNDCWRDN WAYIHEENIA PLIVGEWGGF MDRGDNEKWM
       370        380        390        400        410        420
KALRDYMIEN KISHTFWCYN ANSGDTGGLV YYDFITWDEE KYALLKPALW QTEDGKFIGL
       430        440        450        460        470        480
DHQIPLGSNG ITVTEYYGGY IPEPSPTATV PDVPTPSHSF EIEKGDVNGD GNVNSTDVVW
       490        500        510        520
LRRFLLKLVE DFPVPSGKQA ADMNDDGNIN STDMIALKRK VLKIPI
```

Fig. 1. Predicted Sequence of Cthe_0405, Cel5L

Legend: Predicted protein sequence of Cel5L. Signal sequence (italics) was determined using SignalP software, while Glycosyl Hydrolase family 5 domain (normal) and two dockerin type 1 repeats (underlined) were determined using InterProScan Software as described in Materials and Methods. The start of the truncated Cel5L identified in the clone library is shown in bold

The isolation and enzymatic properties of *C. thermocellum* GH5 cellulases have been described including CelG [28], CelE [41], CelH [27], and CelO [18]. However, the enzymes have never been compared to each other under identical assay conditions. To understand the relationship of the newly identified Cel5L to these enzymes, direct comparisons of activity were undertaken with the purified enzymes. The purified Cel5L had a broad temperature optimum between 70°C and 80°C, and a broad pH optimum between pH 4.0 and 8.0. These values are similar to those observed with *C. thermocellum* CelG, CelO, CelH and CelE when assayed under identical conditions (data not shown).

The enzymatic properties of Cel5L were compared to the properties of CelE, CelG, CelH, and CelO. Enzyme specificity for *endo*-activity was determined using azurine cross-linked-labeled (AZCL) insoluble substrates. The results (Table 2) show all five GH5 cellulases possess *endo*-activity, hydrolyzing AZCL-hydroxyethyl cellulose and AZCL-β-glucan. CelE and CelH also hydrolyze AZCL- arabinoxylan, AZCL-glucomannan, and AZCL-xyloglucan, indicating the ability of CelE and CelH to accept a wide range of polysaccharides in their active site. Cel5L, CelG and CelO have no activity on these three substrates, only linear glucan substrates are accessible to the active sites of these three enzymes. The five cellulases all hydrolyzed β-glucan, with specific activities ranging from 19 to 80 u/mg (Table 2). Cel5L and CelG had similar specific activities, 80 and 75 u/mg respectively. The remaining three enzymes had significantly lower activities, ranging from 19 to 47 u/mg.

Enzyme specificity for *exo*-activitywas determined using 4-methylumbelliferyl-β-D-linked substrates, cellooligosaccharides, and reduced cellooligosaccharides. All five GH5cellulases also possess *exo*-activity, hydrolyzing both 4-methylumbelliferyl-β-D-cellobioside and 4-methylumbelliferyl-β-D-lactoside. All five cellulases have no activity on either 4-methylumbelliferyl-β-D-xylopyranoside or 4-methylumbelliferyl-β-D- glucopyranoside (Table 2). Production of glucose from cellooligosaccharides and reduced cellooligosaccharides was used to determine if Cel5L and the other cellulases were active on the reducing or non-reducing end of cellooligosaccharides. The five enzymes show production of glucose from cellotriose, but not cellotriitol, indicating all five cellulases attack at the non-reducing end of the cellooligosaccharides (Table 3). Cel5L and CelG showed significantly lower production of glucose from cellotetraose than from cellotriose or cellopentaose, suggesting a release of cellobiose from the non-reducing end of the cellooligosaccharides. CelO showed very low activity on any of the cellooligosaccharides or their reduced versions. CelE and CelH showed similar or higher production of glucose from cellotetraose than from cellotriose or cellopentaose, suggesting a release of both cellotriose and cellobiose from cellooligosaccharides.

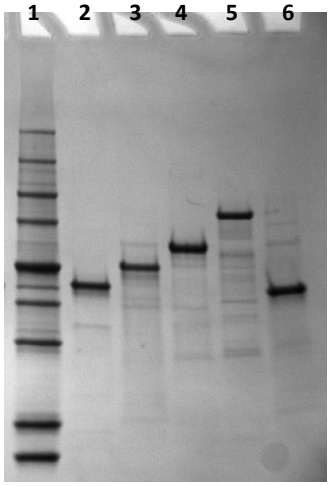

Fig. 2. SDS PAGE of GH5 cellulases used in this work

Legend: SDS PAGE on 4-20% gel as described in Materials and Methods. Lane 1, Promega Broad Range MW Markers; Lane 2, 1 μg Cel5L, Lane 3, 1 μg CelG; Lane 4, 1 μg CelO, Lane 5, 1 μg CelH; Lane 6, 1 μg CelE

The relative activity of the five *C. thermocellum* GH5 cellulases in cellulose conversion was determined in the absence of cellulosomal structures using two cellulosic distinct substrates, Whatman 1 filter paper and Avicel microcrystalline cellulose. The reactions were supplemented with *C. thermocellum* CelI and CelK to provide *exo*-cellulase activity and an excess of thermostable bacterial β-glucosidase to convert cellobiose and celloligosaccharides to glucose. The control set of enzymes, containing β-glucosidase, CelI and CelK converted 27% of the filter paper to glucose after 117 hr (Fig. 3). Addition of *endo*-cellulases would be expected to increase the rate of this cellulose conversion by these *exo*-cellulases [42]. Addition of CBM

domains to cellulases often results in improved cellulase performance [43,44]. To determine if *C. thermocellum* GH5 *endo*-cellulases with attached CBM domains performed better in cellulose hydrolysis than cellulases without CBM domains, we evaluated Cel5L, CelG, and CelE without CBM domains as well as CelH and CelO which possess their native CBM domains. CelH and CelO did not perform better than CelE, CelG, or Cel5L. Addition of CelO to the control set did not result in a significant improvement in conversion. Addition of CelE, CelH, or CelG increased the conversion to 34%, 37%, and 39% respectively. Supplementation with Cel5L produced significantly more glucose from filter paper (45%) than supplementation with CelG, CelE, CelH or CelO. This suggests that Cel5L may be able to bind to insoluble cellulose via a non-traditional binding domain such as the "clamp domain" found in *C. cellulovorans* endoglucanase D [45], resulting in an improved performance of the enzyme. A similar clamp domain may also be employed by CelG.

Conversions of Avicel were significantly slower and yielded lower glucose values. The control set of enzymes, containing β-glucosidase, CelI and CelK only converted 16% of the Avicel to glucose after 162 hr (Fig. 4). Supplementation with GH5 cellulases did not improve glucose yield as significantly as with filter paper as substrate. Addition of CelO, CelE, CelH, or CelG increased the conversion to 18%, 19%, 21%, and 21% respectively. Again, Cel5L produced more glucose than the cellulases with CBM modules. Supplementation with Cel5L again produced significantly more glucose from filter paper (24%) than supplementation with CelG, CelE, CelH or CelO.

To understand the relationship of Cel5L to other *C. thermocellum* GH5 cellulases, a BLASTp search was conducted using the Cel5L protein sequence. The BLASTp search results were used to construct a phylogenetic tree using the closest homologues of Cel5L and other named *C. thermocellum* GH5 cellulases (Fig. 5).

Table 2. Activities of GH5 Cellulases

	Cel5L	**CelE**	**CelH**	**CelO**	**CelG**
Gene locus	Cthe_0405	Cthe_0797	Cthe_1472	Cthe_2147	Cthe_2872
Protein sequence	32-526	30-409	36-832	34-589	37-512
AZCL-BG	+	+	+	+	+
AZCL-HEC	+	+	+	+	+
AZCL-AX	-	+	+	-	-
AZCL-GM	-	+	+	-	-
AZCL-XG	-	+	+	-	-
MUG	-	-	-	-	-
MUX	-	-	-	-	-
MUC	+	+	+	+	+
MUL	+	+	+	+	+
BG Specific Activity (u/mg)	80	25	47	19	75

Legend: Enzymatic activities determined as described in Materials and Methods. BG, β-glucan: HEC, hydroxyethyl cellulose; AX, arabinoxylan; GM, glucomannan; XG, xyloglucan, BG, barley β-glucan; +, strong activity; -, no activity

Table 3. Glucose Produced from Cellooligosaccharides and Reduced Cellooligosaccharides

	C3	**C3red**	**C4**	**C4red**	**C5**	**C5red**
Cel5L	45	0	6.2	5.5	44	7.3
CelG	19	0	2.3	4.3	19	4.0
CelO	4.0	0	0	0	2.3	0.8
CelE	61	0	44	17	63	34
CelH	18	0	142	0	57	23

Legend: Enzymatic activities determined as described in Materials and Methods and expressed as nanomoles glucose formed in 30 minutes. C3, cellotriose; C3red, cellotriitol, C4, cellotetraose; C4red, cellotetraitol, C5, cellopentaose; C5red, cellopentaitol

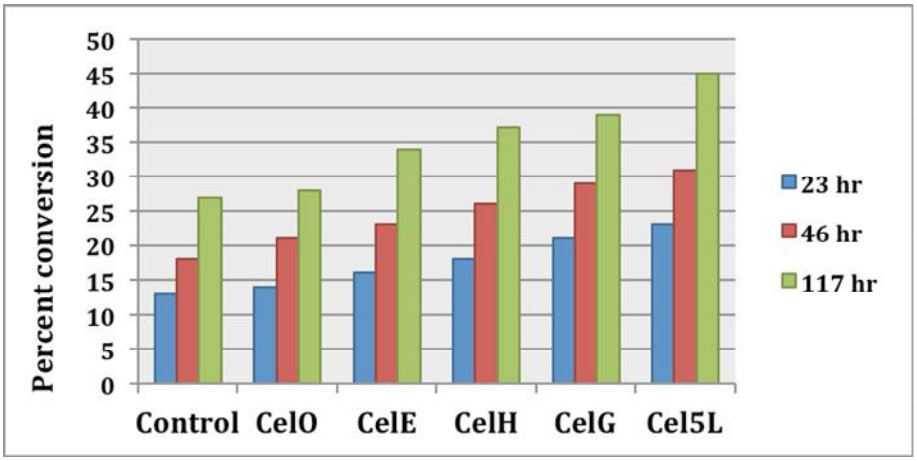

Fig. 3. Hydrolysis of Filter Paper by GH5 Cellulases

Legend: Conversion of Whatman 1 filter paper to glucose using a base set of CthCelI, CthCelK, and thermostable β-glucosidase (Control) as described in Materials and Methods

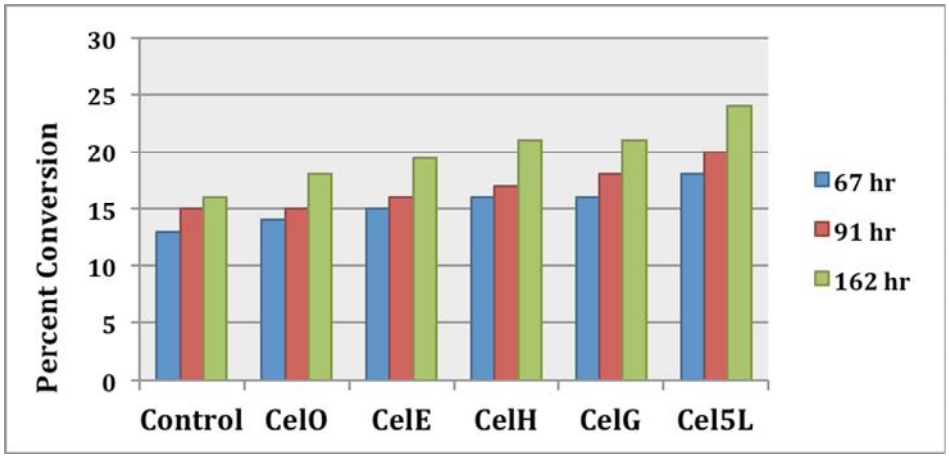

Fig. 4. Hydrolysis of Avicel by GH5 Cellulases

Legend: Conversion of Avicel to glucose using a base set of CthCelI, CthCelK, and thermostable beta-glucosidase (Control) as described in Materials and Methods

The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model. The tree with the highest log likelihood (-4977.7042) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 13 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 259 positions in the final dataset. Evolutionary analyses were conducted in MEGA5. Abbreviations: Cel5L, *C. thermocellum* Cel5L, Clcla, [*Clostridium*] *clariflavum* DSM 19732 YP_005045721; Accel, *Acetivibrio cellulolyticus,* WP_026073968; Cllen, *Cellulosilyticum lentocellum*, WP_013657594; Cerum, *Cellulosilyticum ruminicola* ACZ98599; Caobs, *Caldicellulosiruptor obsidiansis* YP_003840937; Cakro, *Caldicellulosiruptor kronotskyensis* YP_004023549; Cabes, *Caldicellulosiruptor bescii* YP_002573727; CelG, *C. thermocellum* CelG, CelB, *C. thermocellum* CelB, CelO, *C. thermocellum* CelO, CelE, *C. thermocellum* CelE, CelH, *C. thermocellum* CelH.

The phylogenetic tree surprisingly shows that Cel5L is most closely related to cellulases in four mesophiles, *Acetivibrio cellulolyticus*, *Clostridium clariflavum*, *Cellulosilyticum ruminicola* and *Clostridium lentocellum* now renamed *Cellulosilyticum lentocellum* [46]. Cel5L also possesses a high percent identity to GH5 domains in cellulases from the thermophiles *Caldicellulosiruptor obsidiansis, kronotskyensis,* and *bescii*. These Cel5L homologues are incorporated into their respective genomes in a variety of modular constructs (Fig. 6). As evidenced by the presence and absence of dockerin domains, both soluble and cellulosomal homologues of Cel5L are present in these gram-positive microbes [1]. The most complex structures are those of the three *Caldicellulosiruptor* species, where the Cel5L homologue is connected to multiple CBM3 modules and then to an additional GH family module, either a GH5 mannanase or a GH10 xylanase [47-49].

The phylogenetic tree demonstrates that Cel5L is more closely related to these seven enzymes than to its closest homologues in *C. thermocellum*, CelG and CelB. To better understand this finding, the alignment used to generate the phylogenetic tree was examined in detail. The alignment of the GH5 domains (Fig. 7) shows that CelG possesses sequence inserts of 2, 3, 4, 6, and 12 amino acids that are not found in any of the other eight cellulases, supporting the location of CelG within the phylogenetic tree. If Cel5L and CelG are the product of a recent gene duplication event that occurred only in *C. thermocellum,* Cel5L would be more closely related to CelG than to the other seven orthologs. Because these closely-related orthologs of Cel5L are found in these diverse organisms, it is unlikely that Cel5L and CelG are the product of a recent gene duplication event that occurred only in *C. thermocellum.* Further work is needed to unravel the evolutionary path of the multiple cellulases in *C. thermocellum* and other cellulytic organisms.

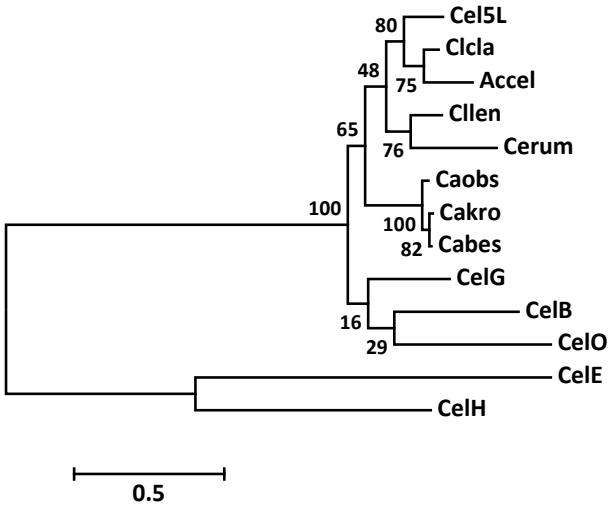

Fig. 5. Phylogenetic Tree of Cel5L

Legend: Molecular Phylogenetic analysis by Maximum Likelihood method

Organism	Structural organization
C. clariflavum	GH5⇒Dockerin
A. cellulolyticus	GH5⇒Dockerin
C. lentocellum	GH5⇒ Dockerin
C. ruminicola	GH5⇒ Dockerin
C. thermocellum Cel5L	GH5⇒ Dockerin⇒ Dockerin
C. thermocellum CelG	GH5⇒ Dockerin⇒ Dockerin
C. thermocellum CelB	GH5⇒ Dockerin⇒ Dockerin
Caldicellulosiruptor kronotskyensis	GH5⇒CBM3⇒CBM3⇒GH5
Caldicellulosiruptor obsidiansis	GH10⇒CBM3⇒CBM3⇒GH5
Caldicellulosiruptor bescii	GH5⇒CBM3⇒CBM3⇒CBM3⇒GH5

Fig. 6. Structural orthologs of Cel5L

```
Cel5L       DMYGNQVWLTGCNWFGFNTGTNVFDGVWSCNMREALKGMADRGINFLRIPISTELLYQWS
Caobs       DKDGRPVWLTGVNWFGYNTGTNVFDGVWSCNLKSTLAEIANRGFNLLRVPISAELILNWS
Cakro       DKDGRPVWLTGINWFGYNTGTNVFDGVWSCNLKDTLAEIANRGFNLLRIPISAEIILNWS
Cabes       DKDGKPVWLTGINWFGYNTGTNVFDGVWSCNLKDTLAEIANRGFNLLRIPISAEIILNWS
Cerum       DKDGNEVWLTGCNWFGFNTDTRMFDGIWAVNLEKAIKGMADHGINFLRIPATSQLLIEWQ
Cllen       DQDGNEVWLTGTNWFGFNTGTNVFDGVWSVNMKKALDEMANRGINLLRIPISTQLIYEWK
Clcla       DMNGNPVWLTGTNWFGFNTGTNVFDGVWSCNMREALTKMADRGINFLRIPVSTEILTGWK
Accel       DMSGKPVWLTGANWFGFNTGTNVFDGVWSCNMKKALIGMADRGINFLRIPVSTEILTKWK
CelG        DMYGNEVWLTGANWFGFNCSENCFHGAWY-DVKTILTSIADRGINLLRIPISTELLYSWM
            *  *. ***** ****:*   . *.* *  ::.  :  :*::*:*:**:* :::::  *

Cel5L       QGIYPK-ANVN------------DFVNPEL-KGKNSLELFDFAVQCCKEFGIKIMVDIHS
Caobs       QGIYPK-PNIN------------YYVNPEL-EGKNSLEVFDIVVQTCKEVGLKIMLDIHS
Cakro       QGIYPK-PNIN------------YYVNPEL-EGKNSLEVFDIVVQTCKEVGLKIMLDIHS
Cabes       QGIYPK-PNIN------------YYVNPEL-EGKNSLEVFDIVVQICKEVGLKIMLDIHS
Cerum       DGSGKK-PNVN------------DFVNPEL-KDMDSLQLFDETVRLCKKYGIKIMLDIHS
Cllen       NGTAKK-ANVN------------EYVNPEL-AGMSSLQLFDEAVKMCKANGMKIMLDIHS
Clcla       NGKPAMPRSLN------------DYVNPEL-KGLNSLELFDFALDVCKEVGIKVMVDVHS
Accel       SGSPPAPGSLN------------DFVNPEL-KGMDSLELFDFALNVCKEEGIKVMLDVHC
CelG        IGKPNPVSSVTasnnppyhvvnpDFYDPETDDVKNSMEIFDIIMGYCKELGIKVMIDIHS
             *      .:.             : :**     .*:::**  :  **  *:*:*:*:*.

Cel5L       PATDAMGHMYPLWYDGQ------FTTEIWISTLEWLTERYKNDDTILALDLKNEPHGTPG
Caobs       IKTDAMGHIYPVWYDDK------YT-EDFYKACEWITNRYKNDDTIIAFDLKNEPHGKPW
Cakro       IKTDAMGHIYPVWYDDK------FTPEDFFKACEWITNRYKKDDTIIAFDLKNEPHGKPW
Cabes       IKTDAMGHIYPVWYDDK------FTPEDFYKACEWITNRYKNDDTIIAFDLKNEPHGKPW
Cerum       ASSDGQRHYYPLWFGEG------YTIDDYHSSLEWFAERYKNEDTILAIDLENEPHGKPY
Cllen       AETNAMGHIYPVWYDEA------HPIEMYHETLEWFTERYKNDDTILAIDLKNEPHGKPG
Clcla       PKSEAMGHNYPVWYDET------YDTEAWISALEWLTERYKNDDTILAIDLKNEPHGKPY
Accel       PQSQAMGHNYPVWYDGA------YNTEAWMSALEWMAERYKNDDTILAFDLKNEPHGKPT
CelG        PDANNSGHNYELWYGKEtstcgvVTTKMWIDTLVWLADKYKNDDTIIAFDLKNEPHGKRG
              ::   * * :*:            . : .:  *::::**::***:*:**:*****.

Cel5L       S-----ELMAKWDGSTDLNNWKHAAETCAKRILAINPNILIVVEGVEVYPKP--GYDYTA
Caobs       Q----DATFAKWDDSTDINNWKYAAETCAKGILNINPNLLIVIEGIEAYPKD--DVTWTS
Cakro       Q----DTTFAKWDNSTDINNWKYAAETCAKRILNINPNLLIVIEGIEAYPKD--DVTWTS
Cabes       Q----DTTFAKWDNSTDINNWKYAAETCAKRILNINPNLLIVIEGIEAYPKD--DVTWTS
Cerum       H----EPLWAKWDGSEDENNWKYAAEVVANKVLDINPNLLVMIEGVEAYPKE--GYDYTT
Cllen       Q----DKVWAKWDNSTDENNWKYAAETAAAKVLAINPNLLIVIEGVEAYPME--GYDYTT
Clcla       E-----KLMAKWDNSTDINNWKYAAETCAKRILKINPNLLIVIEGVEVYPKE--GYDYTA
Accel       D-----TVMAKWDDSKDLNNWKYAAQTCGERVLAINPNLLIMVEGIEAYPKE--GNDYTA
CelG        YtaevPKLLAKWDNSTDENNWKYAAETCAKAILEVNPKVLIVIEGVEQYPKTEKGYTYDT
                     **** * * ****:**:. .  :* :**::*:::**:* **       : :

Cel5L       VDEWGK---ESKYFYNWWGGNLRGVRDYPIDLG-KHQKQLVYSPHDYGPLVHKQPWFYEG
Caobs       KS-------SSDYYSTWWGGNLRGVRKYPINLG-KYQNKVVYSPHDYGPSVYQQPWFYPG
Cakro       KS-------YSDYYSTWWGGNLRGVKKYPINLG-KYQNKVVYSPHDYGPSVYQQPWFYSG
Cabes       KS-------YSDYYSTWWGGNLRGVKKYPINLG-KYQNKVVYSPHDYGPSVYQQPWFYPG
Cerum       QDEYEN----EHYYYTWWGGNLRGVKDYPVTLKPEHQNQVVYSPHDYGPTVYMQNWFKGD
Cllen       RDEYMK----PHYYYNWWGGNLRGVKDYPIDLG-VYQSQVVYSPHDYGPLVYAQPWFEKD
Clcla       VDEWGK---ESRYYYNWWGGNLRGVKDYPIDLG-QYQKQVVYSPHDYGPLVHKQPWFYDG
Accel       VDEWGK---ESHYYFNWWGGNFRGVKDNPLDLG-ENQDQLVYSPHDYGPMVYKQSWFYDG
CelG        PDIWGATGDASPWYSAWWGGNLRGVKDYPIDLG-PLNSQIVYSPHDYGPSVYAQPWFEKD
             .          ::  *****:***:. *: *    :.::********* *: * **

Cel5L       FNKETLYNDCWRDNWAYIHEENIAPLIVGEWGGFMDRGDNEKWMKALRDYMIENKISHTF
Caobs       FTKETLLQDCWRPNWVYIMEENIAPLLIGEWGGYLDGGDNEKWMKCLRDYIIENHIHHTF
Cakro       FTKESLLQDCWRPNWAYIMEENIAPLLIGEWGGYLDGADNEKWMRYLRDYIIENHIHHTF
Cabes       FTKESLLQDCWRPNWAYIMEENIAPLLIGEWGGYLDGADNEKWMRYLRDYIIENHIHHTF
Cerum       FTKESVYNDCWKDNWAYIAESNTAPLLIGEWGGFLDGGKNEQWLNDLRDFIVENKINHTF
Cllen       FTTETLYEDCWGDNWAYINEEGIAPLLMGEWGGFMDGGKNEKWMTLLRDYMIKNKIHHTF
Clcla       FTKETLYNDCWKDNWAYIYEEGIAPLLIGEWGGYMDGGPNEKWMIALRDYIVENRIHHTF
Accel       FTKESVYNDCWKDNWAYIYEDGIAPLLIGEWGGFMDGGDNEKWMTAMRDYIVDNRIHHTF
CelG        FTMQTLLDDYWYDTWAYIHDQGIAPILIGEWGGHMDGGKNQKWMTLLRDYIVQNRIHHTF
            *. ::: :* *  .*.** :.  **:::*****.:* . *::*:  :**:::.*:* ***

Cel5L       WCYNANS
Caobs       WCFNANS
Cakro       WCFNANS
Cabes       WCFNANS
Cerum       WCYNPNS
Cllen       WCYNANS
Clcla       WCYNANS
Accel       WCYNANS
CelG        WCINPNS
            ** * **
```

Fig. 7. Alignment of GH5 domains

Legend: Alignment of GH5 domains using Clustal Omega as described in Methods. Inserts found only in CelG are shown in lower case. Abbreviations: Cel5L, C. thermocellum Cel5L, Clcla, [Clostridium] clariflavum DSM 19732 YP_005045721; Accel, Acetivibrio cellulolyticus, WP_026073968; Cllen, Cellulosilyticum lentocellum, WP_013657594; Cerum, Cellulosilyticum ruminicola ACZ98599; Caobs, Caldicellulosiruptor obsidiansis YP_003840937; Cakro, Caldicellulosiruptor kronotskyensis YP_004023549; Cabes, Caldicellulosiruptor bescii YP_002573727; CelG, C. thermocellum CelG

Recently, a reclassification of clostridial species has been proposed [50]. In this reclassification, *C. thermocellum* and *C. clariflavum* have been transferred to a new family and genus, *Ruminococcaceae* and *Ruminiclostridium*. This reclassification places these two cellulytic, cellulosome-producing organisms together with other cellulosome-producing organisms such as *C. cellulyticum*, *C. leptum*, and *R. albus*. Further work is necessary to fully understand the phylogenetic and physiological relationships among these cellulolytic organisms.

4. CONCLUSION

C. thermocellum Cel5L is a novel thermostable cellulase that possesses both *endo*-cellulase and *exo*-cellulase activities. Based on its action on cellooligosaccharides, the enzyme cleaves from the non-reducing end of cellulose chains to produce predominantly cellobiose. The performance of Cel5L was significantly better than that of the other *C. thermocellum* GH5 enzymes when tested in hydrolysis of insoluble cellulosic substrates. The reason Cel5L is able to degrade cellulose in the absence of a classical CBM module is unclear and a potential area for future study. The potential physiological importance of Cel5L in cellulose degradation is reiterated by the presence of orthologs of Cel5L in other gram-positive cellulolytic organisms, both mesophilic and thermophilic. The presence of these orthologs in a variety of constructs, both soluble and cellulosomal, leaves many unanswered questions about how cellulose degradation, and the cellulases found, today arose in today's cellulytic bacteria.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Brumm P. Bacterial Genomes: What They Teach Us about Cellulose Degradation. Biofuels. 2013;4:669–681.
2. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev. 2002;66:506-577, table of contents.
3. Ng TK, Weimer TK, Zeikus JG. Cellulolytic and physiological properties of Clostridium thermocellum. Arch Microbiol. 1977;114:1-7.
4. Weimer PJ, Zeikus JG. Fermentation of cellulose and cellobiose by Clostridium thermocellum in the absence of Methanobacterium thermoautotrophicum. Appl Environ Microbiol. 1977;33:289-297.
5. Lamed R, Setter E, Bayer EA. Characterization of a cellulose-binding, cellulase-containing complex in Clostridium thermocellum. J Bacteriol. 1983;156:828-836.
6. Bayer EA, Setter E, Lamed R. Organization and distribution of the cellulosome in Clostridium thermocellum. J Bacteriol. 1985;163:552-559.
7. Himmel ME, Xu Q, Luo Y, Ding S, Lamed R, et al. Microbial enzyme systems for biomass conversion: emerging paradigms. Blofuels. 2010;1:323-341.
8. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, et al. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Research. 2009;37:D233-238.
9. Park BH, Karpinets TV, Syed MH, Leuze MR, Uberbacher EC. CAZymes Analysis Toolkit (CAT): web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. Glycobiology. 2010; 20:1574-1584.
10. Henrissat B, Davies G. Structural and sequence-based classification of glycoside hydrolases. Current opinion in structural biology. 1997;7:637-644.
11. Feinberg L, Foden J, Barrett T, Davenport KW, Bruce D, et al. Complete genome sequence of the cellulolytic thermophile Clostridium thermocellum DSM1313. Journal of Bacteriology. 2011;193:2906-2907.
12. Brown SD, Lamed R, Morag E, Borovok I, Shoham Y, et al. Draft genome sequences for Clostridium thermocellum wild-type strain YS and derived cellulose adhesion-defective mutant strain AD2. Journal of Bacteriology. 2012;194:3290-3291.
13. Alzari PM, Souchon H, Dominguez R. The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from Clostridium thermocellum. Structure. 1996;4:265-275.
14. Joliff G, Beguin P, Aubert JP. Nucleotide sequence of the cellulase gene celD encoding endoglucanase D of Clostridium

thermocellum. Nucleic Acids Res. 1986;14: 8605-8613.
15. Hall J, Hazlewood GP, Barker PJ, Gilbert HJ. Conserved reiterated domains in Clostridium thermocellum endoglucanases are not essential for catalytic activity. Gene. 1988;69:29-38.
16. Navarro A, Chebrou MC, Beguin P, Aubert JP. Nucleotide sequence of the cellulase gene celF of Clostridium thermocellum. Res Microbiol. 1991;142:927-936.
17. Hirano N, Hasegawa H, Nihei S, Haruki M. Cell-free protein synthesis and substrate specificity of full-length endoglucanase CelJ (Cel9D-Cel44A), the largest multi-enzyme subunit of the Clostridium thermocellum cellulosome. FEMS Microbiol Lett. 2013;344:25-30.
18. Zverlov VV, Velikodvorskaya GA, Schwarz WH. A newly described cellulosomal cellobiohydrolase, CelO, from Clostridium thermocellum: investigation of the exo-mode of hydrolysis, and binding capacity to crystalline cellulose. Microbiology. 2002;148:247-255.
19. Arai T, Ohara H, Karita S, Kimura T, Sakka K, et al. Sequence of celQ and properties of celQ, a component of the Clostridium thermocellum cellulosome. Appl Microbiol Biotechnol. 2001;57:660-666.
20. Zverlov VV, Schantz N, Schwarz WH A major new component in the cellulosome of Clostridium thermocellum is a processive endo-beta-1, 4-glucanase producing cellotetraose. FEMS Microbiol Lett. 2005;249:353-358.
21. Morag E, Bayer EA, Hazlewood GP, Gilbert HJ, Lamed R. Cellulase Ss (CelS) is synonymous with the major cellobiohydrolase (subunit S8) from the cellulosome of Clostridium thermocellum. Appl Biochem Biotechnol. 1993;43:147-151.
22. Gilad R, Rabinovich L, Yaron S, Bayer EA, Lamed R, et al. CelI, a noncellulosomal family 9 enzyme from Clostridium thermocellum, is a processive endoglucanase that degrades crystalline cellulose. J Bacteriol. 2003;185:391-398.
23. Berger E, Zhang D, Zverlov VV, Schwarz WH. Two noncellulosomal cellulases of Clostridium thermocellum, Cel9I and Cel48Y, hydrolyse crystalline cellulose synergistically. FEMS Microbiol Lett. 2007;268:194-201.
24. Schwarz WH, Schimming S, Rucknagel KP, Burgschwaiger S, Kreil G, et al. Nucleotide sequence of the celC gene encoding endoglucanase C of Clostridium thermocellum. Gene. 1988;63:23-30.
25. Beguin P, Millet J, Aubert JP. The cloned cel (cellulose degradation) genes of Clostridium thermocellum and their products. Microbiol Sci. 1987;4:277-280.
26. Hazlewood GP, Davidson K, Clarke JH, Durrant AJ, Hall J, et al. Endoglucanase E, produced at high level in Escherichia coli as a lacZ' fusion protein, is part of the Clostridium thermocellum cellulosome. Enzyme Microb Technol. 1990;12:656-662.
27. Yague E, Beguin P, Aubert JP. Nucleotide sequence and deletion analysis of the cellulase-encoding gene celH of Clostridium thermocellum. Gene. 1990;89:61-67.
28. Lemaire M, Beguin P. Nucleotide sequence of the celG gene of Clostridium thermocellum and characterization of its product, endoglucanase CelG. J Bacteriol. 1993;175:3353-3360.
29. Gold ND, Martin VJ. Global view of the Clostridium thermocellum cellulosome revealed by quantitative proteomic analysis. J Bacteriol. 2007;189:6787-6795.
30. Brumm P. Bacterial Genomes: What They Teach Us about Cellulose Degradation. BIofuels; 2013.
31. Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual. NY: Cold Spring Harbor Laboratory Press; 1989.
32. Nelson N. A Photometric Adaptation of the Somogyi Method for the Determination of Glucose. J Biological Chemistry. 1944;153: 375-380.
33. Brumm P, Hermanson S, Hochstein B, Boyum J, Hermersmann N, et al. Mining Dictyoglomus turgidum for Enzymatically Active Carbohydrases. Applied Biochemistry and Biotechnology. 2010;1-10.
34. Altschul SF, Madden TL, Schaeffer AA, Zhang J, Zhang Z, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research. 1997;25:3389-3402.
35. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology. 2011;7:539.
36. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell

using TargetP, SignalP, and related tools. Nature Protocols. 2007;2:953-971.
37. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Computer applications in the biosciences: CABIOS. 1992;8:275-282.
38. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular biology and evolution. 2011;28:2731-2739.
39. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42:D490-495.
40. Aspeborg H, Coutinho PM, Wang Y, Brumer H, 3rd, Henrissat B. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). BMC Evol Biol. 2012;12:186.
41. Abdeev RM, Goldenkova IV, Musiychuk KA, Piruzian ES. Exploring the properties of thermostable Clostridium thermocellum cellulase CelE for the purpose of its expression in plants. Biochemistry (Mosc). 2001;66:808-813.
42. Zhang YH, Lynd LR. A functionally based model for hydrolysis of cellulose by fungal cellulase. Biotechnol Bioeng. 2006;94:888-898.
43. Reyes-Ortiz V, Heins RA, Cheng G, Kim EY, Vernon BC, et al. Addition of a carbohydrate-binding module enhances cellulase penetration into cellulose substrates. Biotechnol Biofuels. 2013;6:93.
44. Byrt CS, Cahyanegara R, Grof CP. Plant carbohydrate binding module enhances activity of hybrid microbial cellulase enzyme. Front Plant Sci. 2012;3:254.
45. Bianchetti CM, Brumm P, Smith RW, Dyer K, Hura GL, et al. Structure, Dynamics, and Specificity of Endoglucanase D from Clostridium cellulovorans. Journal of molecular biology. 2013;425:4267-4285.
46. Cai S, Dong X. Cellulosilyticum ruminicola gen. nov., sp. nov., isolated from the rumen of yak, and reclassification of Clostridium lentocellum as Cellulosilyticum lentocellum comb. nov. International Journal of Systematic and Evolutionary Microbiology. 2010;60:845-849.
47. VanFossen AL, Ozdemir I, Zelin SL, Kelly RM Glycoside hydrolase inventory drives plant polysaccharide deconstruction by the extremely thermophilic bacterium Caldicellulosiruptor saccharolyticus. Biotechnol Bioeng. 2011;108:1559-1569.
48. Blumer-Schuette SE, Lewis DL, Kelly RM Phylogenetic, microbiological, and glycoside hydrolase diversities within the extremely thermophilic, plant biomass-degrading genus Caldicellulosiruptor. Appl Environ Microbiol. 2010;76:8084-8092.
49. Dam P, Kataeva I, Yang SJ, Zhou F, Yin Y, et al. Insights into plant biomass conversion from the genome of the anaerobic thermophilic bacterium Caldicellulosiruptor bescii DSM 6725. Nucleic Acids Res. 2011;39:3240-3254.
50. Yutin N, Galperin MY. A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. Environ Microbiol. 2013;15: 2631-2641.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=847&id=3&aid=7625