Transcriptomic Analysis of *Clostridium thermocellum* in Cellulolytic Consortium after Artificial Reconstruction to Enhance Ethanol Production

Hongyu Li,[a,b] Yuanlong Pan,[a,b] Sandra Chang,[b] Ran Du,[a,b] Peipei Li,[a,b] and Shizhong Li [a,b]*

The cellulolytic and ethanologenic bacterial community is a promising candidate for the production of bioethanol from lignocellulose. In this study, by artificially changing the ratio of *Clostridium thermocellum* in the cellulolytic consortium H, ethanol production was increased by 72.7%. Metatranscriptomic analysis was used to elucidate the contribution of *Clostridium thermocellum* to ethanol production. A comprehensive analysis of genes mapped to the *Clostridium thermocellum* ATCC 27405 genome was performed; the identified gene expression differences related to cellulosic ethanol pathways were carefully studied. The results indicated that the majority of genes involved in lignocellulose degradation, sugar transport, cellodextrin breakdown, glycolysis, and ethanol synthesis were up-regulated in *C. thermocellum* when added to H (HCt). More than 18 cellulosome-related genes had 15-fold or greater increased expression. The results illustrate the role of *C. thermocellum* in the cellulolytic consortium H and HCt and provided useful information for identifying genes and preferred pathways. These results will aid in the metabolic and genetic engineering of bacterial strains for more efficient biofuel production.

Keywords: CBP; Bioethanol; Cellulose; Clostridium thermocellum; Transcriptomic

*Contact information: a: Institute of Nuclear and New Energy Technology, Tsinghua University, Tsinghua Garden, Beijing 100084, PR China; b: Beijing Engineering Research Center for Biofuels, Beijing 100084, PR China; *Corresponding author: szli@mail.tsinghua.edu.cn*

INTRODUCTION

Cellulose, the most abundant biomass on earth, has been intensely studied as a renewable resource for bioenergy production (Lynd *et al.* 2005). However, because of the inefficiency and expense issues associated with pretreatment, saccharification, and the fermentation processes, the commercial production of lignocellulosic biofuels remains limited (Lynd *et al.* 2002). Biomass conversion could be achieved through synergistic reactions within bacterial consortia, which provides a variety of enzymes for the degradation of cellulose and hemicellulose from biomass to sugars and the fermentative process from sugars to metabolic products (Zaldivar *et al.* 2001). Most of the screened microbial consortia have outstanding self-stability and the ability to adapt to a broad spectrum of biomass substrates. This indicates that the natural microbial community would be a good candidate for the conversion of biomass (Hui *et al.* 2013). Consolidated bioprocessing (CBP) microbes are becoming the most promising alternative for producing cost-effective biofuel due to their decreased operational expenses and the avoidance of exogenous enzyme supplementation (Lynd 1996; Parisutham *et al.* 2014). However, because natural consortia are complex and only some parts of the component

strains harbor cellulolytic and ethanol-producing enzymes, they tend to be poor ethanol producers (Zuroff and Curtis 2012).

Previous studies have reported that the addition of indigenous or exogenous microorganisms into a screened microbial community can successfully enhance and change the biological activity of the community (Narisawa *et al.* 2007). Du *et al.* (2015) reported the improvement of the ethanol titer by the addition of an indigenous single strain into a natural bacterial consortium.

In our laboratory, a natural consortium capable of effectively degrading a variety of cellulosic materials under anaerobic conditions was developed and several strains from the community were isolated and identified from the community. It was found that *via* the addition of the endogenous strain *Clostridium thermocellum* at a fixed ratio, ethanol productivity was increased by 72%. *Clostridium thermocellum* is one of the most widely studied cellulosome-producing, thermophilic, anaerobic bacteria and has one of the most efficient cellulosome machineries for the degradation of lignocellulose substrates (Hong *et al.* 2014). *C. thermocellum* can also produce ethanol as an end-product, which makes it a promising candidate for consolidated bioprocessing for converting biomass into useful bioenergy (Demain *et al.* 2005). Nevertheless, *C. thermocellum* cannot utilize pentose sugars as carbon hydrates (Lynd *et al.* 2002), and its commercial potential is currently limited by its relatively low ethanol yield.

In this study, bioethanol production of a consortium H was improved significantly by the artificial addition of *C. thermocellum*. To create rational strategies for the further improvement of ethanol yield, it is important to elucidate the transcriptional changes that occurred in *C. thermocellum* when its ratio in the consortium was adjusted to enhance production. Metatranscriptomics (also known as RNA-seq) provide useful insights into changes in gene expression in a community by retrieving and sequencing the total RNA from a microbial ecosystem (Jung *et al.* 2013). Metatranscripomic sequencing was applied to consortium H and HCt (consortium H with the additional *C. thermocellum*) and the analysis of the sequencing data were focused on transcriptional changes of *C. thermocellum* to elucidate the contribution of *C. thermocellum* to the ethanol production. These results will provide useful information for identifying genes and preferred pathways to aid in the metabolic and genetic engineering of strains for more efficient biofuel production.

EXPERIMENTAL

Microbial Source

Consortium H was screened from steppe soils in Inner Mongolia and cultured to a stable condition in our laboratory, which could degrade the filter paper at a degradation ratio of 93% (Du *et al.* 2015). The strain *Clostridium thermocellum* was isolated from H and identified by 16S sequencing.

Co-culture of the Consortium with *C. thermocellum*

Consortium H and *C. thermocellum* were cultured separately at 55 °C in anaerobic conditions until reaching the stationary phase. Duplicated 40-mL fermentation cultures of the consortium and *C. thermocellum* in 50-mL centrifuge tubes were centrifuged at 14,000 x *g* for 5 min, and the supernatant was discarded. The cell pellets were washed with sterile water three times.

The washed pellets were dried in a drying oven to a constant weight, and the biomass was measured. Then consortium H was co-fermented with *C. thermocellum* at a ratio of 17:1 biomass which was the optimal value in the preliminary experiment. The newly combined consortium HCt and H were separately inoculated in 30 mL of PCS medium (1 g yeast extract, 5 g peptone, 5 g NaCl, 2.5 g $CaCO_3$, 0.5 mg $ZnSO_4$, 0.05 mg $CuSO_4$, 0.05 mg $CoSO_4$, 0.05 mg $Na_2B_4O_7$, and 0.05 mg $NaMoO_4$ per liter) at 10%, with 0.5% (m/v) filter paper as the sole carbon source, as previously described (Du *et al.* 2015).

Ethanol Concentration and Filter Paper Assay

Duplicate H and HCt samples were collected and centrifuged at 14,000 x g for 10 min, and the supernatant was filtered through a 0.45-μm filter. Subsequently, the ethanol concentration was measured by HPLC using an Aminex HPX-87H column (Bio-Rad, Hercules, CA, USA), as previously described (Du *et al.* 2015).

Duplicate H and HCt samples were collected as the crude enzyme. The filter paper activity of the crude enzyme was measured using a Whatman Grade 1 filter paper (GE Healthcare, Shanghai, China) as a substrate, and 3,5-dinitrosalicylic acid (DNS) was used to measure the reducing sugars, as described by Du *et al.* (2015).

Total RNA Isolation

The peak of filter paper degradation activity (FPA) was day three, and the peak of the ethanol production was day five, for both H and HCt. The cultures of consortium H on day three and day five (named H-3 and H-5 hereinafter) and consortium HCt on day three and day five (named HCt-3 and HCt-5 hereinafter) were harvested for total RNA extraction. Ten-milliliter samples were centrifuged at 4 °C for 5 min at 8,000 x *g*. The supernatant was discarded.

The total RNA of the cell pellet was isolated using RNeasy Plus Universal Mini Kits (Qiagen, Germany) with the enhanced beads beating lysis process. Briefly, the cell pellet in 900 μL Qiazol reagent was added to a 2-mL Lysing Matrix E tube (MP Biomedicals, USA). The cells were lysed through high-frequency agitation at 4 m/s in two 15-s on/5-s off cycles using a FASTPREP-24 bead beater (MP Biomedicals, USA).

Total RNA Quality Control and rRNA Subtraction

The quality of the total RNA was analyzed for integrity by electrophoresis, Nanodrop (Thermo, USA), and a 2100 Bioanalyzer (Agilent, USA). High-quality samples with RIN values greater than six were treated with DNase I (Takara, Japan) for 30 min at 37 °C. The samples were purified with an RNeasy Cleanup Kit (Qiagen, German) and Ribo-ZeroTM rRNA Removal Kits (Bacteria) (Epicentre, USA) to deplete ribosomal RNA. All operations were performed following the manufacturer instructions.

The Library Construction of cDNA, Sequencing, and Reads Mapping

cDNA libraries were constructed using the RNA-Seq Library Preparation Kit for whole transcriptome discovery (Gnomegen, USA), following the manufacturer instructions. The prepared cDNA was analyzed by a high-sensitivity DNA chip to guarantee cDNA library quality. Resultant cDNA libraries were pair-end sequenced in Illumina HiseqTM 2500.

Sequenced reads were aligned with the closest reference genome sequence of *C. thermocellum* strain ATCC 27405, which was deposited at NCBI (Accession: NC_009012.1) using Bowtie (Langmead *et al.* 2009). According to the NCBI, the genome was 3.8 Mb, with 3173 coding DNA sequences. The best local mode alignment was used with no more than two mismatches. The reads that were not mapped unambiguously were excluded from further analysis, as described by Wei *et al.* (2014). The metatranscriptome sequencing data are publicly available in the NCBI Sequence Read Archive under BioProject ID PRJNA280031, and the 16S rRNA gene sequencing data are publicly available under the GenBank accession number SUB883074 Seq1 KR047885.

Normalization and Identification of Genes Differentially Expressed

After read mapping, the RNA-seq read counts belonging to each gene were calculated. To compare the gene expression level within and between two samples, RPKM (the reads per kilobase of exon model per million mapped reads) was used to quantify and normalize the gene expression levels, as reported by Mortazavi *et al.* (2008).

$$RPKM=10^9 \times \text{Reads number mapped to gene/length of the gene (kb)/total reads number (million reads).}$$

RPKM is a widely recognized indicator of gene expression level, especially as a measure of Solexa read density, which indicates the molar concentration of a transcript (Wei *et al.* 2014). The fold-change for an expressed gene was calculated by DEGseq MARS (MA-plot-based method with random sampling model). The criteria for identifying a gene as "differentially expressed" is the FDR value, which should be less than 0.001.

Biological Analysis of Differentially Expressed Genes

According to the results of the analysis, the differentially expressed genes were classified based on the clusters of orthologous groups (COG) and Carbohydrate Active Enzyme (CAZy) analyses (www.caz.org). GO cluster was analyzed by Gopipe.

Predicted proteins were mapped with Swiss-Prot and TrEMBL database first at the criterion blastp, E value < $1x10^{-5}$. The mapped results were then analyzed in GoPipe by gene2go.

RESULTS AND DISCUSSION

Enzyme Activity and Ethanol Production during Fermentation

In this study, the single strain *C. thermocellum* enhanced the FPA and ethanol production of the consortia H. As can be seen in Fig. 1, the addition of the *C. thermocellum* strain to H (HCt) increased the peak of the FPA on day three by 14.9% and the peak of ethanol production on day five by 72.7%.

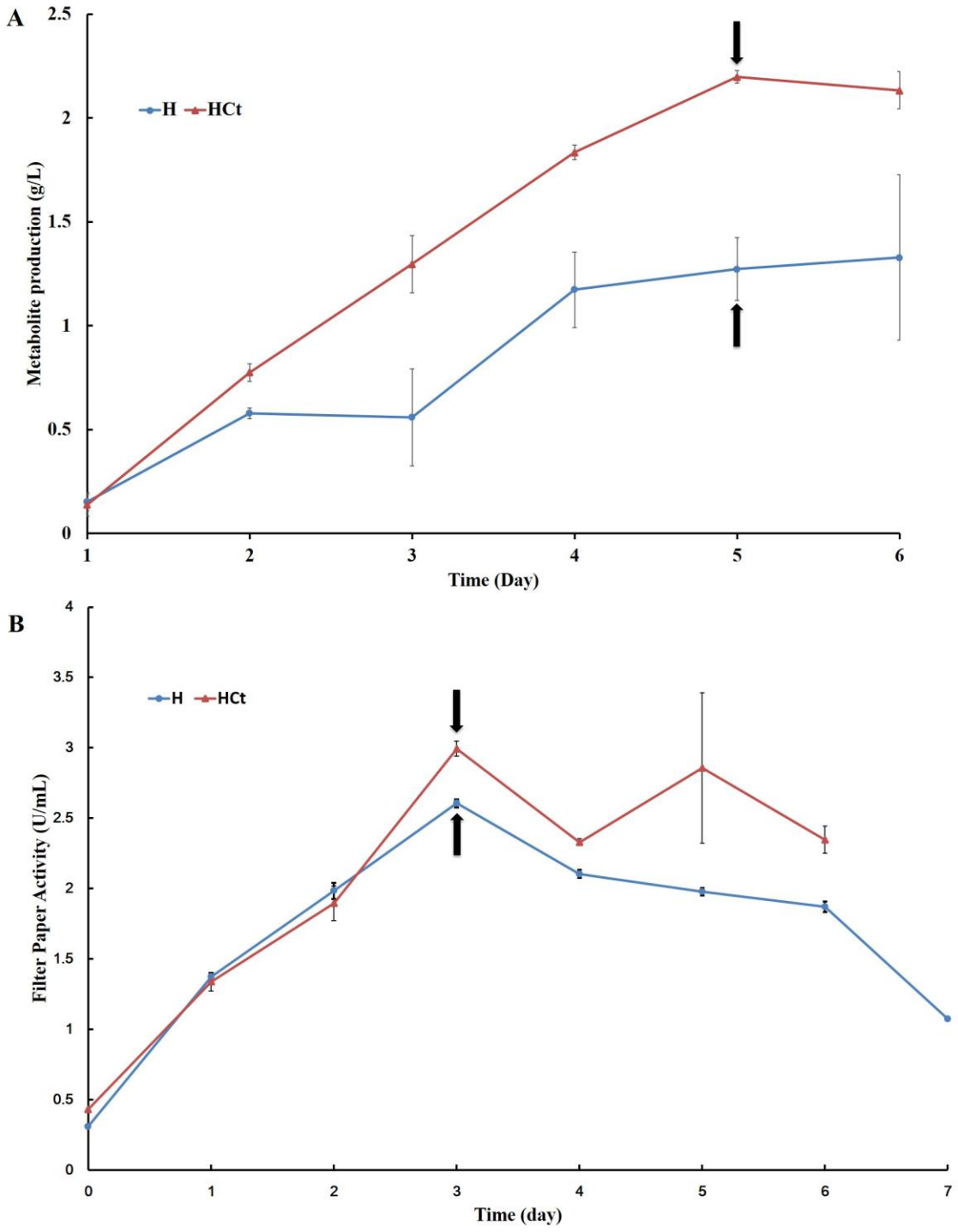
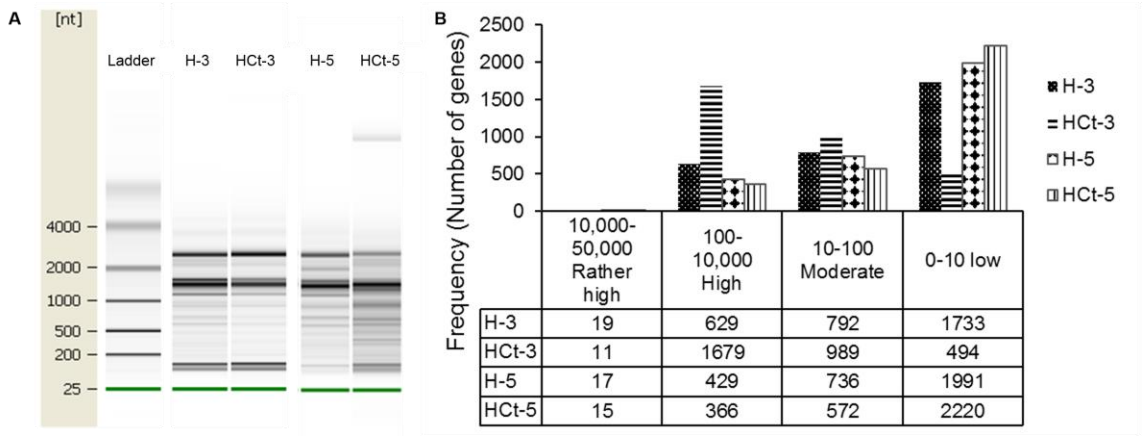


Fig. 1. Filter paper activity and ethanol production of consortia H and HCt. A) Ethanol production during the course of fermentation. Circular and triangle symbols represent consortia H and HCt, respectively. B) Filter paper degrading enzyme activity of consortia H and HCt. Circular and triangular symbols represent consortia H and HCt, respectively. Arrows indicate sampling points for transcriptomic analysis. Error bars are the standard deviation of two replicate fermentations, and measuring methods are described in previous paragraphs.

Total RNA Quality Control and Sequencing Results

The isolated total RNA was assessed on a 2100 Bioanalyzer to guarantee the quality before cDNA library construction, and the analysis results are shown in Fig. 2(a).



	10,000-50,000 Rather high	100-10,000 High	10-100 Moderate	0-10 low
H-3	19	629	792	1733
HCt-3	11	1679	989	494
H-5	17	429	736	1991
HCt-5	15	366	572	2220

Fig. 2. A) Electrophoresis run of total RNA in H-3, HCt-3, H-5, and HCt-5, respectively, using an Agilent 2100 Bioanalyzer with clear 16 S and 23 S bands. B) RPKM frequency histogram of transcripts from fermentation day three and day five in consortia H and HCt. The distribution of number of genes expressed at different RPKM levels is shown in the diagram. The total number of protein coding genes is 3173; the value below each bar indicates the number of genes at specific expression levels in different samples.

The filtered metatranscriptomic reads in four samples were mapped to *C. thermocellum* ATCC 27405 genomes, respectively. For the four cDNA consortium metatranscriptomic samples, 0.07, 1.1, 0.05, and 0.03 million reads were unambiguously mapped to the *C. thermocellum* ATCC27405 genome. The basic results of the gene mapping process are shown in Table 1.

Table 1. Results of Reads Mapping to the *C. thermocellum* ATCC 27405 genome

	H-3	HCt-3	H-5	HCt-5
Average reads length (bp)	101	101	101	101
Mapped RNA reads (million)	0.07	1.10	0.05	0.03
Mapped gene number	1529	2966	1258	965
Mapped Ratio (total 3173)	0.48	0.93	0.40	0.30

The Solexa reads were aligned with the 3173 protein coding genes at the highest ratio of 0.93 in HCt-3. This value was similar to the results from culturing single strain *C. thermocellum* ATCC 27405 in cellobiose and pretreated yellow poplar, with 95% of the genes in *C. thermocellum* transcriptionally active and expressed (Wei *et al.* 2014). The quantification of a gene is represented by the RPKM value, as defined in the Experimental section.

The level of gene expression was classified into four frequency levels: very high, high, moderate, and low. As shown in Fig. 2, after the addition of *C. thermocellum*, high-level expressed genes increased markedly on day three, but were almost unchanged on day five, with a few genes decreasing to much lower levels.

The addition of *C. thermocellum* noticeably altered the level of gene expression during the early fermentation process.

Changes in Global Gene Expression and COG (Clusters of Orthologous Groups) Analysis

The gene expression levels between the original H consortium and the *C. thermocellum*–added consortium on day three (*i.e.*, when the peak point for the FPA occurred) and five (*i.e.*, the peak point for ethanol production) were compared. On days three and five, 129 and 1559 genes were identified as statistically differentially expressed genes, respectively. COG distribution analysis was performed as shown in Fig. 3.

On day three, most of the genes (1367 out of 1559) were up-regulated after the addition of *C. thermocellum.* The functional categories that are essential for the fermentation process related to sugar degradation, such as "amino acid transport and metabolism [T]," "cell envelope biogenesis, outer membrane [Q]," "Carbohydrate transport and metabolism [S]," and "energy production and conversion [L]," were all up-regulated. Half of the genes in the "DNA replication, recombination, and repair [M]" were down-regulated. This possibly pointed to the fact that *C. thermocellum* focused most of the transcriptional activities towards biofuel production and ceased much of the activities for cell expansion.

On the fifth day, the changes between the original consortium and the addition of *C. thermocellum* were not as remarkable as on the third day. Only 128 genes showed a statistically significant difference and most of the genes were down-regulated, again especially in the category of "DNA replication, recombination, and repair [M]."

Transcriptional Changes in Cellulosomal Components and Non-cellulosomal Carbohydrate Active Enzymes

Cellulosome is an important enzyme complex in the fermentation process. It is a macromolecular complex composed of several kinds of organized lignocellulose-degrading enzymes. Furthermore, cellulosome is associated with the cell surface by non-catalytic domains (Hong *et al.* 2014). It was of particular interest to examine the changes of cellulosomal component protein encoding genes in *C. thermocellum* in response to addition to the consortium.

Figure 3 shows that the category "Carbohydrate transport and metabolism [S]," including cellulosomal genes, was up-regulated on day three. There were 81 reported cellulosomal genes in the genome of *C. thermocellum* ATCC 27405, and 79 genes were expressed in the HCt-3 metatranscriptome, while only 61 genes were detected transcriptionally in H-3. Compared with H-3, 58 cellulosomal genes (73%) in HCt-3 were up-regulated and only two dockerin type I protein coding genes were down-regulated (Table 2). In the H-5 and HCt-5 metatranscriptomes, 44 and 35 cellulosomal genes were detected transcriptionally, respectively.

The addition of *C. thermocellum* to the original consortium increased the gene expression of enzymes in *C. thermocellum* that were important for lignocellulose degradation. This was in agreement with the observed increase in FPA on days three and five.

Fig. 3. COG functional category of the genes. A) The expression difference in third-day fermentation sample in H and HCt; B) the expression difference in fifth-day fermentation samples in H and HCt

The cellulosomal system of *C. thermocellum* is divided into cell-bound scaffoldin CBS and cell-free scaffoldin CFS subsystems (Blumer-Schuette *et al.* 2014). CBS are cellulosomes that anchor to the cell wall *via* different anchor proteins and S-layer proteins, while CFS represents two newly identified cell-free cellulosomes with scaffoldins, Cthe_0736 and Cthe_0735, containing seven and one Cohesin II modules, respectively (Raman *et al.* 2009). The simplified schematic of the *C. thermocellum* cellulosomal structure is presented in Fig. 4. On day three after *C. thermocellum* was added, the three CBS S-layer homology domain-containing proteins (containing different

numbers of Cohesin II modules to bind individual Cip scaffoldin *via* C-terminal from Docrin II module, SdbA (Cthe_1307) (Leibovitz *et al.* 1997), Orf2p (Cthe_3079) (Fujino *et al.* 1993), and OlpB (Cthe_3078) (Lemaire *et al.* 1995), increased 6.4-, 6.6-, and 3.8-log fold, respectively. Two other CBS, OlpA (Cthe_3080) (Salamitou *et al.* 1994) and OlpC (Cthe_0452) (Pinheiro *et al.* 2009), which interact with dockrin-containing catalytic units directly, also increased 4.6- and 3.8-fold, respectively. The two cell-free cellulosomes Cthe_0736 and Cthe_0735 both increased 4.9-fold.

Table 2. List of Fold Changes of Core Lignocellulose Degrading-Related Genes

	Genome Loci	Protein Name	HCt/H-3 Log2 (Fold change)	HCt/H-5 Log2 (Fold change)
Structure Module	Cthe3079	Orf2p	6.6	1.9
	Cthe1307	SdbA	6.4	NA
	Cthe0735		4.9	NA
	Cthe0736		4.9	NA
	Cthe3080	OlpA	4.6	0.9
	Cthe3077	CipA	4.1	0.1
	Cthe0452		3.8	NA
	Cthe3078	OlpB	3.8	1.7
	Cthe0044	CseP	2.8	NA
Glucanases	Cthe0412	CelK	1.8	−0.4
	Cthe0413	CbhA	2.3	−0.1
	Cthe2089	CelS	0.5	0.4
	Cthe2147	CelO	2.2	NA
	Cthe0578	CelR	1.0	0.9
	Cthe0043	CelN	1.6	−1.7
	Cthe0211	LicB	0.6	NA
	Cthe0269	CelA	4.6	NA
	Cthe0274	CelP	5.3	NA
	Cthe0405	CelL	2.1	NA
	Cthe0433		2.3	NA
	Cthe0536	CelB	1.3	NA
	Cthe0543	CelF	2.2	1.1
	Cthe0624	CelJ	3.4	1.9
	Cthe0625	CelQ	2.1	0.9
	Cthe0745	CelW	3.3	0.9
	Cthe0825	CelD	3.2	NA
	Cthe1472	CelH	1.8	1.4
	Cthe2360	CelU	1.8	NA
	Cthe2760	CelV	1.1	2.4
	Cthe2761		2.8	NA
	Cthe2812	CelT	4.5	NA
	Cthe2872	CelG	1.3	NA
	Cthe0797	CelE,CtCE2	4.1	NA
	Cthe0660		1.1	NA
	Cthe1235		8.4	1.4
	Cthe0040		1.8	NA
	Cthe0071		1.2	NA
Xylanase	Cthe1963	XynZ	3.3	NA
	Cthe2972	XynA/U	2.9	−1.6
	Cthe0912	XynY	2.8	−2.5
	Cthe2137		2.1	0.9
	Cthe2590	XynD	1.4	NA

	Cthe1398	Xgh74A	1.0	1.9
	Cthe2193	CtXyl5A	0.6	0.9
	Cthe1838	XynC	(0.6)	1.8
Others	Cthe2179		7.1	−3.3
	Cthe1806		7.0	NA
	Cthe0270	ChiA	5.9	NA
	Cthe2038		5.7	0.9
	Cthe1890		5.5	NA
	Cthe2811	ManA	5.3	0.9
	Cthe3132		4.9	NA
	Cthe0661	Ct1, 3Gal43A	4.7	0.9
	Cthe0015		4.0	NA
	Cthe0640		3.7	NA
	Cthe3012		3.4	NA
	Cthe0729		3.0	NA
	Cthe3141		3.0	NA
	Cthe1400		2.8	NA
	Cthe2950		2.5	NA
	Cthe0798		2.1	NA
	Cthe2879		2.1	NA
	Cthe2138		2.0	NA
	Cthe0109		1.5	NA
	Cthe0821	CtMan5A	1.2	NA
	Cthe1271		0.9	0.9
	Cthe2194		0.9	NA
	Cthe0246		0.8	3.2
	Cthe0435		0.6	0.9
	Cthe0032		0.4	NA
	Cthe0032		0.4	NA
	Cthe2139		−0.3	−2.5
	Cthe0438		−0.6	NA
	Cthe2195		−1.0	NA
	Cthe2949		−1.6	NA
	Cthe2549		−2.1	NA
	Cthe2196		−2.4	NA
	Cthe2271		−5.6	2.7

* Negative values indicate a decrease in expression level compared HCt to H, and NA indicates that no expression was detected in H or HCt.

The primary scaffoldin CipA (Cthe_3077) is the most critical structural component in the cellulosome of *C. thermocellum*. On the one hand, CipA interacts with the cohesin components, which anchor the other cellulosome components to the cell wall *via* an S-layer homology (Lamed and Bayer 1988). On the other hand, CipA interacts with the type I dockerin modules of a variety of cellulosomal enzymes *via* the type I cohesin modules, which enhances lignocellulose degradation through the synergistic effects of the multiple enzymes' proximity to each other (Tokatlidis *et al.* 1991; Moraïs *et al*. 2010). The expression of CipA increased 4.1-fold with the addition of *C. thermocellum* on day three.

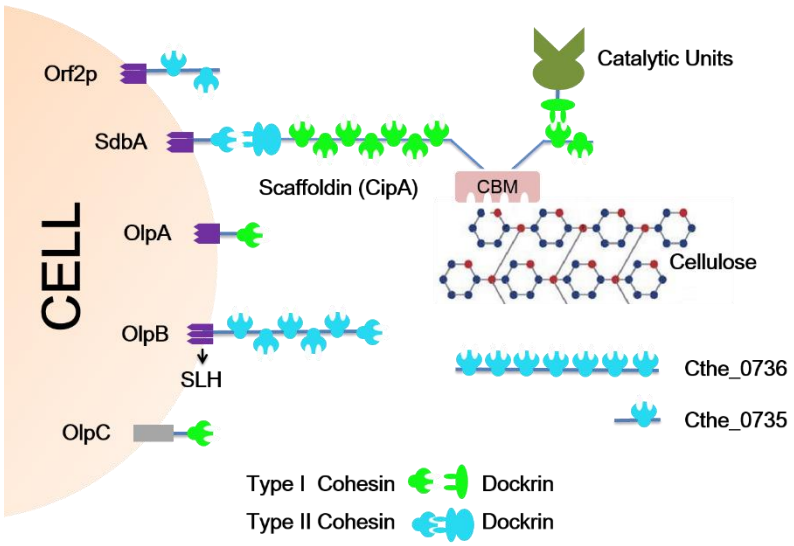


Fig. 4. Simplified schematic representation of *C. thermocellum* cellulosomal structure

CelA (Cthe_0269) is the most spectrally abundant cellulosomal endoglucanase during cell growth on various substrates (Lynd *et al.* 2002; Raman *et al.* 2009). CelA increased 4.6-fold on day three. Two other endoglucanases, CelP (Cthe_0274) and CelT (Cthe_2812), increased 5.3- and 4.5-fold, respectively.

The exoglucanase CelS (Cthe_2089) is one of the most highly expressed cellulosome proteins and was the most abundant cellulosomal-related protein expressed in both H and HCt on day three (Table 2). The addition of *C. thermocellum* did not greatly increase its CelS expression.

Cellulose 1,4-beta-cellobiosidase releases cellobiose from the non-reducing ends of the cellulose chains. The identified cellulose 1,4-beta-cellobiosidase, Cthe_1235, increased 8.4-fold; this was one of the most significant changes in expression between H and HCt on day three. The other two cellulose 1,4-beta-cellobiosidase coding genes, Cthe_0040 and Cthe_0071, increased by 1.8 and 1.2 times, respectively.

Other lignocellulose-degrading-related enzymes, such as xylanases and pectinases, were all expressed at high levels (XynZ, Cthe_1963; XynA/U, Cthe_2972; Pectate lyase, Cthe_2179; Chitinase, Cthe_0270; and mannanase, Cthe_2811) and increased by 3.3, 2.9, 7.1, 5.9, and 5.3-fold, respectively, on day three with the addition of *C. thermocellum*. None of these genes or those mentioned earlier were up-regulated more than 2-fold on day five, which further suggests that the essential genes required for sugar breakdown were expressed during the early phase of the fermentation process.

In general, the expression levels of cellulosomal glycoside hydrolase (GH) genes and the structural modules in HCt were much higher than those in H on day three. The known number of dockerin-bearing enzymes in *C. thermocellum* was reported to be approximately eight times higher than the number of cohesins in the scaffoldin subunit (Shoham *et al.* 1999). A few reports have demonstrated that several cellulosomal catalytic units containing dockerin I modules can bind to cohesin I modules from other *Clostridiums* (Jindou *et al.* 2004; Pinheiro *et al*. 2009; Sakka *et al*. 2009). This may also contribute to the integral increase in FPA.

There are two serine protease inhibitors that protect free cellulosomes from proteolytic cleavage. When the cellulosomes were released from the cell surface, the expressions of protease inhibitor component coding genes increased accordingly (Tracy *et al.* 2012). When *C. thermocellum* was added on day three, the serine protease inhibitor genes, Cthe 0190 and 0191 increased 5.9- and 6.6-fold, respectively. The more stable environment provided by excess serine protease inhibitors for the secreted cellulosomes would be beneficial to the final fermentation yield.

The addition of *C. thermocellum* to H had a significant impact on cellulosomal gene expression on day three. Genes in all aspects of the cellulosome, such as Cohesin II modules, scaffoldin, the enzymes themselves, as well as other accessory genes were up-regulated when *C. thermocellum* was added.

Transcriptional Changes of Cellodextrin Transport-related Genes

Another important group of genes in the "Carbohydrate transport and metabolism [S]" category are cellodextrin transport-related genes. *C. thermocellum* has been reported to use an ABC-type system for transporting cellulose-hydrolyzed oligosaccharides (Strobel *et al.* 1995). ABC-type is an essential energy-conserving mechanism for importing long cellodextrins into cells and can reduce the energy required for transport. Only one mole of ATP is needed for each transport event regardless of the length of the cellodextrin chain (Nataf *et al.* 2009).

Table 3. List of Fold Changes of ABC Transporter System-related Genes

Genome Loci	Gene Description	HCt/H-3 Log2 (Fold change)	HCt/H-5 Log2 (Fold change)
Cthe0391	ABC transporter-like protein	1.4	NA
Cthe0392	inner-membrane translocator	2.4	NA
Cthe0393	sugar ABC transporter sugar-binding protein	2.2	NA
Cthe1019	binding-protein-dependent transport systems inner membrane component	3.5	−1.1
Cthe1020	extracellular solute-binding protein	1.5	0.9
Cthe1862	ABC transporter-like protein	−0.1	−2.2
Cthe2125	binding-protein-dependent transport systems inner membrane component	2.1	NA
Cthe2126	binding-protein-dependent transport systems inner membrane component	4.1	NA
Cthe2128	extracellular solute-binding protein	−0.3	NA
Cthe2446	ABC-type sugar transport system periplasmic component-like protein	5.7	0.9
Cthe2447	ABC transporter-like protein	5.6	NA
Cthe2448	inner-membrane translocator	3.5	NA

* Negative values indicate a decrease in expression level compared HCt to H, and positive values indicate an increase.

A previous study identified and characterized four ABC sugar binding proteins in *C. thermocellum* (CbpA, Cthe0393; CbpB, Cthe1020; CbpC, Cthe2128; and CbpD, Cthe2446) based on their substrate binding features (Nataf *et al.* 2009). According to Raman *et al.* (2009), continuous regions of Cthe_0391-0393 encoded CbpA with binding affinities to G3 beta-1,4-glycans, and continuous regions of Cthe_1018-1020 encoded CbpB with binding affinities to G2-G5 beta-1,4-glycans. CbpA and CbpB expression levels increased 2.2- and 1.5-fold, respectively, with the addition of *C. thermocellum* on

day three. CbpC and CbpD had a binding affinity with G3-G5 cellodextrins (Rydzak *et al*. 2011). CbpC did not exhibit any change in expression. However, CbpD significantly increased by 5.7-fold on day three.

Although the change in the expression of CbpB was small, the expression levels on day three and on day five with or without the addition of *C. thermocellum* were at a high level. This result was consistent with a previous report observing that *C. thermocellum* preferentially imported 4-glucose-unit chains when growing on cellulose (Zhang and Lynd 2005).

Changes in the gene expressions of several cellodextrin transporter genes with the addition of *C. thermocellum* would greatly contribute to the fermentation process.

Transcriptional Changes of Ethanol Producing-related Genes

As shown in Fig. 1, the addition of *C. thermocellum* to H resulted in an increase in ethanol production to a peak level on day five. Figure 5 illustrates the key steps involved in the conversion from cellodextrin to the end-fermentation product ethanol. The key genes involved in cellodextrin phosphorolysis, glycolysis, and end-product synthesis expressed during the course of fermentation were generally up-regulated on day three with the addition of *C. thermocellum*.

For the phosphorolytic cleavage of cellodextrins, three genes are involved: cellodextrin phosphorylase (Cthe_2989) and cellobiose phosphorylases (Cthe_0275 and Cthe_1221) (Wei *et al.* 2014). The expression level of cellodextrin phosphorylase increased 5-fold, while cellobiose phosphorylase Cthe_1221 increased 3.3-fold and Cthe_0275 remained the same. The increased expression of cellodextrin phosphorylases may be related to the previous results indicating that *C. thermocellum* preferentially imported 4-glucose-unit chains and thus, required enzymes to hydrolyze the multi-unit sugars. The addition of *C. thermocellum* also up-regulated the expression of one of the cellobiose phosphorylases, which likely increased the breakdown of carbon sources to glucose, thereby providing more substrates for ethanol fermentation.

C. thermocellum utilizes the Embden-Meyerhof-Parnas pathway to convert glucose to phosphoenolpyruvate (PEP) (Rydzak *et al.* 2012). Most of the essential genes involved in the pathway increased in expression on day three after the addition of *C. thermocellum*. Glucokinase (Cthe_0390), phosphofructokinase (Cthe_1261), and five of the phosphoglycerate mutases (Cthe_0707, Cthe_0946, Cthe_1435, Cthe_2449, and Cthe_1292) were all up-regulated more than 2-fold.

Pyruvate can be converted from PEP *via* three pathways: (i) conversion by pyruvate phosphate dikinase (Cthe_1308); (ii) conversion by PEP synthase; and (iii) conversion to oxaloacetate *via* PEP carboxykinase, followed by a direct conversion by oxaloacetate decarboxylase to pyruvate or indirectly by malate dehydrogenase to malate, and finally by malic enzyme to pyruvate. The expression level of PEP synthase (Cthe_1253) increased 2.7-fold when *C. thermocellum* was added on day three. Interestingly, pyruvate phosphate dikinase (Cthe_1308) and PEP carboxykinase (Cthe_2874) down-regulated during the same period. However, the expression levels of malate dehydrogenase (Cthe_0345) and malic enzyme (Cthe_0344) were up-regulated 3.9- and 4.1-fold, respectively. The opposing differential expressions of genes along the same pathway may reflect the influence of the consortium on the gene expressions of *C. thermocellum*.

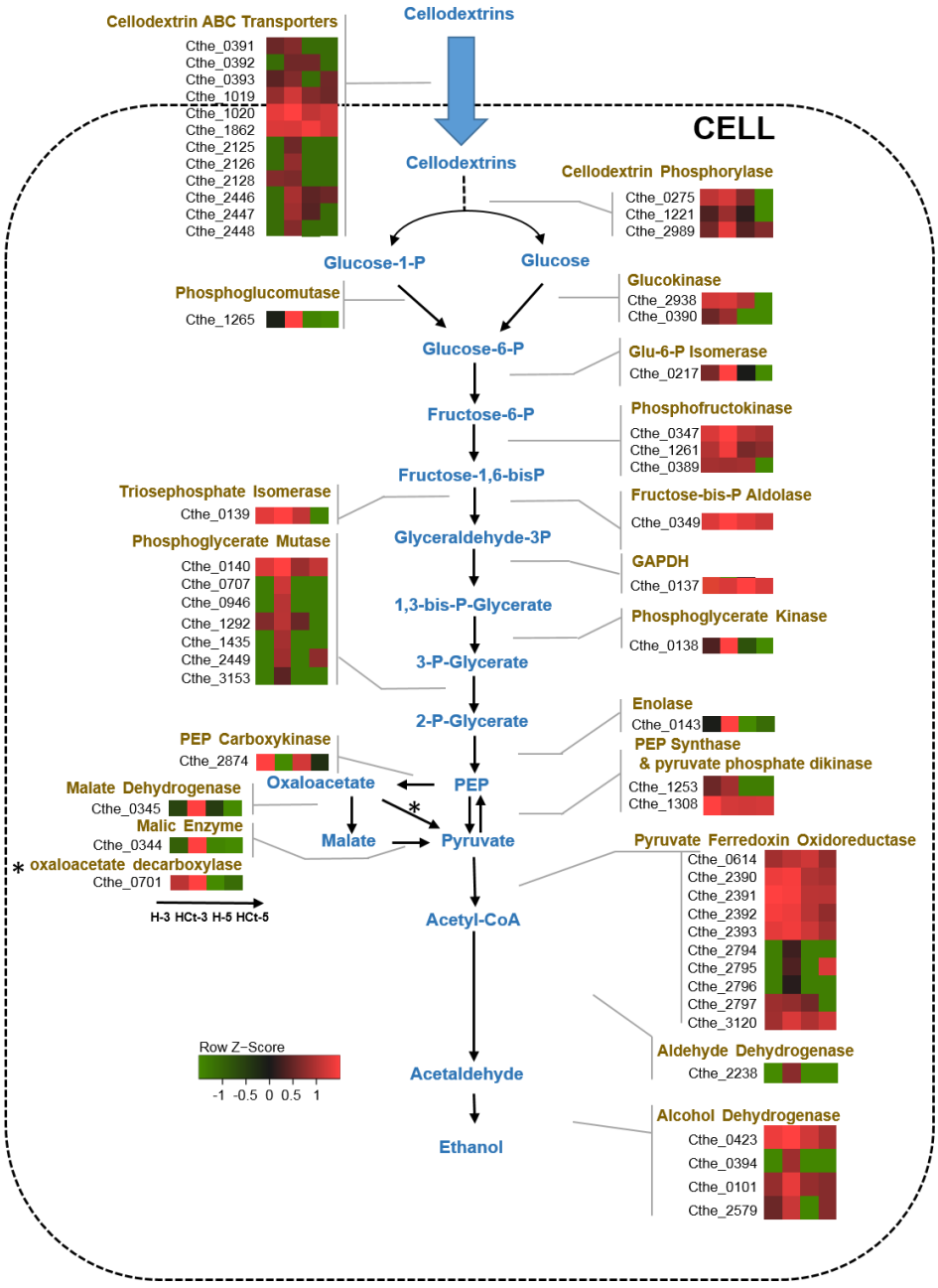


Fig. 5. Diagram of pathways in the conversion from cellodextrins to fermentation products in *C. thermocellum*

Synthesis pathways from pyruvate to ethanol occur through acetyl-CoA. Pyruvate ferredoxin oxidoreductase cluster genes (Cthe_2390-2393) were highly expressed, which is consistent with the observations made in a proteomic study by Rydzak *et al.* (2012).

The conversion from acetyl-CoA to ethanol is first catalyzed by the NADH-dependent aldehyde dehydrogenase (Cthe_2238, *aldH*) and then by three iron-containing alcohol dehydrogenases (Cthe_0101; Cthe_0394, *adhY*; Cthe_2579, *adhZ*) or by the bi-functional acetaldehyde/alcohol dehydrogenase (Cthe_0423, *adhE*) directly. AldH expression was up-regulated 2.5-fold with the addition of *C. thermocellum* on day three. AdhE expression was already high without the addition *C. thermocellum* and only

increased slightly afterward. The three other alcohol dehydrogenases were up-regulated 3- to 4-fold. All gene expressions decreased by day five. The addition of *C. thermocellum* increased the majority of the genes involved in the glycolysis and ethanol synthesis pathways.

Overall, the identification of genes and preferred end-product pathways could aid in the metabolic and genetic engineering of the strain for more efficient biofuel production and provide more information for microbial syntheses in the future.

CONCLUSIONS

1. The addition of *C. thermocellum* to the consortium H enhanced FPA and ethanol production. The ethanol production increased by 72.7%, and FPA increased by 14.9%.
2. Metatranscriptomic sequencing was carried out on the consortium H and HCt on days three and five, which were the peak days of FPA and ethanol production, respectively. Genes likely responsible for the improved activities were identified. Genes involved in lignocellulose degradation, sugar transport, cellodextrin breakdown, glycolysis, and ethanol synthesis were nearly all up-regulated in *C. thermocellum* when added to H. This wide spectrum of gene up-regulation may be the result of interaction with other species in the consortium and contribute to the increase in ethanol production and FPA.

ACKNOWLEDGMENTS

This work was financially supported by grants from the Ministry of Science and Technology, China (the National High Technology Research and Development Program 2012AA101805, the International S&T Cooperation Program 2013DFA60470, and 2012DFG61700), and supported by Tsinghua University Initiative Scientific Research Program 2012Z08127.

REFERENCES CITED

Blumer-Schuette, S. E., Brown, S. D., Sander, K. B., Bayer, E. A., Kataeva, I., Zurawski, J. V., Conway, J. M., Adams, M. W., and Kelly, R. M. (2014). "Thermophilic lignocellulose deconstruction," *FEMS Microbiology Reviews* 38(3), 393-448. DOI: 10.1111/1574-6976.12044

Demain, A. L., Newcomb, M., and Wu, J. D. (2005). "Cellulase, clostridia, and ethanol," *Microbiology and Molecular Biology Reviews* 69(1), 124-154. DOI: 10.1128/MMBR.69.1.124-154.2005

Du, R., Yan, J., Li, S., Zhang, L., Zhang, S., Li, J., Zhao, G., and Qi, P. (2015). "Cellulosic ethanol production by natural bacterial consortia is enhanced by *Pseudoxanthomonas taiwanensis*," *Biotechnology for Biofuels* 8(1), 10. DOI: 10.1186/s13068-014-0186-7

Fujino, T., Béguin, P., and Aubert, J. (1993). "Organization of a *Clostridium thermocellum* gene cluster encoding the cellulosomal scaffolding protein CipA and a protein possibly involved in attachment of the cellulosome to the cell surface," *Journal of Bacteriology* 175(7), 1891-1899.

Hong, W., Zhang, J., Feng, Y., Mohr, G., Lambowitz, A. M., Cui, G.-Z., Liu, Y.-J., and Cui, Q. (2014). "The contribution of cellulosomal scaffoldins to cellulose hydrolysis by *Clostridium thermocellum* analyzed by using thermotargetrons," *Biotechnology for Biofules* 7(1), 1-17. DOI: 10.1186/1754-6834-7-80

Hui, W., Jiajia, L., Yucai, L., Peng, G., Xiaofen, W., Kazuhiro, M., and Zongjun, C. (2013). "Bioconversion of un-pretreated lignocellulosic materials by a microbial consortium XDC-2," *Bioresource Technology* 136, 481-487. DOI: 10.1016/j.biortech.2013.03.015

Jindou, S., Soda, A., Karita, S., Kajino, T., Béguin, P., Wu, J. D., Inagaki, M., Kimura, T., Sakka, K., and Ohmiya, K. (2004). "Cohesin-Dockerin Interactions within and between *Clostridium josui* and *Clostridium thermocellum* binding selectivity between cognate dockerin and cohesin domains and species specificity," *Journal of Biological Chemistry* 279(11), 9867-9874. DOI: 10.1074/jbc.M308673200

Jung, J. Y., Lee, S. H., Jin, H. M., Hahn, Y., Madsen, E. L., and Jeon, C. O. (2013). "Metatranscriptomic analysis of lactic acid bacterial gene expression during kimchi fermentation," *International Journal of Food Microbiology* 163(2–3), 171-179. DOI: 10.1016/j.ijfoodmicro.2013.02.022

Lamed, R., and Bayer, E. A. (1988). "The cellulosome of *Clostridium thermocellum*," *Advances in Applied Microbiology* 33, 1-46. DOI: 10.1016/S0065-2164(08)70203-X

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome," *Genome Biology* 10(3), R25. DOI: 10.1186/gb-2009-10-3-r25

Leibovitz, E., Ohayon, H., Gounon, P., and Beguin, P. (1997). "Characterization and subcellular localization of the *Clostridium thermocellum* scaffoldin dockerin binding protein SdbA," *Journal of Bacteriology* 179(8), 2519-2523.

Lemaire, M., Ohayon, H., Gounon, P., Fujino, T., and Beguin, P. (1995). "OlpB, a new outer layer protein of *Clostridium thermocellum*, and binding of its S-layer-like domains to components of the cell envelope," *Journal of Bacteriology* 177(9), 2451-2459.

Lynd, L. R. (1996). "Overview and evaluation of fuel ethanol from cellulosic biomass: Technology, economics, the environment, and policy," *Annual Review of Energy and the Environment* 21(1), 403-465. DOI: 10.1146/annurev.energy.21.1.403

Lynd, L. R., Weimer, P. J., van Zyl, W. H., and Pretorius, I. S. (2002). "Microbial cellulose utilization: Fundamentals and biotechnology," *Microbiology and Molecular Biology Reviews* 66(3), 506-577. DOI: 10.1128/mmbr.66.3.506-577.2002

Lynd, L. R., Van Zyl, W. H., McBride, J. E., and Laser, M. (2005). "Consolidated bioprocessing of cellulosic biomass: An update," *Current Opinion in Biotechnology* 16(5), 577-583. DOI: 10.1016/j.copbio.2005.08.009

Moraïs, S., Barak, Y., Caspi, J., Hadar, Y., Lamed, R., Shoham, Y., Wilson, D. B., and Bayer, E. A. (2010). "Cellulase-xylanase synergy in designer cellulosomes for enhanced degradation of a complex cellulosic substrate," *MBio* 1(5), e00285-10. DOI: 10.1128/mBio.00285-10

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). "Mapping and quantifying mammalian transcriptomes by RNA-Seq," *Nature Methods* 5(7), 621-628. DOI: 10.1038/nmeth.1226

Narisawa, N., Haruta, S., Cui, Z. J., Ishii, M., and Igarashi, Y. (2007). "Effect of adding cellulolytic bacterium on stable cellulose-degrading microbial community," *Journal of Bioscience and Bioengineering* 104(5), 432-434. DOI: 10.1263/jbb.104.432

Nataf, Y., Yaron, S., Stahl, F., Lamed, R., Bayer, E. A., Scheper, T. H., Sonenshein, A. L., and Shoham, Y. (2009). "Cellodextrin and laminaribiose ABC transporters in *Clostridium thermocellum*," *Journal of Bacteriology* 191(1), 203-209. DOI: 10.1128/JB.01190-08

Parisutham, V., Kim, T. H., and Lee, S. K. (2014). "Feasibilities of consolidated bioprocessing microbes: from pretreatment to biofuel production," *Bioresource Technology* 161, 431-440. DOI: 10.1016/j.biortech.2014.03.114

Pinheiro, B., Gilbert, H., Sakka, K., Fernandes, V., Prates, J., Alves, V., Bolam, D., Ferreira, L., and Fontes, C. (2009). "Functional insights into the role of novel type I cohesin and dockerin domains from *Clostridium thermocellum*," *Biochemical Journal* 424(3), 375-384. DOI: 10.1042/BJ20091152

Raman, B., Pan, C., Hurst, G. B., Rodriguez Jr, M., McKeown, C. K., Lankford, P. K., Samatova, N. F., and Mielenz, J. R. (2009). "Impact of pretreated switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: A quantitative proteomic analysis," *PLOS One* 4(4), e5271. DOI: 10.1371/journal.pone.0005271

Rydzak, T., Levin, D. B., Cicek, N., and Sparling, R. (2011). "End-product induced metabolic shifts in *Clostridium thermocellum* ATCC 27405," *Applied Microbiology and Biotechnology* 92(1), 199-209. DOI: 10.1007/s00253-011-3511-0

Rydzak, T., McQueen, P. D., Krokhin, O. V., Spicer, V., Ezzati, P., Dwivedi, R. C., Shamshurin, D., Levin, D. B., Wilkins, J. A., and Sparling, R. (2012). "Proteomic analysis of *Clostridium thermocellum* core metabolism: Relative protein expression profiles and growth phase-dependent changes in protein expression," *BMC Microbiology* 12(1), 214. DOI: 10.1186/1471-2180-12-214

Sakka, K., Kishino, Y., Sugihara, Y., Jindou, S., Sakka, M., Inagaki, M., Kimura, T., and Sakka, K. (2009). "Unusual binding properties of the dockerin module of *Clostridium thermocellum* endoglucanase CelJ (Cel9D-Cel44A)," *FEMS Microbiology Letters* 300(2), 249-255. DOI: 10.1111/j.1574-6968.2009.01788.x

Salamitou, S., Lemaire, M., Fujino, T., Ohayon, H., Gounon, P., Beguin, P., and Aubert, J.-P. (1994). "Subcellular localization of *Clostridium thermocellum* ORF3p, a protein carrying a receptor for the docking sequence borne by the catalytic components of the cellulosome," *Journal of Bacteriology* 176(10), 2828-2834.

Shoham, Y., Lamed, R., and Bayer, E. A. (1999). "The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides," *Trends in Microbiology* 7(7), 275-281. DOI: 10.1016/S0966-842X(99)01533-4

Strobel, H. J., Caldwell, F. C., and Dawson, K. A. (1995). "Carbohydrate transport by the anaerobic thermophile *Clostridium Thermophile Clostridium thermocellum*," *Applied and Environmental Microbiology* 61(11), 4012-4015.

Tokatlidis, K., Salamitou, S., Béguin, P., Dhurjati, P., and Aubert, J.-P. (1991). "Interaction of the duplicated segment carried by *Clostridium thermocellum* cellulases with cellulosome components," *FEBS Letters* 291(2), 185-188. DOI: 10.1016/0014-5793(91)81279-H

Tracy, B. P., Jones, S. W., Fast, A. G., Indurthi, D. C., and Papoutsakis, E. T. (2012). "Clostridia: The importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications," *Current Opinion in Biotechnology* 23(3), 364-381. DOI: 10.1016/j.copbio.2011.10.008

Wei, H., Fu, Y., Magnusson, L., Baker, J. O., Maness, P.-C., Xu, Q., Yang, S., Bowersox, A., Bogorad, I., Wang, W., Tucker, M. P., Himmel, M. E., and Ding, S.-Y. (2014). "Comparison of transcriptional profiles of *Clostridium thermocellum* grown on cellobiose and pretreated yellow poplar using RNA-Seq," *Frontiers in Microbiology* 5, 1-16. DOI: 10.3389/fmicb.2014.00142

Zaldivar, J., Nielsen, J., and Olsson, L. (2001). "Fuel ethanol production from lignocellulose: A challenge for metabolic engineering and process integration," *Applied Microbiology and Biotechnology* 56(1–2), 17-34. DOI: 10.1007/s002530100624

Zhang, Y.-H. P., and Lynd, L. R. (2005). "Cellulose utilization by *Clostridium thermocellum*: Bioenergetics and hydrolysis product assimilation," *Proceedings of the National Academy of Sciences of the United States of America* 102(20), 7321-7325. DOI: 10.1073/pnas.0408734102

Zuroff, T. R., and Curtis, W. R. (2012). "Developing symbiotic consortia for lignocellulosic biofuel production," *Applied Microbiology and Biotechnology* 93(4), 1423-1435. DOI: 10.1007/s00253-011-3762-9

Article submitted: April 1, 2015; Peer review completed: August 24, 2015; Revised version received and accepted: August 28, 2015; Published: September 3, 2015.
DOI: 10.15376/biores.10.4.7105-7122