

1
2
3 ***Clostridium thermocellum* Cel5L – Cloning and**
4 **Characterization of a New, Thermostable GH5**
5 **Cellulase**

6
7
8
9
10
11 **ABSTRACT**
12

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 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.

13
14 *C. thermocellum*, cellulase Cel5L, Cthe_0405, endoglucanase, biomass

15
16 **1. INTRODUCTION**
17

18 Cellulases are endoglucanases that degrade native (crystalline) cellulose either from the
19 chain ends (*exo*-acting cellulases) or the interior of the chain (*endo*-acting cellulases). This
20 definition is sometimes, but not always, used to distinguish cellulases from other
21 endoglucanases such as *beta*-glucanases (which are active on *beta*-glucan but not native
22 cellulose) [1]. Finding novel thermostable bacterial cellulases has been difficult because
23 growth on cellulosic substrates and production of thermostable cellulases is rare among
24 thermophilic microorganisms [2]. In the 1980s, initial work showed that *Clostridium*
25 *thermocellum* (Cthe), a thermophilic anaerobic cellulose-degrading ethanologen [3,4],
26 utilizes complex, high molecular weight structures termed the cellulosome [5,6] to rapidly
27 degrade cellulose *in vivo*. Progress on understanding the structure and function of the
28 cellulosome has been recently reviewed [1,7] and many questions on the exact mechanism
29 of cellulose degradation by cellulosomal cellulases remains under study. Glycoside
30 hydrolases (GH) are enzymes that catalyze the hydrolysis of the glycosidic linkage of sugar
31 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].

49

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

Gene Number	GH Family	CBM Family	Name	Ref.
Cthe_0405	GH5		Cel5L	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]

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

52

53 The Cthe genome codes for significantly more predicted cellulases than gram-positive
54 organisms producing soluble cellulolytic enzymes, which typically have seven to twelve
55 predicted cellulases [30]. Efficient cellulosomal function may require the larger number of
56 potential cellulases and GH5 cellulases in *C. thermocellum*. It is unclear if this higher number
57 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.

2. MATERIAL AND METHODS

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, Cell, 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-glucopyranoside (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 K₂HPO₄, 2.2 g/l KH₂PO₄, 30 mg/l kanomycin sulfate and 4.0 g/l glycerol added after autoclaving) was used for protein expression in *E. coli* liquid cultures.

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 CellLytic IIB reagent (Sigma). The lysates were assayed for cellulase activity at 70°C in 0.2 ml of 50 mM acetate buffer, pH 5.8 containing 0.2% AZCL-HE-Cellulose.

Enzyme Production and Purification

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

111 Enzyme assays

112 The *endo*-specificity of cellulases was determined in 0.50 ml of 50 mM acetate buffer, pH
113 5.8, containing 0.2% azurine cross-linked-labeled (AZCL) insoluble substrates and 20 µg of
114 enzyme protein. Assays were performed at 70°C, with shaking at 1400 rpm, for 60 minutes
115 in a Thermomixer R (Eppendorf, Hamburg, Germany). Tubes were clarified by centrifugation
116 and absorbance values determined using a Bio-Tek EL_x800 plate reader at 600 nm. The
117 *exo*-specificity of cellulases was determined in 0.10 ml of 50 mM acetate buffer, pH 5.8,
118 containing 1.0 µg of 4-methyl umbelliferyl substrate and 10 µg of enzyme protein. Tubes
119 were incubated at 70°C, with shaking at 500 rpm, for 30 minutes in a Thermomixer R; after
120 incubation, the tubes were examined using a hand-held UV lamp and compared to negative
121 and positive controls.

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

128 Effects of temperature and pH on enzyme activity.

129 The temperature optimum of Cel5L was determined using the reducing sugar assay, with
130 2%β-glucan as substrate at pH 5.8 between 30°C and 90° C. The pH optimum of Cel5L was
131 determined using the reducing sugar assay with 2%β-glucan as substrate at 70°C between
132 pH values of 3.0 and 11.0. All experiments were done in triplicate.

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

139 Cellulose hydrolysis experiments were conducted at 60°C in a final volume of 1.0 ml of 50
140 mM acetate buffer, pH 5.8, containing 5 mM CaCl₂ and either 2.0 mg Avicel microcrystalline
141 cellulose or 3.4 mg of Whatman 1 filter paper and 100 micrograms of each purified cellulase.
142 Thermostable *beta*-glucosidase (100 micrograms was added to all reactions to convert
143 cellodextrin products to glucose. Glucose formed was determined using the Megazyme D-
144 Glucose (GOPOD Format) Assay Kit according to the manufacturer's directions in at least
145 triplicate.

146 Protein concentrations were measured using the Coomassie Plus reagent using bovine
147 serum albumin as standard according to the manufacturer's instructions. Gel electrophoresis

148 was performed using 4-20% acrylamide Mini-PROTEAN® gels according to the
149 manufacturer's instructions and stained with Coomassie R-250.

150 Bioinformatics

151 The GenBank accession number of the sequence reported in this paper is Gene ID:
152 4808408. The complete *Clostridium thermocellum* strain 27405^T genome is available at
153 http://www.ncbi.nlm.nih.gov/nucore/NC_009012. InterProScan Family
154 analysis(<http://www.ebi.ac.uk/Tools/InterProScan/>) was used to identify protein domains
155 within Cel5L and retrieve sequences of *C. thermocellum* cellulases. BLASTp (Basic Local
156 Alignment Search Tool [34]) (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) analysis tools were
157 used to identify closest orthologs of Cel5L. Multiple sequence alignments were done using
158 Clustal Omega [35] at <http://www.ebi.ac.uk/Tools/msa/clustalo/help/>. Glycosyl hydrolase
159 predictions were obtained from http://www.cazy.org/geno/acc_gen.html. Signal sequence
160 predictions were determined using (<http://www.cbs.dtu.dk/services/SignalP/> [36]). The
161 evolutionary history was inferred by using the Maximum Likelihood method based on the
162 JTT matrix-based model [37]. Initial tree(s) for the heuristic search were obtained
163 automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise
164 distances estimated using a JTT model, and then selecting the topology with superior log
165 likelihood value. Evolutionary analyses were conducted in MEGA5 [38]

167 3. RESULTS AND DISCUSSION

168 Screening

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

181

182 Figure 1. Predicted Sequence of Cthe_0405, Cel5L

183

10	20	30	40	50	60
MRKVKALLLG	LIVLAVALLP	TVSFKSPTVA	ADPNDDWLH	VEGNKIVDMY	GNQVWLTGCN
70	80	90	100	110	120
WFGFNTGTNV	FDGVWSCNMR	EALKGMDRG	INFLRIPIST	ELLYQWSQGI	YPKANVNDV
130	140	150	160	170	180
NPELKGKNSL	ELFDFAVQCC	KEFGIKIMVD	IHSPATDAMG	HMYPLWYDGG	FTTEIWISTL
190	200	210	220	230	240
EWLTERYKND	DTILALDLKN	EPHGTPGSEL	MAKWDGSTDL	NNWKHAAETC	AKRILAINPN
250	260	270	280	290	300
ILIVVEGVEV	YPKPGYDYTA	VDEWGKESKY	FYNWGGNLR	GVRDYPIDLG	KHQQQLVYSP
310	320	330	340	350	360
HDYGPLVHKQ	PWFYEGFNKE	TLYNDWCWRDN	WAYIHEENIA	PLIVGEWGGF	MDRGDNEKWM
370	380	390	400	410	420
KALRDYMIEN	KISHTFCYN	ANSGDTGGLV	YYDFITWDEE	KYALLKPALW	QTEDGKFIGL
430	440	450	460	470	480
DHQIPLGSNG	ITVTEYGGY	IPEPSPTATV	PDVPTPSHSF	EIEKGDVNGD	GNVNSTDVVW
490	500	510	520		
LRRFLKLVE	DFPVPSPGKA	ADMNDGNIN	STDMIALKRK	VLKIPI	

184

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

190

191 Cel5L Cloning, Purification and Characterization

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

194 Cel5L F: TACGGTAATCAGGTCTGGCTGACCGGCT-3' (506023-506050)

195 Cel5L R: TTACCCCTTCTCTATTTGAAAGAATGCGA-3' (507244-507270)

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

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

202

203 **Figure 2. SDS PAGE of GH5 cellulases used in this work**



204

205

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

209

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

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

229

230

231 **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

232

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

236

237 Enzyme specificity for **exo-activity** was determined using 4-methylumbelliferyl- β -D-linked
238 substrates, cellooligosaccharides, and reduced cellooligosaccharides. **All five GH5** cellulases
239 also possess *exo*-activity, hydrolyzing both 4-methylumbelliferyl- β -D-cellobioside and 4-
240 methylumbelliferyl- β -D-lactoside. All five cellulases have no activity on either 4-
241 methylumbelliferyl- β -D-xylopyranoside or 4-methylumbelliferyl- β -D- glucopyranoside (Table
242 2). Production of glucose from cellooligosaccharides and reduced cellooligosaccharides
243 was used to determine if Cel5L and the other cellulases were active on the reducing or non-
244 reducing end of cellooligosaccharides. The five enzymes show production of glucose from
245 cellotriose, but not cellotriitol, indicating all five cellulases attack at the non-reducing end of
246 the cellooligosaccharides (Table 3). Cel5L and CelG showed significantly lower production
247 of glucose from cellotetraose than from cellotriose or cellopentaose, suggesting a release of
248 cellobiose from the non-reducing end of the cellooligosaccharides. CelO showed very low
249 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.

Table 3. Glucose Produced from Cellooligosaccharides and Reduced Cellooligosaccharides

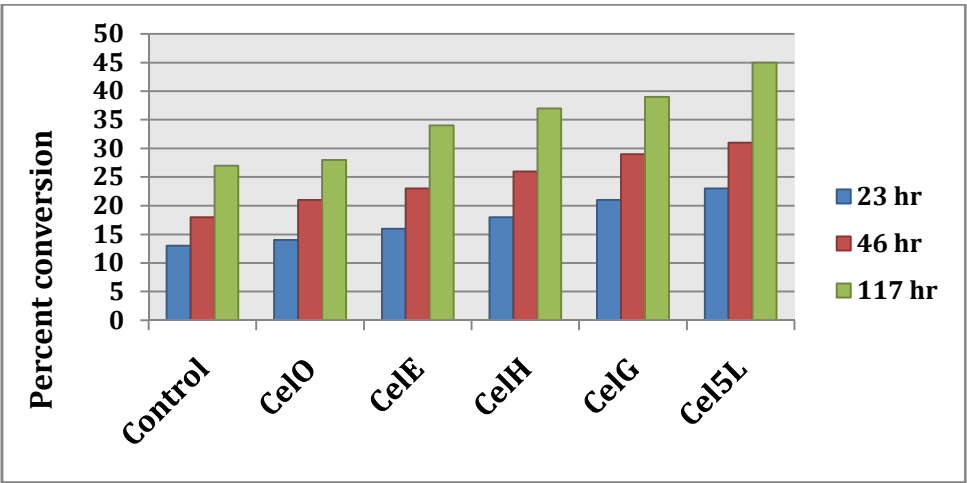
	C3	C3red	C4	C4red	C5	C5red
Cel5L	45	0	6.2	5.5	44	7.3
CelG	19	0	2.3	4.3	19	4.0
CelO	4.0	0	0	0	2.3	0.8
CelE	61	0	44	17	63	34
CelH	18	0	142	0	57	23

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.

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* CelL and CelK to provide exo-cellulase activity and an excess of thermostable bacterial β -glucosidase to convert cellobiose and cellooligosaccharides to glucose. The control set of enzymes, containing β -glucosidase, CelL and CelK converted 27% of the filter paper to glucose after 117 hr (Figure 4). 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 lyse 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.

283 **Figure 4 Hydrolysis of Filter Paper by GH5 Cellulases**

284



285

286

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

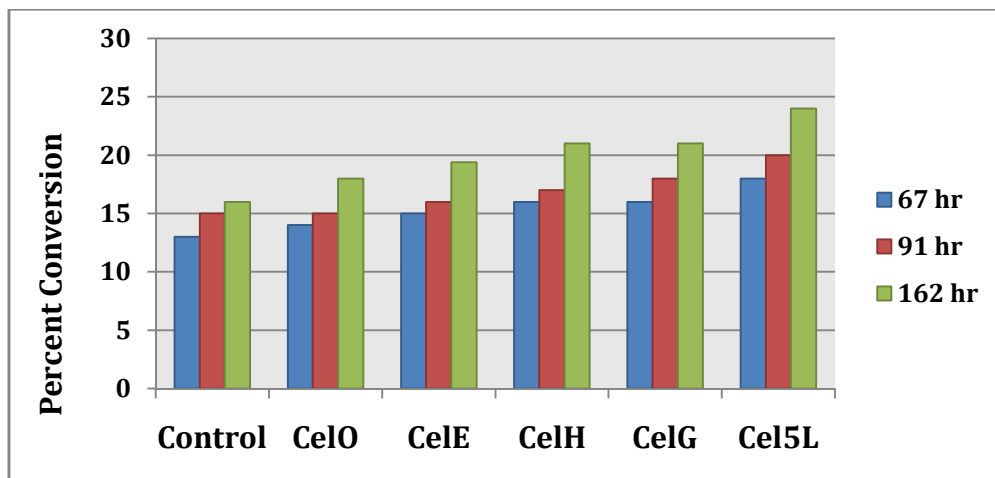
289

290 Conversions of Avicel were significantly slower and yielded lower glucose values. The
291 control set of enzymes, containing β -glucosidase, CelL and CelK only converted 16% of the
292 Avicel to glucose after 162 hr (Figure 5). Supplementation with GH5 cellulases did not
293 improve glucose yield as significantly as with filter paper as substrate. Addition of CelO,
294 CelE, CelH, or CelG increased the conversion to 18%, 19%, 21%, and 21% respectively.
295 Again, Cel5L produced more glucose than the cellulases with CBM modules.
296 Supplementation with Cel5L again produced significantly more glucose from filter paper
297 (24%) than supplementation with CelG, CelE, CelH or CelO.

298

299 **Figure 5 Hydrolysis of Avicel by GH5 Cellulases**

300



301

302

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

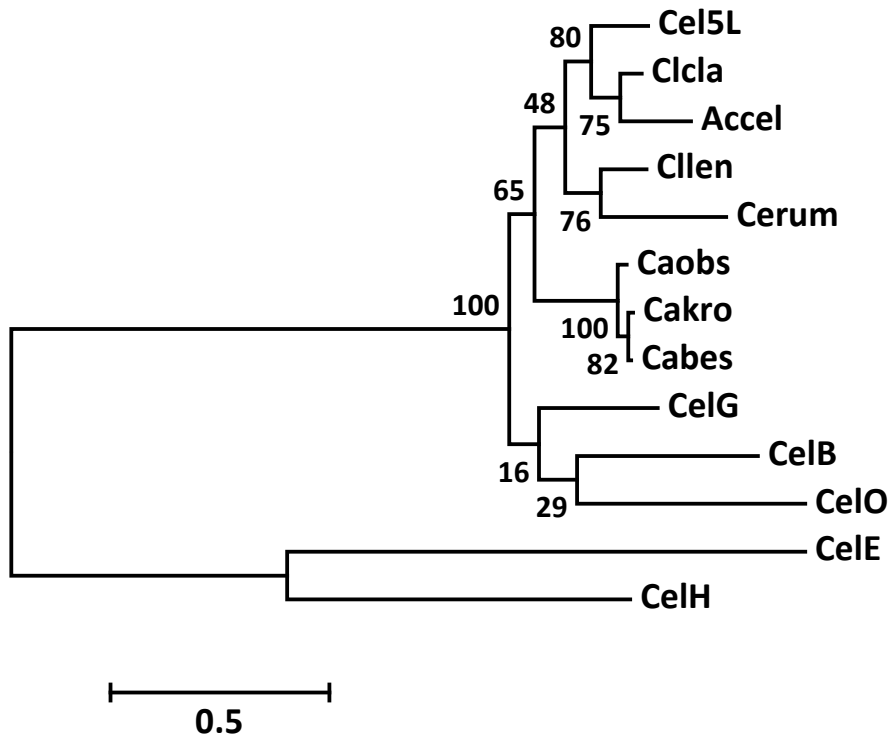
305

306 To understand the relationship of Cel5L to other *C. thermocellum* GH5 cellulases, a BLASTp
 307 search was conducted using the Cel5L protein sequence. The BLASTp search results were
 308 used to construct a phylogenetic tree using the closest homologues of Cel5L and other
 309 named *C. thermocellum* GH5 cellulases (Figure 6).

310

311 **Figure 6. Phylogenetic Tree of Cel5L**

312



313

314 Legend: Molecular Phylogenetic analysis by Maximum Likelihood method

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

332

333 The phylogenetic tree surprisingly shows that Cel5L is most closely related to cellulases in
 334 four mesophiles, *Acetivibrio cellulolyticus*, *Clostridium clariflavum*, *Cellulosilyticum ruminicola*
 335 and *Clostridium lentocellum* now renamed *Cellulosilyticum lentocellum*[46]. Cel5L also
 336 possesses a high percent identity to GH5 domains in cellulases from the thermophiles
 337 *Caldicellulosiruptor obsidiansis*, *kronotskyensis*, and *bescii*. These Cel5L homologues are

incorporated into their respective genomes in a variety of modular constructs (Figure 7). 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].

344

345 **Figure 7. Structural Orthologs of Cel5L**

Organism	Structural Organization
<i>C. clariflavum</i>	GH5⇒Dockerin
<i>A. cellulolyticus</i>	GH5⇒Dockerin
<i>C. lentocellum</i>	GH5⇒Dockerin
<i>C. ruminicola</i>	GH5⇒Dockerin
<i>C. thermocellum</i> Cel5L	GH5⇒Dockerin⇒Dockerin
<i>C. thermocellum</i> CelG	GH5⇒Dockerin⇒Dockerin
<i>C. thermocellum</i> CelB	GH5⇒Dockerin⇒Dockerin
<i>Caldicellulosiruptor kronotskyensis</i>	GH5⇒CBM3⇒CBM3⇒GH5
<i>Caldicellulosiruptor obsidiansis</i>	GH10⇒CBM3⇒CBM3⇒GH5
<i>Caldicellulosiruptor bescii</i>	GH5⇒CBM3⇒CBM3⇒CBM3⇒GH5

346

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 (Figure 8) 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.

359

362

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

370

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

378

379 4. CONCLUSION

380

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

393

394

395

396 REFERENCES

397

- 398 1. Brumm P (2013) Bacterial Genomes: What They Teach Us about Cellulose Degradation.
399 Biofuels 4: 669–681.
- 400 2. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization:
401 fundamentals and biotechnology. Microbiol Mol Biol Rev 66: 506-577, table of
402 contents.
- 403 3. Ng TK, Weimer TK, Zeikus JG (1977) Cellulolytic and physiological properties of
404 *Clostridium thermocellum*. Arch Microbiol 114: 1-7.
- 405 4. Weimer PJ, Zeikus JG (1977) Fermentation of cellulose and cellobiose by *Clostridium*
406 *thermocellum* in the absence of *Methanobacterium thermoautotrophicum*. Appl
407 Environ Microbiol 33: 289-297.
- 408 5. Lamed R, Setter E, Bayer EA (1983) Characterization of a cellulose-binding, cellulase-
409 containing complex in *Clostridium thermocellum*. J Bacteriol 156: 828-836.

- 410 6. Bayer EA, Setter E, Lamed R (1985) Organization and distribution of the cellulosome in
411 *Clostridium thermocellum*. *J Bacteriol* 163: 552-559.
- 412 7. Himmel ME, Xu Q, Luo Y, Ding S, Lamed R, et al. (2010) Microbial enzyme systems for
413 biomass conversion: emerging paradigms. *Biofuels* 1: 323-341.
- 414 8. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, et al. (2009) The
415 Carbohydrate-Active EnZymes database (CAZy): an expert resource for
416 Glycogenomics. *Nucleic Acids Research* 37: D233-238.
- 417 9. Park BH, Karpinets TV, Syed MH, Leuze MR, Uberbacher EC (2010) CAZymes Analysis
418 Toolkit (CAT): web service for searching and analyzing carbohydrate-active
419 enzymes in a newly sequenced organism using CAZy database. *Glycobiology* 20:
420 1574-1584.
- 421 10. Henrissat B, Davies G (1997) Structural and sequence-based classification of glycoside
422 hydrolases. *Current opinion in structural biology* 7: 637-644.
- 423 11. Feinberg L, Foden J, Barrett T, Davenport KW, Bruce D, et al. (2011) Complete genome
424 sequence of the cellulolytic thermophile *Clostridium thermocellum* DSM1313.
425 *Journal of Bacteriology* 193: 2906-2907.
- 426 12. Brown SD, Lamed R, Morag E, Borovok I, Shoham Y, et al. (2012) Draft genome
427 sequences for *Clostridium thermocellum* wild-type strain YS and derived cellulose
428 adhesion-defective mutant strain AD2. *Journal of Bacteriology* 194: 3290-3291.
- 429 13. Alzari PM, Souchon H, Dominguez R (1996) The crystal structure of endoglucanase
430 CelA, a family 8 glycosyl hydrolase from *Clostridium thermocellum*. *Structure* 4: 265-
431 275.
- 432 14. Joliff G, Beguin P, Aubert JP (1986) Nucleotide sequence of the cellulase gene *celD*
433 encoding endoglucanase D of *Clostridium thermocellum*. *Nucleic Acids Res* 14:
434 8605-8613.
- 435 15. Hall J, Hazlewood GP, Barker PJ, Gilbert HJ (1988) Conserved reiterated domains in
436 *Clostridium thermocellum* endoglucanases are not essential for catalytic activity.
437 *Gene* 69: 29-38.
- 438 16. Navarro A, Chebrou MC, Beguin P, Aubert JP (1991) Nucleotide sequence of the
439 cellulase gene *celF* of *Clostridium thermocellum*. *Res Microbiol* 142: 927-936.
- 440 17. Hirano N, Hasegawa H, Nihei S, Haruki M (2013) Cell-free protein synthesis and
441 substrate specificity of full-length endoglucanase CelJ (Cel9D-Cel44A), the largest
442 multi-enzyme subunit of the *Clostridium thermocellum* cellulosome. *FEMS Microbiol*
443 *Lett* 344: 25-30.
- 444 18. Zverlov VV, Velikodvorskaya GA, Schwarz WH (2002) A newly described cellulosomal
445 cellobiohydrolase, CelO, from *Clostridium thermocellum*: investigation of the exo-
446 mode of hydrolysis, and binding capacity to crystalline cellulose. *Microbiology* 148:
447 247-255.
- 448 19. Arai T, Ohara H, Karita S, Kimura T, Sakka K, et al. (2001) Sequence of *celQ* and
449 properties of *celQ*, a component of the *Clostridium thermocellum* cellulosome. *Appl*
450 *Microbiol Biotechnol* 57: 660-666.
- 451 20. Zverlov VV, Schantz N, Schwarz WH (2005) A major new component in the cellulosome
452 of *Clostridium thermocellum* is a processive endo-beta-1,4-glucanase producing
453 cellotetraose. *FEMS Microbiol Lett* 249: 353-358.
- 454 21. Morag E, Bayer EA, Hazlewood GP, Gilbert HJ, Lamed R (1993) Cellulase Ss (CelS) is
455 synonymous with the major cellobiohydrolase (subunit S8) from the cellulosome of
456 *Clostridium thermocellum*. *Appl Biochem Biotechnol* 43: 147-151.
- 457 22. Gilad R, Rabinovich L, Yaron S, Bayer EA, Lamed R, et al. (2003) Cell, a
458 noncellulosomal family 9 enzyme from *Clostridium thermocellum*, is a processive
459 endoglucanase that degrades crystalline cellulose. *J Bacteriol* 185: 391-398.
- 460 23. Berger E, Zhang D, Zverlov VV, Schwarz WH (2007) Two noncellulosomal cellulases of
461 *Clostridium thermocellum*, Cel9I and Cel48Y, hydrolyse crystalline cellulose
462 synergistically. *FEMS Microbiol Lett* 268: 194-201.

- 463 24. Schwarz WH, Schimming S, Rucknagel KP, Burgschwaiger S, Kreil G, et al. (1988)
464 Nucleotide sequence of the celC gene encoding endoglucanase C of *Clostridium*
465 *thermocellum*. *Gene* 63: 23-30.
- 466 25. Beguin P, Millet J, Aubert JP (1987) The cloned cel (cellulose degradation) genes of
467 *Clostridium thermocellum* and their products. *Microbiol Sci* 4: 277-280.
- 468 26. Hazlewood GP, Davidson K, Clarke JH, Durrant AJ, Hall J, et al. (1990) Endoglucanase
469 E, produced at high level in *Escherichia coli* as a lacZ' fusion protein, is part of the
470 *Clostridium thermocellum* cellulosome. *Enzyme Microb Technol* 12: 656-662.
- 471 27. Yague E, Beguin P, Aubert JP (1990) Nucleotide sequence and deletion analysis of the
472 cellulase-encoding gene celH of *Clostridium thermocellum*. *Gene* 89: 61-67.
- 473 28. Lemaire M, Beguin P (1993) Nucleotide sequence of the celG gene of *Clostridium*
474 *thermocellum* and characterization of its product, endoglucanase CelG. *J Bacteriol*
475 175: 3353-3360.
- 476 29. Gold ND, Martin VJ (2007) Global view of the *Clostridium thermocellum* cellulosome
477 revealed by quantitative proteomic analysis. *J Bacteriol* 189: 6787-6795.
- 478 30. Brumm P (2013) Bacterial Genomes: What They Teach Us about Cellulose Degradation.
479 *Biofuels*.
- 480 31. Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. NY:
481 Cold Spring Harbor Laboratory Press.
- 482 32. Nelson N (1944) A Photometric Adaptation of the Somogyi Method for the Determination
483 of Glucose. *J Biological Chemistry* 153: 375-380.
- 484 33. Brumm P, Hermanson S, Hochstein B, Boyum J, Hermersmann N, et al. (2010) Mining
485 *Dictyoglomus turgidum* for Enzymatically Active Carbohydrases. *Applied*
486 *Biochemistry and Biotechnology*: 1-10.
- 487 34. Altschul SF, Madden TL, Schaeffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST
488 and PSI-BLAST: A new generation of protein database search programs. *Nucleic*
489 *Acids Research* 25: 3389-3402.
- 490 35. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, et al. (2011) Fast, scalable
491 generation of high-quality protein multiple sequence alignments using Clustal
492 Omega. *Molecular Systems Biology* 7: 539.
- 493 36. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell
494 using TargetP, SignalP, and related tools. *Nature Protocols* 2: 953-971.
- 495 37. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data
496 matrices from protein sequences. *Computer applications in the biosciences* :
497 *CABIOS* 8: 275-282.
- 498 38. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular
499 evolutionary genetics analysis using maximum likelihood, evolutionary distance, and
500 maximum parsimony methods. *Molecular biology and evolution* 28: 2731-2739.
- 501 39. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The
502 carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42:
503 D490-495.
- 504 40. Aspeborg H, Coutinho PM, Wang Y, Brumer H, 3rd, Henrissat B (2012) Evolution,
505 substrate specificity and subfamily classification of glycoside hydrolase family 5
506 (GH5). *BMC Evol Biol* 12: 186.
- 507 41. Abdeev RM, Goldenkova IV, Musiyhuk KA, Piruzian ES (2001) Exploring the properties
508 of thermostable *Clostridium thermocellum* cellulase CelE for the purpose of its
509 expression in plants. *Biochemistry (Mosc)* 66: 808-813.
- 510 42. Zhang YH, Lynd LR (2006) A functionally based model for hydrolysis of cellulose by
511 fungal cellulase. *Biotechnol Bioeng* 94: 888-898.
- 512 43. Reyes-Ortiz V, Heins RA, Cheng G, Kim EY, Vernon BC, et al. (2013) Addition of a
513 carbohydrate-binding module enhances cellulase penetration into cellulose
514 substrates. *Biotechnol Biofuels* 6: 93.

- 515 44. Byrt CS, Cahyanegara R, Grof CP (2012) Plant carbohydrate binding module enhances
516 activity of hybrid microbial cellulase enzyme. *Front Plant Sci* 3: 254.
- 517 45. Bianchetti CM, Brumm P, Smith RW, Dyer K, Hura GL, et al. (2013) Structure,
518 Dynamics, and Specificity of Endoglucanase D from *Clostridium cellulovorans*.
519 *Journal of molecular biology* 425: 4267-4285.
- 520 46. Cai S, Dong X (2010) *Cellulosilyticum ruminicola* gen. nov., sp. nov., isolated from the
521 rumen of yak, and reclassification of *Clostridium lentocellum* as *Cellulosilyticum*
522 *lentocellum* comb. nov. *International Journal of Systematic and Evolutionary*
523 *Microbiology* 60: 845-849.
- 524 47. VanFossen AL, Ozdemir I, Zelin SL, Kelly RM (2011) Glycoside hydrolase inventory
525 drives plant polysaccharide deconstruction by the extremely thermophilic bacterium
526 *Caldicellulosiruptor saccharolyticus*. *Biotechnol Bioeng* 108: 1559-1569.
- 527 48. Blumer-Schuette SE, Lewis DL, Kelly RM (2010) Phylogenetic, microbiological, and
528 glycoside hydrolase diversities within the extremely thermophilic, plant biomass-
529 degrading genus *Caldicellulosiruptor*. *Appl Environ Microbiol* 76: 8084-8092.
- 530 49. Dam P, Kataeva I, Yang SJ, Zhou F, Yin Y, et al. (2011) Insights into plant biomass
531 conversion from the genome of the anaerobic thermophilic bacterium
532 *Caldicellulosiruptor bescii* DSM 6725. *Nucleic Acids Res* 39: 3240-3254.
- 533 50. Yutin N, Galperin MY (2013) A genomic update on clostridial phylogeny: Gram-negative
534 spore formers and other misplaced clostridia. *Environ Microbiol* 15: 2631-2641.
- 535