AMERICAN SOCIETY FOR MICROBIOLOGY | **Applied and Environmental Microbiology®**

LacI Transcriptional Regulatory Networks in *Clostridium thermocellum* DSM1313

Charlotte M. Wilson,[a,b] Dawn M. Klingeman,[a,b] Caleb Schlachter,[a,b] Mustafa H. Syed,[a,b] Chia-wei Wu,[a,b] Adam M. Guss,[a,b] Steven D. Brown[a,b]

BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA[a]; Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA[b]

ABSTRACT Organisms regulate gene expression in response to the environment to coordinate metabolic reactions. *Clostridium thermocellum* expresses enzymes for both lignocellulose solubilization and its fermentation to produce ethanol. One LacI regulator termed GlyR3 in *C. thermocellum* ATCC 27405 was previously identified as a repressor of neighboring genes with repression relieved by laminaribiose (a β-1,3 disaccharide). To better understand the three *C. thermocellum* LacI regulons, deletion mutants were constructed using the genetically tractable DSM1313 strain. DSM1313 *lacI* genes Clo1313_2023, Clo1313_0089, and Clo1313_0396 encode homologs of GlyR1, GlyR2, and GlyR3 from strain ATCC 27405, respectively. Growth on cellobiose or pretreated switchgrass was unaffected by any of the gene deletions under controlled-pH fermentations. Global gene expression patterns from time course analyses identified glycoside hydrolase genes encoding hemicellulases, including cellulosomal enzymes, that were highly upregulated (5- to 100-fold) in the absence of each LacI regulator, suggesting that these were repressed under wild-type conditions and that relatively few genes were controlled by each regulator under the conditions tested. Clo1313_2022, encoding lichenase enzyme LicB, was derepressed in a $\Delta glyR1$ strain. Higher expression of Clo1313_1398, which encodes the Man5A mannanase, was observed in a $\Delta glyR2$ strain, and α-mannobiose was identified as a probable inducer for GlyR2-regulated genes. For the $\Delta glyR3$ strain, upregulation of the two genes adjacent to *glyR3* in the *celC-glyR3-licA* operon was consistent with earlier studies. Electrophoretic mobility shift assays have confirmed LacI transcription factor binding to specific regions of gene promoters.

IMPORTANCE Understanding *C. thermocellum* gene regulation is of importance for improved fundamental knowledge of this industrially relevant bacterium. Most LacI transcription factors regulate local genomic regions; however, a small number of those genes encode global regulatory proteins with extensive regulons. This study indicates that there are small specific *C. thermocellum* LacI regulons. The identification of LacI repressor activity for hemicellulase gene expression is a key result of this work and will add to the small body of existing literature on the area of gene regulation in *C. thermocellum*.

KEYWORDS EMSA, LacI, RNA-seq, *Ruminiclostridium*, transcriptomics, consolidated bioprocessing, gene regulation

*C*lostridium thermocellum is an anaerobic thermophile with an elaborate and highly efficient extracellular enzyme system for solubilizing lignocellulosic biomass. *C. thermocellum* is of biotechnological interest due to its potential for lignocellulose conversion to fuels. *C. thermocellum* produces cellulosomes, which are multiprotein structures consisting of a protein platform (scaffoldins) with attached cellulolytic enzymes. These enzymes are bound to scaffoldins via type 1 cohesion-dockerin domain interactions, and this large complex is anchored to the cell wall by secondary scaffol-

dins and type 2 cohesin-dockerin interations (1). In addition, the bacterium produces free enzymes and "cell-free" cellulosomes (2) and it is a candidate for cellulosic ethanol production (3). Wild-type *C. thermocellum* can utilize hexose sugars but not pentose sugars as fermentation substrates, despite having eight genes encoding xylanases (1). The organism has characterized ABC transporters for the uptake of cellodextrins ranging from G2 to G5 with one, CbpA, specific for cellotriose and another specific for laminaribiose (4). Understanding *C. thermocellum* gene regulation is of importance for improved fundamental knowledge of this industrially relevant bacterium.

   C. thermocellum regulates gene expression in response to carbon sources and growth rates (5–7). The expression of the cellulosomal proteins is influenced by substrate availability and extracellular polysaccharides (8). Most cellulase and hemicellulase genes of *C. thermocellum* are scattered throughout the genome in small clusters, in contrast to the larger operons in mesophilic cellulolytic clostridia (9–16). Methods of sensing and coordinating gene expression must therefore exist to allow the organism to utilize cellodextrins in plant biomass (17). Altered gene expression in response to substrate availability is expected to affect the composition of the cellulosome in terms of binding available dockerin-containing enzymes to the cohesion domains on the scaffoldin (CipA) backbone (6). *C. thermocellum* DSM1313 was used in this study due to the existence of a genetic system and the absence of an equivalent system in the type strain, *C. thermocellum* ATCC 27405. The genomes for strains DSM1313 and ATCC 27405 are very similar, with average nucleotide identities (ANIs) of 99.6 and 99.3% in reciprocal genome comparisons (18). Strain DSM1313 encodes more than 100 transcription regulators. However, only a small proportion of *C. thermocellum* transcriptional regulators have been studied. Previous work has demonstrated the inducible expression of the *celC-glyR3-licA* operon (11, 19), the involvement of alternative sigma factors with their cognate anti-sigma factors in the regulation of cellulolytic genes, including *celS* (17, 20, 21), and histidine kinase genes involved in the regulation of sporulation (22).

   C. thermocellum strains DSM1313 and ATCC 27405 each encode three LacI transcriptional regulators (GlyR1 to GlyR3), and the homologs of these regulators in the two strains are 100% identical in respective comparisons. Among *C. thermocellum* strains, comparisons between LacI proteins give similarity values of ~24 to 42% identity at the amino acid level. LacI family members can sense sugar and noncarbohydrate effectors via a C-terminal effector binding domain, with the majority regulating carbohydrate utilization genes, and have a characteristic N-terminal helix-turn-helix DNA binding domain (23). The number of LacI regulators varies per genome, with most having fewer than 10 such regulators and some actinobacteria having up to 32. The characterized ATCC 27405 genes *celC-glyR3-licA* (Cthe_2807, Cthe_2808, and Cthe_2809) are homologous to DSM1313 genes Clo1313_0395, Clo1313_0396, and Clo1313_0397. CelC and LicA, a glycoside hydrolase 5 family member and glucan endo-1,3-beta-D-glucosidase, respectively, are not cellulosome associated and are instead members of the *C. thermocellum* free cellulolytic enzyme system (1). The GlyR3 regulator on this three-gene operon was demonstrated to have repressor activity, which was relieved in the presence of laminaribiose, a β-1,3-linked glucose dimer. GlyR3 was shown to bind to an 18-bp near-palindromic sequence upstream of the start codon of Cthe_2807 (19).

   To date, the majority of *C. thermocellum* genetic engineering studies have focused on improving fuel production (24–30), but little genetic work has targeted understanding regulatory pathways. Here, we combine use of gene deletions with transcriptomics and DNA binding assays to gain insights into the regulatory networks of the three LacI transcription factors in the genome of *C. thermocellum* DSM1313.

RESULTS

Deletion of the LacI regulators. Three genes annotated in the *C. thermocellum* DSM1313 genome as LacI transcriptional regulators were deleted as previously described (31). Deletion of each locus was confirmed by PCR and Sanger sequencing analysis. In addition, genotypes were analyzed by resequencing analysis by mapping respective Illumina reads to the *C. thermocellum* DSM1313 reference genome. In total,
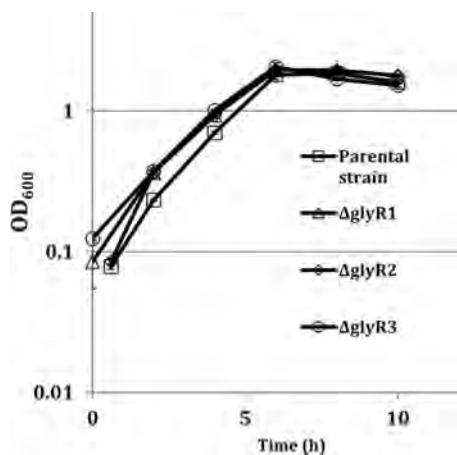
FIG 1 Growth of the parental and mutant strains on $5 \ g \cdot liter^{-1}$ cellobiose in 1-liter bioreactors for RNA-Seq analysis. Plotted are average OD_{600} values from triplicate fermentations with error bars (within symbols and shown as standard deviations). Transcriptomics samples were taken at mid-log phase (OD_{600}, 0.7 to 0.95), late log phase (OD_{600}, 1.8 to 2.0), and in stationary phase approximately 30 min after cultures stopped producing acid (OD_{600}, 2.0 to 2.2).

112 differences were identified by resequencing the three deletion strains and comparing the results to the wild-type reference genome (see Table S1 in the supplemental material). Many single nucleotide polymorphisms (SNPs) were present in multiple strains, with 67 distinct mutations occurring across the three deletion strains, some of which may have also been present in the parental strain. Sixteen of these uniquely identified mutations were predicted to result in nonsynonymous changes to coding regions and validated by Sanger sequencing of the region. Three of these 16 mutations were miscalled during Illumina sequencing analysis and were the same as in the DSM1313 reference genome. Six nonsynonymous mutations occurred in homopolymer repeat regions and were the same as in the ATCC 27405 genome, and they are possibly errors in the DSM1313 reference genome. The remaining seven mutations were confirmed as present in at least one of the strains.

Mutations present in the coding regions of each of the strains were as follows: GlyR1/Clo1313_2023 had an SNP in Clo1313_0908, encoding adenine phosphoribosyltransferase, an SNP in Clo1313_2711, encoding a hypothetical protein, and a transposon insertion in Clo1313_1891, encoding a protein with a predicted uridine kinase function (see Fig. S1 in the supplemental material). GlyR2/Clo1313_0089 had mutations that included a transposon insertion into Clo1313_0145, encoding an aminotransferase (see Fig. S1 in the supplemental material), an SNP in Clo1313_0908, encoding adenine phosphoribosyltransferase, which was distinct from the mutation in GlyR1, and an SNP in Clo1313_0638, encoding a protein with a histone family DNA-binding domain. Clo1313_0396 had a confirmed SNP in uracil phosphoribosyltransferase (Clo1313_0185) and the same SNP in Clo1313_0908 as the GlyR2 strain. Mutations were not found in genes directly related to sugar metabolism and fermentation.

Growth and harvesting for transcriptomic analysis. Little growth difference on the soluble cellobiose substrate was detected between the three *lacI* deletion strains and the parental strain (Fig. 1), and the deletion strains had no apparent phenotype differences when pretreated switchgrass was provided as the carbon source (see Fig. S2 in the supplemental material). For transcriptome sequencing (RNA-Seq) analyses, triplicate one-liter bioreactor cultures were grown for each strain to monitor growth and harvest samples. Each strain reached a similar turbidity before transitioning to stationary phase. Samples were collected from each bioreactor at mid-exponential phase (optical density at 600 nm [OD_{600}], 0.7 to 1.0), late exponential phase (OD_{600}, 1.8 to 2.0), and early stationary phase (OD_{600}, 2.0 to 2.2) to extract RNA for transcriptome profiling by RNA sequencing. Reads per samples ranged from 15.5 million to 36.5 million with an

TABLE 1 Number of expression differences in comparisons with Δ*hpt* parent[a]

Growth stage of sampling	Summary of gene expression differences	No. of differentially expressed genes[b]		
		GlyR1	GlyR2	GlyR3
Mid-exponential	Upregulated 2-fold	2	17	8
	Downregulated 2-fold	11	24	23
Late exponential	Upregulated 2-fold	20	8	23
	Downregulated 2-fold	14	18	42
Stationary	Upregulated 2-fold	572	8	30
	Downregulated 2-fold	520	38	45

[a]Significant differences were defined as changes of ±2-fold and adjusted *P* values of <0.05.
[b]Genes corresponding to GlyR1, GlyR2, and GlyR3 are ΔClo1313_2023, ΔClo1313_0089, and ΔClo1313_0396, respectively.

average of 25 million reads across the 36 samples (see Table S2 in the supplemental material). Between 97.2% and 98.9% of the reads per sample were uniquely mapped to the DSM1313 reference sequence with genome coverages ranging from 194× to 460× (Table S2). Mapped reads were imported into the DESeq2 workflow for normalization and determination of differential gene expression between each strain at a given time point and the parental strain at the same growth stage. Genes were considered differentially expressed if results showed a 2-fold difference in gene expression between a mutant and wild-type strain at a given growth stage and a false-discovery rate (FDR; Benjami-Hochberg method)-adjusted *P* value of less than 0.05 (see Table S3 in the supplemental material).

LacI regulons. A distribution analysis of mean read unique read counts for the Δ*hpt* strain at each sampling point revealed that counts for the three *lacI* genes were in the upper 50th percentile under all growth conditions. The Δ*hpt* strain is the parental strain used in the production of the *lacI* deletion mutants. Deletion of the *hpt* gene in *C. thermocellum* is necessary to create markerless deletion strains. At mid-exponential, late exponential, and early stationary phases, for *glyR1*, there were an average of 3,042, 2,956, and 3,864 reads, respectively; for *glyR2*, 4,651, 4,648, and 2,232 reads, respectively; and for *glyR3*, 1,362, 1,413, and 3,864 reads, respectively. In each deletion strain, a dramatic upregulation of at least one gene occurred in the absence of a particular regulator. This was not dependent on growth stage, as levels of expression remained consistently upregulated by greater than 5-fold across the growth stage transcriptome profiling. Numbers of differentially expressed genes varied between strains and growth stages (Table S3). The focus of this study was derepressed genes under all growth conditions.

GlyR1. Deletion of the *glyR1* gene (Clo1313_2023) resulted in 13, 34, and 1,092 differentially expressed genes at mid-exponential, late exponential, and early stationary phases, respectively (Table 1 shows a summary; see Table S3 in the supplemental material for complete details). Substantial upregulation (>20-fold) of Clo1313_2022 occurred across all three time points in the absence of *glyR1* (Table 2). Clo1313_2022 is divergently transcribed to *glyR1* and encodes LicB, a lichenase enzyme with a dockerin I domain, indicating that this enzyme can bind to the cellulosome scaffoldins. Clo1313_2024, downstream from and on the strand opposite to *glyR1*, did not show differential expression. The only other gene (Clo1313_2865) upregulated at mid-exponential growth phase encoded a protein of unknown function (DUF214). Downregulated across all three time points sampled was an operon, Clo1313_1886-1891, on the reverse strand. Mapped RNA-Seq data suggest that these genes are transcribed as a single unit from the complement strand, as reads were mapped across the intergenic regions of these six genes. Strain resequencing revealed an IS*3* transposon insertion near the start of Clo1313_1891 (see Fig. S1 in the supplemental material), which likely disrupted transcription of this operon in the *glyR1* deletion strain. The Clo1313_1886-1891 genes were the most downregulated genes across all three time points, with at least 40-fold less expression occurring in the stationary phase than in the parental

TABLE 2 Summary of major gene expression differences in the three *lacI* deletion strains

Locus tag	Protein encoded	SNP (Y/N)	Growth phase	Log$_2$ differential expression		
				Δ*glyR1*–Δ*hpt*	Δ*glyR2*–Δ*hpt*	Δ*glyR3*–Δ*hpt*
Clo1313_0089	GlyR2	N	Mid-log	−0.01	Del.[a]	0.09
			Late log	−0.01	Del.	−0.19
			Early stationary	−1.47	Del.	0.09
Clo1313_0395	CelC	N	Mid-log	−0.12	−0.29	3.89
			Late log	0.03	−0.25	3.82
			Early stationary	−0.65	−0.21	2.53
Clo1313_0396	GlyR3	N	Mid-log	0.12	−0.38	Del.
			Late log	0.44	−0.09	Del.
			Early stationary	−0.67	−0.38	Del.
Clo1313_0397	LicA	N	Mid-log	−0.09	−0.19	3.82
			Late log	0.01	−0.08	3.75
			Early stationary	−0.18	−0.07	2.33
Clo1313_1398	Man5A	N	Mid-log	−0.05	4.65	0.06
			Late log	0.02	4.58	0.06
			Early stationary	−0.08	5.66	0.04
Clo1313_2022	LicB	N	Mid-log	4.33	0.70	0.8
			Late log	4.44	0.47	0.39
			Early stationary	6.64	0.22	0.29
Clo1313_2023	GlyR1	N	Mid-log	Del.	−0.49	−0.27
			Late log	Del.	−0.22	−0.29
			Early stationary	Del.	0.11	0.21

[a]Del., deleted (the gene was deleted in that particular strain).

strain. The Clo1313_1886-1891 genes are poorly characterized, unfortunately, with predicted functions including an Fe-S cluster domain, a putative anti-sigma regulatory factor. If the anti-sigma regulatory factor annotation is correct, this gene could affect the expression of other *C. thermocellum* genes and be contributing to the large set of differentially expressed genes in the stationary-phase comparison with the parental strain.

GlyR2. Deletion of the *glyR2* gene (Clo1313_0089) resulted in 41, 26, and 46 genes significantly differentially expressed at mid-exponential, late exponential, and early stationary phases, respectively (see Table 1 for a summary and Table S3 in the supplemental material for complete details). Clo1313_0090 encodes an S-layer domain-containing protein and is downstream of and transcribed divergently from *glyR2*. The Clo1313_0090 gene showed approximately 2-fold-higher expression in the *glyR2* mutant strain than in the parent strains only in stationary phase, and its conditional derepression indicates other regulation. Clo1313_0088 encodes a hypothetical protein, is on the same strand, and is immediately downstream of *glyR2*, and it did not show differential expression under the conditions tested (Fig. S3 and Table S3 in the supplemental material). In the Δ*glyR2* strain, upregulation of Clo1313_1398 (Cthe_0821) was observed across all three sampling time points (Fig. 2). Clo1313_1398 encodes the Man5A protein, which contains a dockerin I domain required for attachment to the cellulosome scaffoldin proteins. Man5A also has GH5 and CBM32 domains and has been demonstrated to have a preference for manno-oligosaccharides. During the mid-exponential phase, deletion of the *glyR2* gene resulted in an upregulation of a further nine genes with carbohydrate-active enzyme (CAZyme)-related functions and a gene encoding a glycogen/starch synthase. The expression patterns of three of these genes was maintained at a higher level in the *glyR2* deletion strain than in the parental strain during late log phase; however, by stationary phase these differences between the two strains had diminished, aside from the dramatic upregulation of Clo1313_1398.

Eight genes with predicted functions in the pentose phosphate pathway and central carbon metabolism, Clo1313_0073 to Clo1313_0080, encoding two transketolase domain proteins, phosphoglycerate mutase, and a glycerol kinase, were downregulated across all three time points sampled for both *glyR2* and *glyR3* deletion strains compared to the parental strain (Table S3). One gene, Clo1313_0638, annotated as encoding a histone family protein DNA-binding protein, was downregulated more than 4-fold;
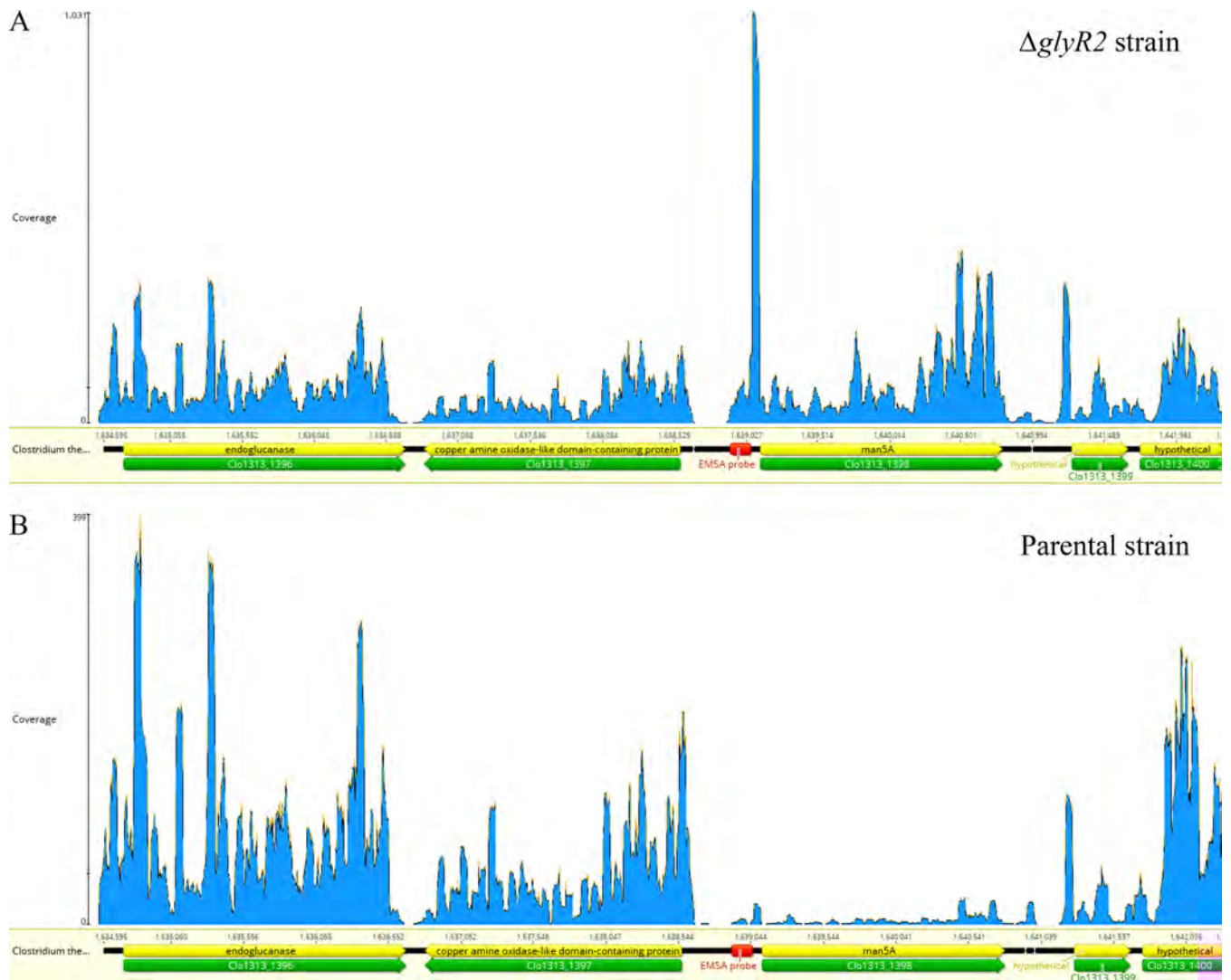
FIG 2 RNA-seq analysis for the *man5A* region in Δ*glyR2* mutant and parent strains. Approximately 30 million reads for deletion (A) and parent (B) strains were mapped to the reference genome with the read numbers mapping indicated by coverage.

however, the strain resequencing had identified an SNP present in the coding region (GI 724743). This SNP led to an isoleucine-to-threonine change at the amino acid level. The *glyR3* deletion strain had the same mutation and also had a downregulation of the gene similar to the one seen in the parental strain (see below), suggesting that this pattern of gene expression was related to the SNP and not the respective *lacI* gene deletions.

GlyR3. Deletion of the previously described *C. thermocellum* LacI transcriptional regulator gene, *glyR3* (Cthe_2808/Clo1313_0396), resulted in 31, 65, and 75 genes considered significantly differentially expressed at mid-exponential, late exponential, and early stationary phases, respectively (Table 1). Upregulation of the two genes adjacent to *glyR3* in the *celC-glyR3-licA* operon was consistent with the activity of this regulator as previously reported (11, 19), and each showed similar trends across all three growth stages (Fig. S3 and Table S3 in the supplemental material). The *orf4* gene (Clo1313_0398) showed 1.5 $\log_2$-fold-higher expression in the *glyR3* mutant only for stationary-phase samples. Otherwise, no upregulation of the downstream genes (*orf4-manB-celT*) was detected in the *glyR3* deletion strain, similar to what was observed when *C. thermocellum* was grown on laminarin (11). Clo1313_2013, annotated as encoding a hypothetical protein, was the only other gene that maintained a greater
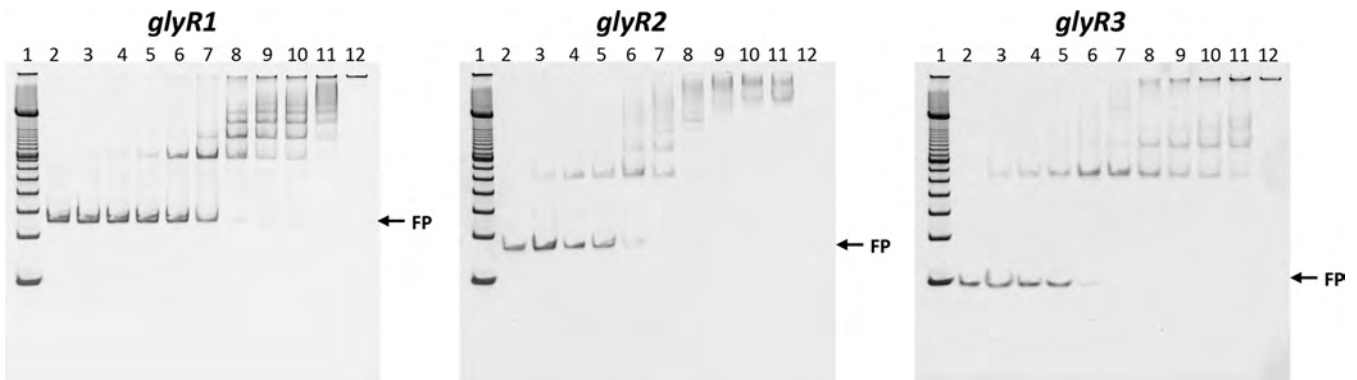
FIG 3 Electrophoretic mobility shift assays using purified LacI transcriptional regulators with DNA fragments from the promoters of regulated genes. Approximately 0.03 μM, 0.02 μM, and 0.03 μM DNA fragments from the upstream genomic regions of Clo1313_2022, Clo1313_1398, and Clo1313_0395 were incubated with increasing amounts of purified protein encoded by Clo1313_2023, Clo1313_0089, and Clo1313_0396, respectively. Lanes were loaded as follows: lane 1, 100-bp marker (Life Technologies); lane 2, DNA only; lanes 3 to 11, DNA plus LacI regulator (for GlyR1 and GlyR2, 0.016, 0.03, 0.06, 0.13, 0.25, 0.50, 0.75, 1.0, 1.5 μM, respectively; for GlyR3, 0.06, 0.13, 0.25, 0.50, 0.75, 1.0, 1.5, 2.0, 2.5 μM); lane 12, protein only. FP indicates where the free probe is running on the gel.

than 2-fold upregulation of gene expression across the three growth stages when *glyR3* was deleted. The most downregulated gene (Clo1313_0638) at mid-exponential and late-exponential growth phases contained the same SNP as the Δ*glyR2* strain. This was the only SNP in the coding region of a gene in the Δ*glyR3* strain that displayed differential gene expression. Another gene with CAZyme-related functions is Clo1313_1424 (*celX*), encoding a predicted acetylxylan esterase cellulosomal protein, which was upregulated 2-fold in stationary phase. Two members of the free-cellulase system in *C. thermocellum* were upregulated by stationary phase, Clo1313_1336 (GH23) and Clo1313_2473 (GH18), with possible substrates of chitin and/or peptidoglycan.

A histidine kinase (Clo1313_1711), predicted to be involved in regulating sporulation in *C. thermocellum*, was downregulated in late exponential and early stationary phases in the *glyR3* deletion strain compared to the parental strain. Other genes with sporulation annotations (Clo1313_0205, SpoIIID; Clo1313_1773, SigE; Clo1313_1774, SpoIIGA; Clo1313_1409, Spo0A; Clo1313_1548, stage II sporulation protein M; Clo1313_1381, stage III sporulation protein AC; Clo1313_1381, stage III sporulation protein AF) were downregulated at stationary phase. The average read counts for the triplicate samples for many of these genes were low in both the parental and mutant strains across all time points sampled: SpoIIID, 3 to 49; SigE, 123 to 1,143; SpoIIGA, 12 to 55; Spo0A, 128 to 648; stage II sporulation protein M, 5 to 35; stage III sporulation protein AC, 7 to 29; stage III sporulation protein AF, 23 to 128. Spores were not identified during culturing of the *lacI* mutants; however, *C. thermocellum* is a poor sporeformer, with rates of less than 10% under typical conditions that would induce sporulation and less than 0.002% under typical lab conditions (32).

LacI regulators bind to regulatory sequences of target genes. To test whether each LacI regulator has a direct role in the regulation of the genes that displayed increased expression in the transcriptomic analysis, each regulator was assessed for the ability to bind the DNA fragments from promoter regions of the genes of interest via several independent electrophoretic mobility shift assays (EMSAs). Fluorescent EMSA was more sensitive than other EMSAs, and these results are presented. Chemiluminescent EMSA data were also collected using independent probe sequences (not shown). Purified LacI transcription factors were confirmed to be able to bind to DNA fragments from particular promoter regions, with increasing amounts of protein in the assay binding increasing amounts of DNA fragments (Fig. 3). GlyR3 bound to the region as previously described, GlyR1 bound upstream of *licB* (Clo1313_2022), and the GlyR2 regulator bound to a region upstream of *man5A* (Clo1313_1398).

LacI proteins typically dissociate from DNA upon binding of the effector molecule. While laminaribiose has been shown to be an effector molecule for GlyR3, effectors for

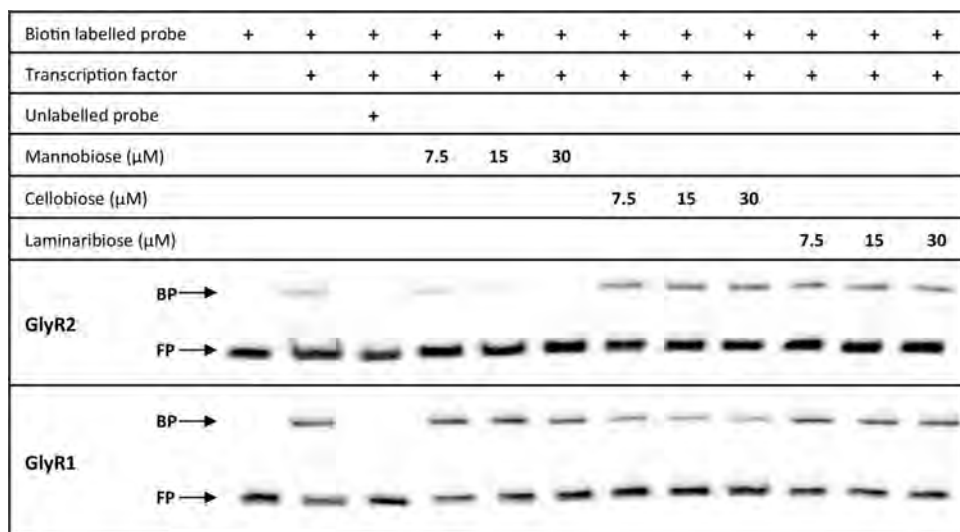| Biotin labelled probe	+	+	+	+	+	+	+	+	+	+	+	+
Transcription factor		+	+	+	+	+	+	+	+	+	+	+
Unlabelled probe		+										
Mannobiose (µM)				7.5	15	30						
Cellobiose (µM)							7.5	15	30			
Laminaribiose (µM)										7.5	15	30

FIG 4 Sugar interaction tests for GlyR1 and GlyR2 via EMSA with biotin-labeled DNA fragments from the promoters of regulated genes. Twenty femtomoles of biotin-labeled probe and 15 nM purified transcription factor were used. Control competitive binding reaction mixtures included 4 pmol of unlabeled probe. Bound probe (BP) and free probe (FP) are indicated, and sugars were included in experiments at the designated quantities.

GlyR1 and GlyR2 are unknown. Various sugars were tested with biotin-labeled fragments to identify sugars that would disrupt binding. Cellobiose, laminaribiose (Fig. 4), maltose, sucrose, glucose, lactose, galactose, arabinose, and xylose (data not shown) did not affect DNA binding by either GlyR1 or GlyR2, and α-mannobiose had no impact on GlyR1 binding. GlyR2 was no longer able to bind in the presence of α-mannobiose. This suggests that α-mannobiose is a probable inducer for GlyR2-regulated genes (Fig. 4), consistent with the function of Clo1313_1398 as a mannanase. Individual sugars did not yield a candidate inducer molecule for GlyR1, and the possible inducer molecule for this transcription factor remains unknown.

DISCUSSION

This study sought to identify the regulons of three LacI family transcription factors to better understand the regulation of genes involved in plant biomass deconstruction in *C. thermocellum* DSM1313. The *C. thermocellum* genome encodes a large repertoire of enzymes, both cellulases and hemicellulases that can target plant cell walls, degrading the cellulose and hemicellulose components to liberate soluble β-1,4- and β-1,3-linked glucans that the bacterium will ferment. The composition of the cellulosome has been shown to vary depending on the substrate (8), and cellulase gene expression can be affected by the growth rate of the organism (5). However, the regulatory mechanisms employed by the cell to determine the cellulosome composition and *C. thermocellum* regulation in general remain relatively understudied.

LacI family members can sense sugar and noncarbohydrate effectors, with the majority regulating carbohydrate utilization genes (23). A comparative study of 1,300 LacI transcription factors found that nearly 90% of these were local regulators of sugar utilization pathways, with regulatory effects impacting only genes located near the regulator gene in the genome, while the other 10% of LacI regulators control large sets of genes and are considered to be global regulators. LacI family members can also be integrated into complex regulatory networks, as is the case for *Mesorhizobium loti* FixV, which activates genes for symbiotic nitrogen fixation (33), as an example.

All three LacI transcription factors affected the regulation of genes encoding hemicellulases. Hemicelluloses constitute significant proportions of plant biomass polysaccharides, between one-quarter and one-half of the mass of many types of cell walls, and they vary tremendously depending upon the plant species and tissue type (34). The four main kinds of hemicelluloses are xyloglucan, xylans, mixed-linkage glucans, and

mannans, which contain different backbones with various side groups, bonds, degrees of branching, and polymerization. Xyloglucan is often the most abundant hemicellulose in primary cell walls, and it consists of a β-1,4-glucan backbone with xylose-containing side chains, which may also contain other sugars (e.g., arabinose). However, in grass primary cell walls, glucuronoarabinoxylan is a dominant hemicellulose, and it consists of mixed-linkage glucans. Xylan is a polymer of β-1,4-linked xylose with possible arabinan and glucuronic acid side chains. Mixed-linkage glucans are made of glucose residues linked by β-1,3 and β-1,4 bonds. A mannose backbone joined by β-1,4 linkages makes up mannans or similarly linked glucose and mannose residues comprise gluco-mannans. Wild-type *C. thermocellum* strains are unable to use these hemicellulose sugars as carbon sources. Nevertheless, *C. thermocellum* encodes a variety of enzymes needed to break hemicellulose bonds in order to access cellulose from a variety of sources and consume it for growth. We have identified new LacI repressor activity for genes with hemicellulase functions, which is a key result of this work.

The characterized ATCC 27405 genes *celC-glyR3-licA* (Cthe_2807, 2808, 2809) are homologous to DSM1313 genes Clo1313_0395, 0396, 0397. Expression studies across growth stages have identified a peak in the gene expression from this operon at late exponential phase as the cell transitions into early stationary phase (11, 35, 36). The activity of the GlyR3 regulator on this three-gene operon was demonstrated to have repressor activity, which was relieved in the presence of laminaribiose, a β-1,3-linked glucose dimer. GlyR3 was shown to bind to an 18-bp near-palindromic sequence upstream of the start codon of Cthe_2807 (19). GlyR3 has also been used to create a laminaribiose-inducible expression system (37). We find that all three LacI transcription factors in *C. thermocellum* DSM1313 appear to be local rather than global regulators and each LacI controlled the expression of a small number of genes encoding CAZyme functions. Deletion of each LacI transcription factor and transcriptome profiling across three different growth stages revealed that the deletion of each transcription factor had a dramatic effect on the expression of at least one other gene across all time points sampled. This indicated that the absence of the regulator relieved the transcription repression and this was not confounded by the growth stage of the culture, consistent with roles as transcriptional repressors.

The regulon of GlyR3 appears to be limited to the *celC-glyR3-licA* operon, as previously described (11, 19). GlyR1 affected the expression of *licB* (Clo1313_2022), containing a glycoside hydrolase 16 domain. This gene encodes a lichenase (β-1,3-1,4-glucanase), which has been demonstrated to localize to the cellulosome (38) and cleaves β-1,4 linkages adjacent to β-1,3 linkages in mixed β-glucans, such as barley β-glucans and lichenin (39, 40). LicB has no activity against β-1,3-β-1,6 linkages such as laminarin (39) and so has a function distinct from that of the LicA/CelC enzymes regulated by GlyR3. Purified GlyR1 bound to a 232-bp region upstream of the *licB* gene, and this binding was dose dependent. Monomeric and multimeric sugar molecules were used to determine an inducer molecule for this regulator; however, the binding of GlyR1 to the *licB* promoter region could not be disrupted and the inducer molecule remains unknown. Future studies that could fractionate different components of the lignocellulosic substrate or a higher-throughput assay would be useful to narrow down possible candidate effector molecules for GlyR1.

Deletion of *glyR2* led to significant (>2-fold) upregulation of 17 genes. Many had CAZyme activities, and the expression of nine of these gene were previously reported to display a decreasing level of expression when the growth rate of *C. thermocellum* ATCC 27405 on cellobiose was increased (5). In the *glyR2* mutant strain, these genes had higher levels of expression at exponential growth phase than did the parental strain, and this trend was not observed in either of the other two strains, suggesting that this effect is specific to this *lacI* deletion. Clo1313_1398, encoding Man5A, had substantially higher expression when *glyR2* was deleted. Interestingly, the regulation of Clo1313_1398 (*man5A*) in the two, similar *C. thermocellum* strains ATCC 27405 and DSM1313 appears to differ. Cthe_0821, the homolog of Clo1313_1398, was one of the most highly expressed genes when the cells were grown on biomass (7) and was in the top 6% of expressed genes

grown on cellulose (5), while Clo1313_1398 was only in the top 30% of expressed genes when DSM1313 was grown on cellulose (2). The genomic architecture of this region does differ between the two strains, with DSM1313 lacking two transposase genes immediately upstream of Cthe_0821. This difference between the two strains has disrupted a σ^A recognition motif and might be the reason for the strain differences. EMSAs revealed that the purified GlyR2 bound to a 149-bp region upstream of the start site of the Clo1313_1398 gene in a dose-dependent manner. This binding could be disrupted by the presence of mannobiose. This, along with the RNA-Seq data revealing high *man5A* expression levels in the *glyR2* deletion strain, is indicative of the transcription factor acting as an inducible repressor. The ATCC 27405 homolog displays endomannanase activity and over an extended incubation period can hydrolyze mannopentaose to mannobiose (35). As mannobiose is both an inducer and a product of the activity of Man5A, this suggests that feedback regulation could occur in the cell. Due to the genomic architecture of the region, this may be more important in the DSM1313 strain than the ATCC 27405 strain, where the gene is already expressed at a constitutive level.

Deletions strains were resequenced, which identified a series of SNPs and other mutations, including transposon insertions in two of the strains. The occurrence of mutations and insertion in phosphoribosyltransferases may result from selective pressure on nucleotide pathways during the mutagenesis strategy employed to make these deletion strains. Fortunately, mutations were identified, and the effects associated with these mutations could be excluded from the interpretation of the transcriptome analyses. It is recommended that mutants be routinely resequenced to increase awareness of the genomic context of targeted deletions.

This study adds to the small body of knowledge for *C. thermocellum* gene regulation and identified negative regulation affecting the expression of cellulosome-related genes in this bacterium. Previous studies have demonstrated negative regulation of noncellulosomal enzymes by GlyR3 transcription factor (11, 19). Positive regulation of cellulosome enzymes through cellulose and xylan interactions with anti-σ factors has been reported (17, 20, 21). Further studies are required to refine knowledge of *C. thermocellum* gene-regulatory networks contributing to the control and composition of the cellulosome and the lignocellulolytic abilities of the organism, regulation of important physiological responses, and cross talk between networks.

MATERIALS AND METHODS

Bacterial strains and mutant construction. Plasmids were constructed with standard methods (41), including yeast gap repair (42) and Gibson assembly (43) using the Gibson assembly cloning kit (New England BioLabs, Ipswich, MA, USA), resulting in plasmids pAMG488 for deletion of Clo1313_0396, pCS2 for deletion of Clo1313_0089, and pCS6 for deletion of Clo1313_2023. Plasmids were isolated from *Escherichia coli* strain BL21, such that the DNA was not methylated by Dcm methylase in *E. coli* prior to transformation into *C. thermocellum* (44). *C. thermocellum* transformations and gene deletions were performed as previously described (25, 31) using primers shown for construction and strain verification (Table 3). Annotated plasmid maps are provided in the supplemental material. Briefly, plasmid DNA was electroporated into *C. thermocellum* Δ*hpt* and plated on solid CTFUD medium containing 5 μg ml^{-1} thiamphenicol (TM; Sigma-Aldrich, St. Louis, MO) to select for transformation. Colonies were picked into liquid CTFUD medium supplemented with TM, followed by plating on solid CTFUD medium containing TM and 50 μg ml^{-1} 5-fluoro-2′-deoxyuridine (Sigma-Aldrich, St. Louis, MO) to select for merodiploids strains. After single-colony purification, colonies were picked into liquid CTFUD medium containing TM. Strains were subcultured in CTFUD medium without antibiotics, followed by plating on solid CTFUD medium supplemented with 500 μg ml^{-1} 8-azahypoxanthine (Acros Organics, Pittsburgh, PA). Strains were again single-colony purified and picked into liquid CTFUD medium, and the resulting unmarked gene deletions were verified by PCR.

Fermentation growth conditions. Bioreactors were prepared for *C. thermocellum* culture by autoclaving cellobiose, water, and resazurin (total volume, 800 ml) in 1-liter Biostat Q-Plus bioreactors (Sartorius Stedim, Gottingen, Germany) for 30 min. Reactors were sparged with nitrogen gas for 1 h, and a sterile cocktail of the remaining medium for thermophilic clostridia (MTC) components was prepared (45). A cocktail (100 ml) was added to each reactor, and sparging with nitrogen continued overnight with observable loss of color from reduced resazurin. The initial vessel pH was checked postautoclaving using an external pH probe, and slight adjustments were made to the pH calibration, controlling the reactors as necessary. The fermenters were run at 58°C and 100 rpm, and pH was controlled at 7.0 with 3 N NaOH. Immediately prior to inoculation, the gas was switched from 100% nitrogen to an 80% nitrogen–20% CO_2 mix, and the pH of the reactor was allowed to drop to ~6.2. The automatic base addition pump was switched on, and the pH of the reactor was brought back up to pH 7.0. The bioreactors were

TABLE 3 Oligonucleotides used in this study

Primer use and name	Sequence (description)
Mutant construction	
Cthe0210up-f-4AMG550Bam	TGATCCACTAGTAAGCTTACCCCTGCGAATACTGAGATTT
Cthe0210up-r-for0210down	GATTTTAAGTTGAGACAAATCAGTTTTCTTTGTATTT
Cthe0210down-f-for0210up	ATTTGTCTCAACTTAAAATCAGCCTTTTAAGTAAAGAA
Cthe0210down-r-4AMG550Bam	GGCCTCGCGAAGGAGTTCTTCCGTCGTAACGT
Cthe0210in-f-4AMG550Eco	CCTTTTGTTTTGGTACCGTCAAGCAATAATTCAATCGAGC
Cthe0210in-r-4AMG550Eco	AAACGCTGAGGCGCGCCGATCGTCGGAGTTTCAAGAAG
Cthe2505up-f-4AMG550Bam	GATCCACTAGTAAGCTTGCCTATGGTTCCGCAACTTTC
Cthe2505up-r-for2505down	TCAGCCTATTCGCTTTACTGACATTGTGACTTTCTTTGCCATT
Cthe2505down-f-for2505up	AAGAAAGTCACAATGTCAGTAAAGCGAATAGGCTGAT
Cthe2505down-r-4MG550Bam	ACTGGCCGGCCTCGCGAGTTGGGAACATTTCTGTGAATTC
Cthe2505in-f-4AMG550EcoRI	CCTTTTGTTTTGGTACCGAGCAAGTCAAATTCAAGAACCG
Cthe2505in-r-4AMG550EcoRI	AAACGCTGAGGCGCGCCGTCATCGGGAACGGATATACC
Ct2808down-r-4AMG550	ATTGATACACCCATGTCAGAATACTGGCCGGCCTCGCGAGGATCCGATTTAACGACAACACTTTCAACC
Ct2808down-f-42808up	TAATTTAAAAAATACAACAGCTGTTGATTGATTGTTTTATAATCATCCAAAGATGTACTC
Ct2808up-f- 4AMG550	GGCGCTACAGGGCGCGTGGGGATGATCCACTAGTAAGCTTGGATCCATTACGGAGAAGGACATTGAAACTATT
Ct2808up-r- 42808down	ACATCTTTGGATGATTATAAAACAATCAATCAACAGCTGTTGTATTTTTTAAATTACAAA
Ct2808in-f- 4AMG550(Smal)	TCCTTTTATAGGCGTTTTGAAACCTGAAATGTACAGCTAAGGAGGTGGGCCCATGACCAGTGAAGAAATAGCAAAATTAT
Ct2808in-r- 4AMG550(Smal)	AAGGCCCAGTCTTCCGACTGAGCCTTTTGTTTTCTCGAGGTCAGAATTCCAAAGCCCTCTTGGT
Mutant verification	
F_AAACCTGTCTCGCAAGAATTG	AAACCTGTCTCGCAAGAATTG (upstream of Che_2808)
R_CGCGTTTCCTCTTTTACGTT	CGCGTTTCCTCTTTTACGTT (reverse within Cthe_2808)
R_GTACGGATATTGGCGACCAG	GTACGGATATTGGCGACCAG (downstream at end of Cthe_2809)
F_CTTGTTTACCGCATCTGCAA	CTTGTTTACCGCATCTGCAA (5′ of Cthe_2505)
R_AATTCCTCAACTGCCCACAG	AATTCCTCAACTGCCCACAG (reverse within Cthe_2505)
R_TTTGCCTGCATTACAACAGC	TTTGCCTGCATTACAACAGC (3′ of Cthe_2505)
F_TTTCATGGCCTTCATGTTGA	TTTCATGGCCTTCATGTTGA (5′ of Cthe_0210)
R_CATGGCTATCGGAGCATACA	CATGGCTATCGGAGCATACA (reverse within Cthe_0210)
R_AACACTTCAGCAAAGCTCCTG	AACACTTCAGCAAAGCTCCTG (3′ of Cthe_0210)
Cloning	
F_pTXB1 Clo1313_2023/0210	TAATTTTGTTTAACTTTAAGAAGGAGATATACAATGAACAGCAAGGATATAGC
R_pTXB1_CLO1313_2023/0210	GGGTAGGGCAACTAGTGCATCTCCCGTGATGCAGACAATTTTTTTACATGAATTACGC
F_pTXB1_CLO1313_0089/2505	TAATTTTGTTTAACTTTAAGAAGGAGATATACAATGTCACAATG
R_pTXB1_Clo1313_0089/2505	GGGTAGGGCAACTAGTGCATCTCCCGTGATGCATATTCGCTTTACTGAC
F_pTXB1_Clo1313_0396/2808	TAATTTTGTTTAACTTTAAGAAGGAGATATACAATGACCAGTGAAGAAATAGC
R_pTXB1_Clo1313_0396/2808	GGGTAGGGCAACTAGTGCATCTCCCGTGATGCAGAATTCCAAAGCCCTCTTG
F_pTXB1 across insertion	TAATACGACTCACTATAGGG
R_pTXB1 across insertion	GTAGGGCAACTAGTGCATCT
R. pTXB1 A up of insertion	TGTATATCTCCTTCTTAAAG
F. pTXB1 A down of insertion	ATCACGGGAGATGCACTAGT
Fluorescent EMSA	
F_EMSA_288bp_2022	GCAAAGCTCCTGTAACAATTCA
R_EMSA_288bp_2022	TGCCACACCACCTTCATATCA
F_EMSA_100bp_0395	CCGAATAAAAACTGGACAGAG
R_EMSA_100bp_0395	TGAAACCATTTAACACTGGATTAT
F_EMSA_refined_int1397:98	TTCAATAAAGCGGAATTTGAAGA
R_EMSA in refined 1398	CGGAGGACTTGTGTCGGTT
Chemiluminescent EMSA	
F_EMSA refined_int1397:98Biotin	TTCAATAAAGCGGAATTTGAAGA
R_EMSA in refined 1398	CGGAGGACTTGTGTCGGTT
F_EMSA 288bp_2022	GCAAAGCTCCTGTAACAATTCA
R_EMSA_2022_6_Biotin	TAGTTGACAATCTGAAGTTTA
SNP detection	
F_Clo1313_0908_1058322_1058	CAAGGAGGATTTAACATGGATTT
R_Clo1313_0908_1058322_1058	ACCACCCTTTGACCCTTTTT
F_Clo1313_1891_2208653_2208	TAGCATCCGGACTGGTTTTT
R_Clo1313_1891_2208653_2208	ATGAATGAAAACAATCTAAATA
F_Clo1313_0145_159861_15986	CAGATCCCAGTTGGTTTGGA
R_Clo1313_0145_159861_15986	CAGGAAAATACAGGCTGTTGG
F_Clo1313_0185_199883_19988	GTTGGCAAAATCGGTGAGAT
R_Clo1313_0185_199883_19988	CATCGGATCGAGCACTACAA

subsequently inoculated using overnight inoculum cultures of *C. thermocellum* DSM1313 strains that had been grown anaerobically in 500-ml bottles of MTC containing 5 g/liter cellobiose. Inoculum aliquots of 100 ml were subcultured into each reactor (total volume, 1 liter) for a final inoculum of ~10%. Postinoculation, the gas was switched from passing through the growth medium to just passing through the headspace of the reactor and was used only to create positive vessel pressure during sampling; otherwise, the gas sparge remained off during bacterial growth. Triplicate fermentations were performed for each strain. Growth was tracked by base addition, and OD_{600} measurements were taken every 2 h. Samples (50 ml) were removed at mid-log phase (OD_{600}, 0.7 to 0.95), late log phase (OD_{600}, 1.8 to 2.0), and stationary phase approximately 30 min after cultures stopped producing acid (OD_{600}, 2.0 to 2.2).

Genomic DNA preparation and RNA extraction. Genomic DNA for each of the *lacI* deletion strains was prepared using a Wizard Genomic DNA purification kit (Promega Corp., Madison, WI, USA). For total RNA preparation, cells pelleted from a 50-ml sample drawn from each fermenter were resuspended in 3 ml of TRIzol reagent (Invitrogen, Carlsbad, CA, USA). A 1-ml aliquot of the TRIzol cell suspension was used for cell lysis by bead beating with 0.8 g of 0.1-mm glass beads (BioSpec Products, Bartlesville, OK, USA) and three 20-s bead-beating treatments at 6,500 rpm in a Precellys 24 high-throughput tissue homogenizer (Bertin Technologies, Montigny-le-Bretonneux, France). The RNA from each cell lysate was purified, DNase I treated, and quantity and quality assessed as previously described (7). Purified RNA was depleted of rRNA using the Ribo-Zero rRNA removal kit for Gram Positive Bacteria (Epicentre, Madison, WI, USA). The rRNA-depleted RNA sample was concentrated with RNA Clean & Concentrate-5 (Zymo Research, Irvine, CA, USA) according to the manufacturer's protocol.

Library preparation and DNA sequence data generation. For genome resequencing, libraries were prepared using a TruSeq LT kit that employed a gel-based size selection and subsequently sequenced on an MiSeq instrument (Illumina, San Diego, CA, USA) using version 2 chemistry and a 2 × 250-bp paired-end configuration. Depleted RNA was used as the starting material for the ScriptSeq v2 RNA-Seq Library Preparation kit (Epicentre, Illumina compatible) utilizing the Fail Safe PCR Enzyme mix (Epicentre) and observing the manufacturer's protocol. cDNA tagged with Script-seq adaptors (1–12) was eluted with 20 μl of buffer EB, provided in the Min-Elute PCR purification kit (Qiagen, Valencia, CA), according to the Script-Seq mRNA-Seq Library preparation kit protocol. Twelve PCR cycles were used during library amplification, and samples were purified using Agencourt AMPure XP beads (Beckman Coulter) according to the library protocol and eluted with 20 μl of nuclease-free water. The final mRNA-Seq library was quantified with a Qubit Fluorometer (Invitrogen), and library quality was assessed with Agilent Bioanalyzer DNA 7500 Chip for DNA (Agilent, Santa Clara, CA, USA). Samples were pooled into three groups to make a stock of 20 nM in 72 μl. A 40-μl aliquot of the three pooled libraries was sequenced using an SR50 sequencing run on an Illumina HiSeq 2500 platform (HudsonAlpha Genomic Services Laboratory, Huntsville, AL).

Resequencing and RNA-seq analysis. The FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) was used for quality control of resequencing data. Illumina reads were trimmed if the median quality of reads at 5′ or 3′ ends was below 30, and any reads with more than 10% of positions below a quality score of 30 were filtered out. Bowtie2 (46) was used to align reads to the *C. thermocellum* DSM1313 genome sequence (GenBank accession number CP002416.1) with default parameters, which generated sam files, and subsequently samtools (47) was used to call variants (with minimum read depth of 20 being specified). Putative SNPs were verified by PCR and Sanger sequencing analysis using the described primers (Table 3). Raw resequencing data are available from the NCBI Sequence Read Archive (SRA) (see below). For RNA-seq, reads were mapped to the *C. thermocellum* DSM1313 genome using CLC Genomics Workbench version 8 (CLCbio) using the default settings for prokaryote genomes. Counts of uniquely mapped reads were analyzed for differential gene expression by DESeq2 (48). Filtering was applied to identify those genes with an FDR of <0.05 and >$\log_2$ of ±1 for differential gene expression.

Construction of recombinant expression plasmids. Each *lacI* gene was cloned into plasmid pTXB1 (New England Biolab [NEB], Ipswich, MA, USA) using the indicated primers (Table 3). The native stop codon was left off the construct so that a fusion chitin binding domain tag would be expressed on cloning into pTXB1. These were transformed into 5-α High Efficiency Competent *E. coli* (NEB), and constructs were verified by PCR and Sanger sequencing. Plasmids were isolated and transformed into strain T7 Express High Efficiency Competent *E. coli* (NEB) for expression.

Protein purification. A single colony from an LB plate was used to inoculate 10 ml of LB medium with ampicillin (100 μg/ml). The culture was grown overnight at 37°C with 200 rpm shaking. A 2-ml aliquot of the overnight culture was used to inoculate a 500-ml volume of LB medium, and the culture was grown at 37°C with 200 rpm shaking until an OD_{600} of ~0.6 was reached. Protein expression was induced with filter-sterilized isopropyl-β-D-thiogalactopyranoside (IPTG; final concentration, 0.5 mM), and incubation followed overnight at 15°C with 200 rpm shaking. Cells were harvested by centrifugation in two lots of 250-ml culture at 8,000 × *g* for 15 min at 4°C. Cell pellets were resuspended in 25 ml of column buffer (20 mM Tris-HCl [pH 8.5], 0.5 M NaCl, 1 mM EDTA, 0.1% Tween 20) and transferred to 50-ml tubes. Cells were lysed by sonication on ice using a microtip, with 60 cycles of 15 s on and 30 s off with the power set at 4 (Misonix Sonicator 4000; Misonix Inc., Farmingdale, NY, USA). Cell debris was removed by centrifugation at 7,919 × *g* for 30 min at 4°C. The fusion protein was purified using chitin resin (NEB) according to the manufacturer's protocols. After binding the fusion protein to the column and washing of the bound material, the LacI protein was cleaved from the CBD fusion protein by incubating the column with column buffer containing 50 mM dithiothreitol (DTT) ~20 h at room temperature. Eluted fractions (1.5 ml) were analyzed by SDS-PAGE, fractions observed to contain cleaved protein were pooled, and the column buffer was exchanged for phosphate-buffered saline (PBS) using Amicon Ultra

0.5-ml 10-kDa spin columns (EMD Millipore, Billerica, MA, USA). Protein concentration was determined by the Bradford assay using bovine serum albumin (BSA) standards.

Electrophoretic mobility shift assay. *In vitro* binding of each LacI transcription factor to candidate promoters of genes showing differential patterns of gene expression was assessed by EMSA using a fluorescence-based detection EMSA kit (Life Technologies, Carlsbad, CA, USA). Target DNA fragments were chosen based on the results of the RNA-seq experiment with candidate genes showing strong upregulation in the absence of a particular LacI transcription factor. Genomic regions upstream of candidate responsive genes were amplified by PCR (see Table 3 for primers) and purified using a PCR purification kit (Qiagen). Increasing amounts of each purified LacI transcription factor (final concentrations: 0089, 0.016 to 1.5 μM; 0396, 0.06 to 2.5 μM; 2023, 0.016 to 1.5 μM) were incubated with purified target DNA (0.02 to 0.03 μM) for 30 min at room temperature using 1$\times$ EMSA binding buffer provided in the kit and 10-μl reaction mixtures. After the 30-min incubation, 2 μl of DNA loading dye was added to each tube, and the samples were separated on a 6% Tris-borate-EDTA (TBE) nondenaturing polyacrylamide gel (Life Technologies) at 125 V for 60 min in prechilled 0.5$\times$ TBE buffer. Gels were rinsed with deionized water and stained with SYBR green for 30 min in the dark at room temperature on a shaking platform set to 50 rpm. The gels were rinsed with ~150 ml deionized water twice and visualized using a transilluminator GelDoc (UVP) with a UV wavelength of 302 nm and a SYBR filter.

Sugar disruption assays. The ability of various sugars (mannobiose, cellobiose, laminaribiose, maltose, sucrose, glucose lactose, galactose, arabinose, xylose) to disrupt LacI proteins binding to DNA fragments were tested in EMSAs. EMSAs were conducted using a Lightshift EMSA kit (Thermo Fisher Scientific, Waltham, MA, USA). Probes were made by using biotin-labeled primers (IDT) to create labeled probes (Table 3). Purified transcription factors were the same as those used in the fluorescent EMSAs. EMSAs were performed according to the manufacturer's instructions with each reaction mixture containing 1$\times$ binding buffer, 2.5% (vol/vol) glycerol, 5 mM MgCl$_2$, 50 ng $\cdot$ μl^{-1} poly(dI-dC), and 0.05% (vol/vol) NP-40. Each reaction mixture included 20 fmol of biotin-labeled probe and 15 nM purified transcription factor. Control competitive binding reaction mixtures included 4 pmol of unlabeled probe. Sugars were included in experiments at indicated quantities. Assays were incubated at room temperature for 30 min and analyzed on 4% polyacrylamide gels in 0.5% TBE buffer. EMSA gels were electroblotted onto Biodyne membranes and signal developed using the Chemiluminescent nucleic acid detection module kit (Thermo Fisher Scientific).

Accession number(s). Raw RNA-seq data have been deposited in NCBI SRA under accession number SRP057818 and gene expression data under NCBI GEO accession number GSE68423.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.02751-16.

TEXT S1, PDF file, 0.8 MB.
DATA SET S1, XLSX file, 0.01 MB.
DATA SET S2, XLSX file, 0.01 MB.
DATA SET S3, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Blumer-Schuette SE, Brown SD, Sander KB, Bayer EA, Kataeva I, Zurawski JV, Conway JM, Adams MW, Kelly RM. 2014. Thermophilic lignocellulose deconstruction. FEMS Microbiol Rev 38:393–448. https://doi.org/10.1111/1574-6976.12044.

2. Xu Q, Resch MG, Podkaminer K, Yang S, Baker JO, Donohoe BS, Wilson C, Klingeman DM, Olson DG, Decker SR, Giannone RJ, Hettich RL, Brown SD, Lynd LR, Bayer EA, Himmel ME, Bomble YJ. 2016. Dramatic performance of *Clostridium thermocellum* explained by its wide range of cellulase modalities. Sci Adv 2:e1501254. https://doi.org/10.1126/sciadv.1501254.

3. Lynd LR, van Zyl WH, McBride JE, Laser M. 2005. Consolidated bioprocessing of cellulosic biomass: an update. Curr Opin Biotechnol 16:577–583. https://doi.org/10.1016/j.copbio.2005.08.009.

4. Nataf Y, Yaron S, Stahl F, Lamed R, Bayer EA, Scheper TH, Sonenshein AL, Shoham Y. 2009. Cellodextrin and laminaribiose ABC transporters in *Clostridium thermocellum*. J Bacteriol 191:203–209. https://doi.org/10.1128/JB.01190-08.

5. Riederer A, Takasuka TE, Makino S-I, Stevenson DM, Bukhman YV, Elsen NL, Fox BG. 2011. Global gene expression patterns in *Clostridium ther-*

mocellum as determined by microarray analysis of chemostat cultures on cellulose or cellobiose. Appl Environ Microbiol 77:1243–1253. https://doi.org/10.1128/AEM.02008-10.

6. Yaniv O, Fichman G, Borovok I, Shoham Y, Bayer EA, Lamed R, Shimon LJ, Frolow F. 2014. Fine-structural variance of family 3 carbohydrate-binding modules as extracellular biomass-sensing components of *Clostridium thermocellum* anti-sigma factors. Acta Crystallogr D Biol Crystallogr 70:522–534. https://doi.org/10.1107/S139900471302926X.

7. Wilson CM, Rodriguez M, Jr, Johnson CM, Martin SL, Chu TM, Wolfinger RD, Hauser LJ, Land ML, Klingeman DM, Syed MH, Ragauskas AJ, Tschaplinski TJ, Mielenz JR, Brown SD. 2013. Global transcriptome analysis of *Clostridium thermocellum* ATCC 27405 during growth on dilute acid pretreated *Populus* and switchgrass. Biotechnol Biofuels 6:179. https://doi.org/10.1186/1754-6834-6-179.

8. Raman B, Pan C, Hurst GB, Rodriguez M, Jr, McKeown CK, Lankford PK, Samatova NF, Mielenz JR. 2009. Impact of pretreated switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis. PLoS One 4:e5271. https://doi.org/10.1371/journal.pone.0005271.

9. Bagnara-Tardif C, Gaudin C, Belaich A, Hoest P, Citard T, Belaich JP. 1992. Sequence analysis of a gene cluster encoding cellulases from *Clostridium cellulolyticum*. Gene 119:17–28. https://doi.org/10.1016/0378-1119(92)90062-T.

10. Kakiuchi M, Isui A, Suzuki K, Fujino T, Fujino E, Kimura T, Karita S, Sakka K, Ohmiya K. 1998. Cloning and DNA sequencing of the genes encoding *Clostridium josui* scaffolding protein CipA and cellulase CelD and identification of their gene products as major components of the cellulosome. J Bacteriol 180:4303–4308.

11. Newcomb M, Millen J, Chen CY, Wu JH. 2011. Co-transcription of the *celC* gene cluster in *Clostridium thermocellum*. Appl Microbiol Biotechnol 90:625–634. https://doi.org/10.1007/s00253-011-3121-x.

12. Nolling J, Breton G, Omelchenko MV, Makarova KS, Zeng Q, Gibson R, Lee HM, Dubois J, Qiu D, Hitti J, Wolf YI, Tatusov RL, Sabathe F, Doucette-Stamm L, Soucaille P, Daly MJ, Bennett GN, Koonin EV, Smith DR. 2001. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. J Bacteriol 183:4823–4838. https://doi.org/10.1128/JB.183.16.4823-4838.2001.

13. Sabathe F, Belaich A, Soucaille P. 2002. Characterization of the cellulolytic complex (cellulosome) of *Clostridium acetobutylicum*. FEMS Microbiol Lett 217:15–22. https://doi.org/10.1111/j.1574-6968.2002.tb11450.x.

14. Schwarz WH. 2001. The cellulosome and cellulose degradation by anaerobic bacteria. Appl Microbiol Biotechnol 56:634–649. https://doi.org/10.1007/s002530100710.

15. Tamaru Y, Doi RH. 2000. The *engL* gene cluster of *Clostridium cellulovorans* contains a gene for cellulosomal *manA*. J Bacteriol 182:244–247. https://doi.org/10.1128/JB.182.1.244-247.2000.

16. Tamaru Y, Karita S, Ibrahim A, Chan H, Doi RH. 2000. A large gene cluster for the *Clostridium cellulovorans* cellulosome. J Bacteriol 182:5906–5910. https://doi.org/10.1128/JB.182.20.5906-5910.2000.

17. Nataf Y, Bahari L, Kahel-Raifer H, Borovok I, Lamed R, Bayer EA, Sonenshein AL, Shoham Y. 2010. *Clostridium thermocellum* cellulosomal genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors. Proc Natl Acad Sci U S A 107:18646–18651. https://doi.org/10.1073/pnas.1012175107.

18. Sander K, Wilson CM, Rodriguez M, Klingeman DM, Rydzak T, Davison BH, Brown SD. 2015. *Clostridium thermocellum* DSM 1313 transcriptional responses to redox perturbation. Biotechnol Biofuels 8:211. https://doi.org/10.1186/s13068-015-0394-9.

19. Newcomb M, Chen CY, Wu JH. 2007. Induction of the *celC* operon of *Clostridium thermocellum* by laminaribiose. Proc Natl Acad Sci U S A 104:3747–3752. https://doi.org/10.1073/pnas.0700087104.

20. Bahari L, Gilad Y, Borovok I, Kahel-Raifer H, Dassa B, Nataf Y, Shoham Y, Lamed R, Bayer EA. 2011. Glycoside hydrolases as components of putative carbohydrate biosensor proteins in *Clostridium thermocellum*. J Ind Microbiol Biotechnol 38:825–832. https://doi.org/10.1007/s10295-010-0848-9.

21. Kahel-Raifer H, Jindou S, Bahari L, Nataf Y, Shoham Y, Bayer EA, Borovok I, Lamed R. 2010. The unique set of putative membrane-associated anti-sigma factors in *Clostridium thermocellum* suggests a novel extracellular carbohydrate-sensing mechanism involved in gene regulation. FEMS Microbiol Lett 308:84–93. https://doi.org/10.1111/j.1574-6968.2010.01997.x.

22. Mearls EB, Lynd LR. 2014. The identification of four histidine kinases that influence sporulation in *Clostridium thermocellum*. Anaerobe 28:109–119. https://doi.org/10.1016/j.anaerobe.2014.06.004.

23. Ravcheev DA, Khoroshkin MS, Laikova ON, Tsoy OV, Sernova NV, Petrova SA, Rakhmaninova AB, Novichkov PS, Gelfand MS, Rodionov DA. 2014. Comparative genomics and evolution of regulons of the LacI-family transcription factors. Front Microbiol 5:294. https://doi.org/10.3389/fmicb.2014.00294.

24. Papanek B, Biswas R, Rydzak T, Guss AM. 2015. Elimination of metabolic pathways to all traditional fermentation products increases ethanol yields in *Clostridium thermocellum*. Metab Eng 32:49–54. https://doi.org/10.1016/j.ymben.2015.09.002.

25. Argyros DA, Tripathi SA, Barrett TF, Rogers SR, Feinberg LF, Olson DG, Foden JM, Miller BB, Lynd LR, Hogsett DA, Caiazza NC. 2011. High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes. Appl Environ Microbiol 77:8288–8294. https://doi.org/10.1128/AEM.00646-11.

26. Tripathi SA, Olson DG, Argyros DA, Miller BB, Barrett TF, Murphy DM, McCool JD, Warner AK, Rajgarhia VB, Lynd LR, Hogsett DA, Caiazza NC. 2010. Development of *pyrF*-based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a *pta* mutant. Appl Environ Microbiol 76:6591–6599. https://doi.org/10.1128/AEM.01484-10.

27. Biswas R, Prabhu S, Lynd LR, Guss AM. 2014. Increase in ethanol yield via elimination of lactate production in an ethanol-tolerant mutant of *Clostridium thermocellum*. PLoS One 9:e86389. https://doi.org/10.1371/journal.pone.0086389.

28. Biswas R, Zheng T, Olson DG, Lynd LR, Guss AM. 2015. Elimination of hydrogenase active site assembly blocks H_2 production and increases ethanol yield in *Clostridium thermocellum*. Biotechnol Biofuels 8:20. https://doi.org/10.1186/s13068-015-0204-4.

29. Tian L, Papanek B, Olson DG, Rydzak T, Holwerda EK, Zheng T, Zhou J, Maloney M, Jiang N, Giannone RJ, Hettich RL, Guss AM, Lynd LR. 2016. Simultaneous achievement of high ethanol yield and titer in *Clostridium thermocellum*. Biotechnol Biofuels 9:116. https://doi.org/10.1186/s13068-016-0528-8.

30. Lin PP, Mi L, Morioka AH, Yoshino KM, Konishi S, Xu SC, Papanek BA, Riley LA, Guss AM, Liao JC. 2015. Consolidated bioprocessing of cellulose to isobutanol using *Clostridium thermocellum*. Metab Eng 31:44–52. https://doi.org/10.1016/j.ymben.2015.07.001.

31. Olson DG, Lynd LR. 2012. Transformation of *Clostridium thermocellum* by electroporation. Methods Enzymol 510:317–330. https://doi.org/10.1016/B978-0-12-415931-0.00017-3.

32. Mearls EB, Izquierdo JA, Lynd LR. 2012. Formation and characterization of non-growth states in *Clostridium thermocellum*: spores and L-forms. BMC Microbiol 12:180. https://doi.org/10.1186/1471-2180-12-180.

33. Sullivan JT, Brown SD, Ronson CW. 2013. The NifA-RpoN regulon of *Mesorhizobium loti* strain R7A and its symbiotic activation by a novel LacI/GalR-family regulator. PLoS One 8:e53762. https://doi.org/10.1371/journal.pone.0053762.

34. Mohnen D, Bar-Peled M, Somerville C. 2009. Cell wall polysaccharide synthesis, p 94–187. *In* Himmel ME (ed), Biomass recalcitrance. Blackwell Publishing Ltd., Oxford, England. https://doi.org/10.1002/9781444305418.ch5.

35. Mishra S, Beguin P, Aubert JP. 1991. Transcription of *Clostridium thermocellum* endoglucanase genes *celF* and *celD*. J Bacteriol 173:80–85. https://doi.org/10.1128/jb.173.1.80-85.1991.

36. Raman B, McKeown CK, Rodriguez M, Jr, Brown SD, Mielenz JR. 2011. Transcriptomic analysis of *Clostridium thermocellum* ATCC 27405 cellulose fermentation. BMC Microbiol 11:134. https://doi.org/10.1186/1471-2180-11-134.

37. Mearls EB, Olson DG, Herring CD, Lynd LR. 2015. Development of a regulatable plasmid-based gene expression system for *Clostridium thermocellum*. Appl Microbiol Biotechnol 99:7589–7599. https://doi.org/10.1007/s00253-015-6610-5.

38. Zverlov VV, Fuchs KP, Schwarz WH, Velikodvorskaya GA. 1994. Purification and cellulosomal localization of *Clostridium thermocellum* mixed linkage β-1,4-β-D-glucanase (LicB) (1,3-1,4-β-D-glucanase). Biotechnol Lett 16:29–34. https://doi.org/10.1007/BF01022619.

39. Schimming S, Schwarz WH, Staudenbauer WL. 1991. Properties of a thermoactive beta-1,3-1,4-glucanase (lichenase) from *Clostridium thermocellum* expressed in *Escherichia coli*. Biochem Biophys Res Commun 177:447–452. https://doi.org/10.1016/0006-291X(91)92004-4.

40. Schimming S, Schwarz WH, Staudenbauer WL. 1992. Structure of the *Clostridium thermocellum* gene *licB* and the encoded beta-1,3-1,4-glucanase. A catalytic region homologous to *Bacillus* lichenases joined

to the reiterated domain of clostridial cellulases. Eur J Biochem 204: 13–19.

41. Maniatis T, Fritsch EF, Sambrook J. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

42. Shanks RMQ, Kadouri DE, MacEachran DP, O'Toole GA. 2009. New yeast recombineering tools for bacteria. Plasmid 62:88–97. https://doi.org/10.1016/j.plasmid.2009.05.002.

43. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:-343–U341. https://doi.org/10.1038/nmeth.1318.

44. Guss AM, Olson DG, Caiazza NC, Lynd LR. 2012. Dcm methylation is detrimental to plasmid transformation in *Clostridium thermocellum*. Biotechnol Biofuels 5:30. https://doi.org/10.1186/1754-6834-5-30.

45. Kridelbaugh DM, Nelson J, Engle NL, Tschaplinski TJ, Graham DE. 2013.

Nitrogen and sulfur requirements for *Clostridium thermocellum* and *Caldicellulosiruptor bescii* on cellulosic substrates in minimal nutrient media. Bioresour Technol 130:125–135. https://doi.org/10.1016/j.biortech.2012.12.006.

46. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.

47. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

48. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8.