Applied and Environmental Microbiology, Dec. 2011, p. 8288–8294 Vol. 77, No. 23
0099-2240/11/$12.00 doi:10.1128/AEM.00646-11


# High Ethanol Titers from Cellulose by Using Metabolically Engineered Thermophilic, Anaerobic Microbes▽†‡

D. Aaron Argyros,[1]§ Shital A. Tripathi,[1]§¶ Trisha F. Barrett,[1] Stephen R. Rogers,[1] Lawrence F. Feinberg,[1] Daniel G. Olson,[1,2] Justine M. Foden,[1] Bethany B. Miller,[1] Lee R. Lynd,[1,2] David A. Hogsett,[1]* and Nicky C. Caiazza[1]#

*Mascoma Corporation, 67 Etna Road, Suite 300, Lebanon, New Hampshire 03766,[1] and Thayer School of Engineering, Dartmouth College, 8000 Cummings Hall, Hanover, New Hampshire 03755[2]*

Received 18 March 2011/Accepted 20 September 2011

**This work describes novel genetic tools for use in *Clostridium thermocellum* that allow creation of unmarked mutations while using a replicating plasmid. The strategy employed counter-selections developed from the native *C. thermocellum hpt* gene and the *Thermoanaerobacterium saccharolyticum tdk* gene and was used to delete the genes for both lactate dehydrogenase (Ldh) and phosphotransacetylase (Pta). The Δ*ldh* Δ*pta* mutant was evolved for 2,000 h, resulting in a stable strain with 40:1 ethanol selectivity and a 4.2-fold increase in ethanol yield over the wild-type strain. Ethanol production from cellulose was investigated with an engineered coculture of organic acid-deficient engineered strains of both *C. thermocellum* and *T. saccharolyticum*. Fermentation of 92 g/liter Avicel by this coculture resulted in 38 g/liter ethanol, with acetic and lactic acids below detection limits, in 146 h. These results demonstrate that ethanol production by thermophilic, cellulolytic microbes is amenable to substantial improvement by metabolic engineering.**

Lignocellulosic biomass is the most abundant feedstock for biofuel production, and at $50/metric ton it is one of the cheapest, equivalent on an energy basis to oil at $17/barrel (16). However, a significant cost is attributed to the saccharolytic enzymes required to release soluble sugars from lignocellulose (16). In order to match the current economics of starch hydrolysis from corn kernels, the cost of cellulase enzymes would need to decrease nearly an order of magnitude (1). A promising solution to reduce enzyme cost is consolidated bioprocessing (CBP), which requires an organism or collection of organisms capable of both releasing and fermenting cellulosic sugars without added enzymes and at high ethanol yield (17).

Expression of saccharolytic enzymes has been demonstrated in microorganisms with established genetic systems such as *Saccharomyces cerevisiae* (8, 15, 29, 32), but the overall enzymatic activity has not rivaled that of natively cellulolytic organisms. Alternatively, metabolic engineering has produced high-yield ethanol fermentation in model organisms that can be genetically manipulated (24, 27, 33). Therefore, metabolic engineering of a naturally cellulolytic microbe should make it possible to create a strain that can both hydrolyze and ferment the products of cellulose hydrolysis into ethanol as the major fermentation end product.

An attractive candidate for metabolic engineering is the thermophilic anaerobe *Clostridium thermocellum*, which is capable of hydrolyzing cellulose at rates approaching 2.5 g/liter/h (10). This high rate of cellulose hydrolysis results from the activity of a multiprotein complex called the cellulosome. This cell-associated, organelle-like appendage has been extensively reviewed and contains an assemblage of structural proteins and enzymes designed to access, bind, and hydrolyze cellulose (2, 6). It is widely recognized that development of *C. thermocellum* as a CBP organism has been limited by the genetic tools required to create a stable strain with high ethanol-to-organic acid ratios (1, 18). Recent reports have described development of basic genetic tools for *C. thermocellum*, including transformation of plasmid DNA and the ability to make marked gene deletions (19, 28, 30). However, these tools do not provide the versatility associated with creating unmarked genetic mutations, and while they could be used to eliminate organic acid production, they would not be ideal. In particular, these approaches rely on replacement of gene targets with antibiotic resistance cassettes, limiting the number of gene deletions to the available number of antibiotic resistance markers. Furthermore, this strategy can be performed only using the *C. thermocellum* Δ*pyrF* strain which has a growth defect and is not an optimal background for future genetic manipulation (28). The development of additional genetic tools, in particular, counter-selectable markers, is thus desirable.

Nucleic acid metabolism has been the basis for numerous counter-selectable markers. The cellular activity of hypoxanthine phosphoribosyl transferase (Hpt) reassimilates purines such as hypoxanthine, xanthine, and guanine for the purpose of DNA and RNA synthesis (25) but can lead to cellular toxicity in the presence of purine antimetabolites such as 8-azahypoxanthine (AZH). Recently *hpt* was developed into a useful genetic marker for counter-selection in *Archaea* (20). The cellular toxicity of fluoro-deoxyuracil (FUDR) is dependent on

* Corresponding author. Mailing address: Mascoma Corporation, 67 Etna Road, Suite 300, Lebanon, NH 03766. Phone: (603) 676-3320, ext. 1103. Fax: (603) 676-3321. E-mail: dhogsett@mascoma.com.
¶ Present address: Total Gas and Power, 5858 Horton Street, Suite 253, Emeryville, CA 94608.
# Present address: Synthetic Genomics, 11149 North Torrey Pines Road, LaJolla, CA 92037.
§ D.A.A. and S.A.T. contributed equally to this work.
† Supplemental material for this article may be found at http://aem.asm.org/.
▽ Published ahead of print on 30 September 2011.
‡ The authors have paid a fee to allow immediate free access to this article.

TABLE 1. Plasmids and strains used in this study

| Plasmid or strain | Description and/or relevant characteristics[a] | Source or reference |
|---|---|---|
| Plasmids | | |
| pMU1647 | Cloning vector; contains *cpb*p-*tdk* and *gapDH*p-*cat-hpt* operon | This study |
| pMU1657 | HPT KO vector with *cpb*p-*hpt* and *gapDH-cat* outside deletion flank | This study |
| pMU1758 | Lactate dehydrogenase deletion vector | This study |
| pMU1777 | Lactate dehydrogenase deletion vector without yeast machinery | This study |
| pMU1817 | Phosphotransacetylase deletion vector | This study |
| Strains | | |
| *C. thermocellum* | | |
| M0003 | Wild-type DSM 1313 | DSMZ[b] |
| M1354 | DSM 1313 Δ*hpt* | This study |
| M1375 | M1354 Δ*ldh* | This study |
| M1448 | M1354 Δ*pta* | This study |
| M1434 | M1375 Δ*pta* | This study |
| M1570 | M1434 evolved | This study |
| *T. saccharolyticum* | | |
| ALK2 | Δ*ldh* Δ*pta-ack* | 24 |

[a] HPT, hypoxanthine phosphoribosyl transferase; KO, knockout.
[b] Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Germany.

the presence of two enzymes involved in pyrimidine metabolism: thymidine kinase (Tdk) and thymidylate synthetase (ThyA). Tdk converts FUDR to fluoro-dUMP (F-dUMP) which is a covalent inhibitor of ThyA and the basis for counterselection in a variety of eukaryotic organisms (5, 9, 12, 26).

*C. thermocellum* uses the Embden-Myerhof pathway to ferment sugars to pyruvate, with branched end product metabolism resulting in formation of ethanol, acetic acid, and lactic acid (13). As reviewed elsewhere (17) a substantial effort has been devoted to strain isolation, optimization of culture conditions (14), and strain development via mutation and selection for improved thermophilic fermentation of cellulose and/or xylose to ethanol at high yields. However, these efforts did not result in robust strains that consistently produce ethanol at high yields under a broad range of conditions and in the hands of different investigators. Metabolic engineering using molecular techniques has been applied to manipulate end product metabolism toward increased ethanol yield in the noncellulolytic, thermophilic bacterium *Thermoanaerobacterium saccharolyticum* (23, 24) but not in the highly cellulolytic *C. thermocellum*. The benefit of a coculture of these two organisms would result from a combination of the cellulose hydrolysis by *C. thermocellum* with the broad sugar utilization and high ethanol yield of *T. saccharolyticum.* Prior efforts along these lines have been recently reviewed (6), but until now organic acid production by *C. thermocellum* has inhibited growth and decreased ethanol yield. In this study, we show that ethanol production by *C. thermocellum* was amenable to substantial improvement by metabolic engineering and demonstrate the potential of a coculture between ethanologenic versions of these two organisms.

## MATERIALS AND METHODS

**Molecular techniques.** All plasmid constructions in this work were created using yeast gap repair cloning. A detailed description of the particular techniques and yeast strains used for plasmid construction is given by Shanks et al. (22). A complete list of strains, plasmids, and primers used in this study to generate knockouts is given in Tables 1 and 2. Primer design was based on the *C. thermocellum* 27405 genome (http://www.ncbi.nlm.nih.gov). Transformation was performed as described in Tripathi et al. (28).

**Hpt deletion vector (pMU1657).** A sequence including 65 bp of coding sequence and 1,121 bp upstream of *hpt* was amplified using primers X107384 and X07385, and a sequence including the 10 bp of coding sequence and 1,025 bp

TABLE 2. Oligonucleotides used in this study

| Primer | Sequence 5′–3′ |
|---|---|
| XO7384 | TATATTTTTTAGTCCATATCTTCTTTGTCCGTATAAACAAGTTCCTCTCTGGTAACCAA |
| XO7385 | CGTGTAAGTTACAGGCAAGCGATCGCGGCCGCGGTACGAGGCACAGGTCTTGACGGACT |
| XO7383 | TAAAGAAATTTTGGTTACCAGAGAGGAACTTGTTTATACGGACAAAGAAGATATGGACT |
| XO7503 | ACTAGGGCTCGCCTTTGGGAAGTTTGAAGGGCTACCGTCCATTTCAACCAACAGCTGAT |
| XO8205 | GAATCTTTTCCTCTCTTTCGGAAAAGAAATACAATTTCTTCATCGTTGAAAGGCACGTT |
| XO8206 | GTGTAGAAAGTGCCATGAAGTCCCGCGGACTTAAGGTTCCACCACAGCTTATACATTGA |
| XO8207 | CTTTTAGAGTGTTTCCGGACTTTCTGAGAAGCTGTACAAAGCCTGCACCAACTACGGTT |
| XO8208 | TATATTGCTATAAAGAATGAGGAGGGAACTAGTTGAATGACACTATCCTGTATCCTGAT |
| XO8209 | TACCCGGGGATCCTCTAGAGTCGACCTGCAGCCATGCCTGGGAGGCTCTGTATAGAGAA |
| XO8210 | AAGTAACCGTAGTTGGTGCAGGCTTTGTACAGCTTCTCAGAAAGTCCGGAAACACT |
| XO9013 | AGAATCTTTTCCTCTCTTTCGGAAAAGAAATACCAATGGCGGCATCCACCTGAAGTTCT |
| XO9014 | AGGTGTAGAAAGTGCCATGAAGTCCCGCGGACTTAAGGGAATTGCAAAGGTTGTACTGA |
| XO9015 | GTACCCGGGGATCCTCTAGAGTCGACCTGCAGTGTCATCTCCCTTTTCTGCGGCATCCT |
| XO9016 | GAAAGTACGGATCTGAGGGTTATTAAAGCCGCCGGTTCAGGCTCAATATGTCAAGGCAT |
| XO9017 | AATTTATGCCTTGACATATTGAGCCTGAACCGGCGGCTTTAATAACCCTCAGATCCGTA |
| XO9018 | AGGTATCGTTATATGGATACTGATAATTATCGCCAGGCAAAGTCCAACTATGCATTGGT |

downstream of *hpt* was amplified using primers X07503 and X07383. These fragments were cloned by gap repair into AatII-digested pMU749 (28) along with a cassette containing the *hpt* gene expressed from the *C. thermocellum* cellobiose phosphorylase promoter and an antibiotic resistance cassette consisting of the *cat* gene from pNW33N fused to the *C. thermocellum* glyceraldehyde 3-phosphate dehydrogenase promoter.

**Deletion vector backbone (pMU1647).** The entire annotated sequence of plasmid pMU1647 is described pictorially Fig. SA1 in the supplemental material. An operon was created to link the positive (*cat*) and negative (*hpt*) selectable markers. A 15-bp segment located between the first two genes of the *C. thermocellum* 1313 urease operon was used as a spacer and ribosome binding site for the *hpt* gene. To facilitate subsequent cloning of homologous flanks used for targeted deletions, the *gapDH*p-*cat-hpt* operon was flanked by FspI restriction sites. For counter-selection against the plasmid and selection for chromosomal integration of the operon, the *tdk* gene was amplified from *T. saccharolyticum* and fused to the *C. thermocellum* cellobiose phosphorylase promoter.

**Lactate dehydrogenase (LDH) deletion vector (pMU1758).** To create plasmid pMU1758, plasmid pMU1647 was digest with FspI, which cuts immediately upstream and downstream of the *gapDH*p-*cat-hpt* cassette. An 840-bp fragment located inside the lactate dehydrogenase (*ldh*) open reading frame (ORF) was amplified from *C. thermocellum* genomic DNA using primers X08205 and X08206 with tails homologous to vector sequence surrounding the upstream FspI site in pMU1647. An 824-bp region upstream of the *ldh* ORF was amplified using primers X08207 and X08208, and a 952-bp region downstream of the *ldh* ORF was amplified using primers X08209 and X08210. These two fragments had homologous tails for each other and the sequence surrounding the downstream FspI site. The yeast origins and selectable markers were removed by digestion of pMU1758 with NotI, gel purification, and ligation of remaining plasmid to create pMU1777.

**Phosphotransacetylase (PTA) deletion vector.** To create plasmid pMU1817, plasmid pMU1647 was digested with FspI. A 632-bp fragment located inside the *pta* ORF was amplified from *C. thermocellum* 1313 genomic DNA using primers X09013 and X09014, which contain tails homologous to vector sequence surrounding the upstream FspI. An 843-bp region upstream of the *pta* ORF was amplified using primers X09015 and X0915, and a 1,035-bp region downstream of the *pta* ORF was amplified using primers X09017 and X09018 with homologous tails for each other and the sequence surrounding the downstream FspI site.

**Medium and culture conditions.** Routine growth of *C. thermocellum* for genetic manipulations was done at 55°C using modified DSMZ medium 122 containing 5 g/liter cellobiose. The modifications included a reduction in the $K_2HPO_4$ concentration to 1.8 g/liter, and the glutathione was replaced with 0.5 g/liter cysteine-HCl. Analysis of mutant strains was done in CM3 medium (3) containing Avicel at concentrations indicated in the figure legends. The inocula for all fermentations were taken from an overnight culture in CM3 medium containing 10 g/liter Avicel to induce cellulosome production, with the exception of those used in the batch reactor, which where pregrown in chemically defined MTC medium (34–36).

**Evolution of Δ*hpt* Δ*ldh* Δ*pta* strain.** The initial Δ*hpt* Δ*pta* Δ*ldh* strain was evolved for faster growth by serial transfer in CM3 medium containing 10 g/liter Avicel for 10 transfers, 20 g/liter for the second 20 transfers, 40 g/liter for the third 20 transfers, and 60 g/liter for a final 10 transfers. Transfers were by done by subculturing at a final dilution of 1:100 every 24 to 48 h.

**Antimetabolite selection media.** Plating and selection against the *hpt* gene was done using a final concentration of 500 μg/ml 8-azahypoxanthine (202590050; Acros Organics) in defined DSMZ medium 122, which contained previously described vitamin concentrations as a substitute for yeast extract (11). Selection against *tdk* was done in modified DSMZ medium 122 containing 10 μg/ml 5-fluoro-2′-deoxy-uridine (F0503; Sigma). All selections were done using cultures grown to mid-log phase.

**Cocultures.** *C. thermocellum* and *T. saccharolyticum* strains were initially grown for 18 to 20 h at 55°C in monoculture on CM3 medium containing 25 g/liter cellobiose at pH 6.8. The cultures were transferred twice in this manner prior to inoculation. Coculture experiments were conducted using duplicate 50-ml final volumes of CM3 containing ~17.2 g/liter Avicel and 5 g/liter calcium carbonate at pH 6.8 in 160-ml serum bottles purged with 5% carbon dioxide–95% nitrogen gas mix prior to autoclaving. A 5% inoculum of each strain was introduced at the outset. Cultures were incubated at 55°C and shaken at 300 rpm. Well-mixed samples were taken aseptically using a 3-ml syringe fitted with a 21-gauge needle.

For reactor batch cultures, inocula for *C. thermocellum* and *T. saccharolyticum* strains were initially grown for 18 to 24 h at 55°C in monoculture on MTC medium containing, respectively, 10 g/liter Avicel or cellobiose. The coculture reactor was initiated using a 10% inoculum of each strain into a 2-liter Sartorius Biostat Bplus bioreactor (Sartorius BBI Systems, Inc., Bethlehem, PA) with a 1-liter working volume containing MTC medium supplemented with 5 g/liter calcium carbonate and 100 g/liter (wet weight) Avicel. Incubations were conducted anaerobically at 55°C with stirring at 300 rpm and a starting pH 6.3. Well-mixed samples were taken periodically from the sampling port using a syringe.

**Analytical techniques.** Residual cellulose from Avicel was estimated gravimetrically and used to compute carbon balance as follows. Mixed 1.5-ml culture samples were centrifuged (10,000 × *g*) in 2-ml tubes for 3 min, and the supernatants were discarded. To solubilize residual carbonate, 1 ml of a 10% (wt/vol) citric acid solution was added and allowed to stand for 5 min after thorough mixing. The samples were then centrifuged and washed three times in 1 ml of distilled water as described above prior to transfer to aluminum weighing tins. Drying overnight at 55 to 60°C was followed by 45 min at 105°C prior to weighing. The substrate concentrations were then computed based on the processed sample volumes. Yield (carbon balance) was calculated as the ratio of products to residual substrate (g/liter glucose equivalents) in the fermentation samples according to the following equation: yield = [(1.053 Cb) + Glu + LA + (AA/0.667) + (EtOH/0.511)]/[1.111($S_0 - S_f$)], where Cb, Glu, LA, AA, and EtOH are the concentrations (g/liter) of cellobiose, glucose, lactic acid, acetic acid, and ethanol, respectively, in the fermentation broth as determined by high-performance liquid chromatography (HPLC) and $S_0$ and $S_f$ are the initial and final substrate concentrations, respectively, measured as above. This equation includes carbon contributions from carbon dioxide, accounted for by stoichiometric correlation to ethanol and acetic acid formation, and cells, included with residual substrate. The maximum difference of computing converted substrate under the latter assumption is ±2%, which is well within the sampling error (~5%) at low substrate concentrations. Fermentation metabolites were analyzed by HPLC methods described in Shaw et al. (24).

**Nucleotide sequence accession number.** The entire annotated sequence of plasmid pMU1647 was deposited in the GenBank database under accession number HQ264098.

## RESULTS

**Development of selectable markers for use in *C. thermocellum*.** Plating experiments revealed that *C. thermocellum* is sensitive to AZH (see Table SA1 in the supplemental material), and subsequent genetic analysis confirmed that open reading frame (ORF) Cthe2254 is an *hpt* homolog. Figure 1a illustrates deletion of *hpt* from *C. thermocellum* using AZH counter-selection. PCR analysis (Fig. 1b) of the *hpt* locus indicated that AZH-resistant colonies contained an in-frame deletion of *hpt* that was confirmed by DNA sequencing. A cassette was constructed on pMU1647 (see Fig. SA1 in the supplemental material) that contained the *C. thermocellum hpt* gene cloned in an operon with *cat* expressed from the *C. thermocellum gapDH* promoter (*gapDH*p). The functionality of the cassette was verified by transforming the Δ*hpt* strain with pMU1647, selecting on thiamphenicol (TM), and demonstrating that the resulting transformants were sensitive to AZH while the parental Δ*hpt* strain was not (see Table SA1).

Analysis of the *C. thermocellum* genome indicated that ORF Cthe1227 is a *thyA* homolog, but a homolog of *tdk* could not be found. Consistent with the absence of *tdk*, *C. thermocellum* was insensitive to FUDR (see Table SA1). *T. saccharolyticum tdk* was expressed from the *C. thermocellum* cellobiose phosphorylase promoter (*cbp*p) and transformed into *C. thermocellum* using pMU1647. Counter-selection in the presence of FUDR was evident in the form of a 96% reduction in the CFU count (see T SA1).

**Deletion of genes responsible for organic acid formation by using a removable marker system.** Deletion of *ldh* and *pta*, encoding lactate dehydrogenase (Ldh) and phosphotransacetylase (Pta), respectively, was undertaken pursuant to develop-

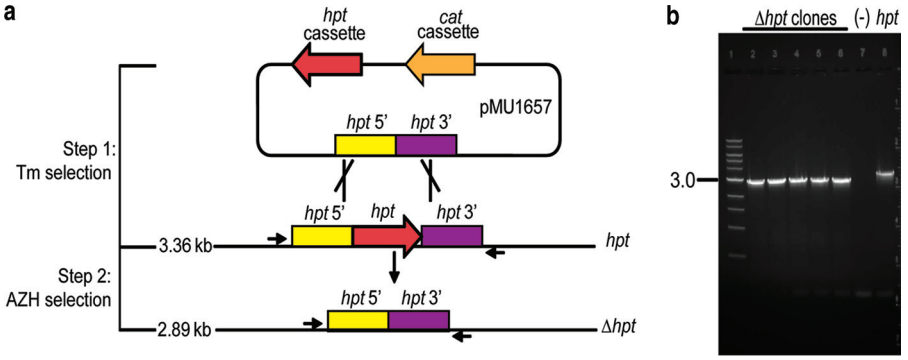


FIG. 1. Deletion of the *C. thermocellum hpt* gene. (a) Cartoon schematic illustrating the deletion of *hpt*. The plasmid used to delete *hpt*, pMU1657, contained the following: (i) a deletion cassette comprised of *hpt* 5′ and 3′ flanking DNA; (ii) a *cat* cassette, with *cat* driven by *C. thermocellum gapDH* promoter encoding Tm resistance; (iii) an *hpt* cassette, with *hpt* driven by the *C. thermocellum cbp* promoter encoding sensitivity to AZH. Step 1 depicts plasmid transformation. *C. thermocellum* transformants were selected using Tm resistance encoded by *cat* cassette. Step 2 depicts plasmid curing and deletion of chromosomal copy of *hpt.* Plating on AZH selected for plasmid loss, mediated by the *hpt* cassette and deletion of chromosomal *hpt* by homologous recombination. Depiction of chromosomal regions, as labeled on the right, indicate primer binding sites (→) used in diagnostic PCR. Expected amplicon size is shown to the left. (b) DNA gel showing results of diagnostic PCR at the *hpt* locus after step 2. Gel image contains a 1-kb ladder (lane 1), diagnostic PCR of clones subjected to selection conditions (lanes 2 to 6), no-DNA negative control (lane 7), and *C. thermocellum* 1313 genomic DNA used as a template to amplify the wild-type *hpt* locus as a positive control (lane 8).

ment of a strain of *C. thermocellum* with high ethanol yield. Figure 2a illustrates deletion of *ldh* using the removable marker system described above. After step 2, Tm- and FUDR-resistant colonies were screened by PCR, and the dominant product was 5.7 kb, consistent with insertion of the integration cassette onto the chromosome (Fig. 2b). After step 3, AZH-resistant colonies were screened by PCR, and the dominant product was 2.0 kb, consistent with deletion of the integration cassette and *ldh* from the chromosome (Fig. 2c), an event that was confirmed by DNA sequencing. These selections were very stringent as positive results were observed with 95% and 100% of colonies screened ($n = 20$) after step 1 and step 2, respectively.

The strategy outlined in Fig. 2a was used to delete *pta* in both the Δ*hpt* and Δ*hpt* Δ*ldh* strains. Using the same primer set, PCR amplification shows a 1.0-kb difference between the

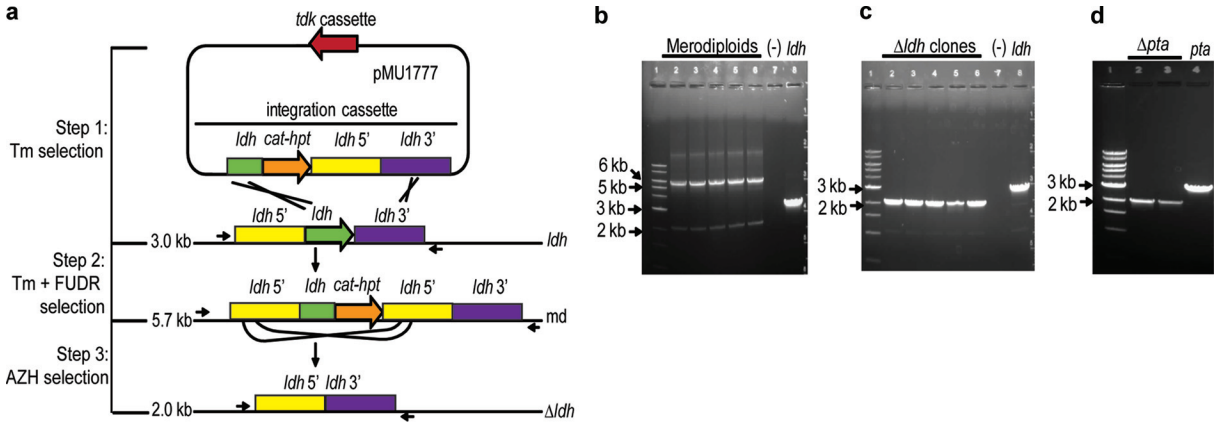


FIG. 2. Deletion of the *C. thermocellum ldh* and *pta* genes. (a) Cartoon schematic illustrating the deletion of *ldh*. The plasmid used to delete *ldh*, pMU1777, contains two major features: (i) *tdk* cassette, with *T. saccharolyticum tdk* driven by *C. thermocellum cbp* promoter, and (ii) an integration cassette, including an internal fragment of the *ldh* gene, transcriptional fusion of *cat* and *hpt* driven by the *C. thermocellum gapDH* promoter, and 5′ and 3′ flanking DNA labeled as *ldh* 5′ and *ldh* 3′. The first step consists of transforming the *C. thermocellum* Δ*hpt* strain with pMU1777 and selecting on Tm. The second step combines Tm and FUDR, which selects for recombination of the integration cassette onto the chromosome and for loss of the plasmid backbone containing the *tdk* cassette. This creates a merodiploid (md) strain in which the *ldh* 5′ flanking DNA is duplicated. The third step utilizes AZH counter-selection to isolate cells that have lost the *cat-hpt* cassette through recombination between the duplicated *ldh* 5′ flanking DNA, resulting in deletion of *ldh*. Chromosomal regions are labeled to the right and indicate primer binding sites (→) used in diagnostic PCR. Expected amplicon size is shown to the left. (b) DNA gel showing results of diagnostic PCR at the *ldh* locus after step 2. (c) DNA gel showing results of diagnostic PCR at the *ldh* locus after step 3. Both gel images contain a 1-kb DNA Ladder (lane 1), diagnostic PCR of clones subjected to selection conditions (lanes 2 to 6), no-DNA negative control (lane 7), and Δ*hpt* strain genomic DNA used as template to generate wild-type *ldh* locus as a positive control (lane 8). (d) Gel image showing molecular confirmation of Δ*hpt* Δ*pta* and Δ*hpt* Δ*ldh* Δ*pta* strains. The same selections as those shown in Fig. 2a were performed to delete *pta* in both the Δ*hpt* and Δ*hpt* Δ*ldh* strains. Lanes 2 to 4 show the results of diagnostic PCR at the *pta* locus for the Δ*hpt* Δ*pta* strain (lane 2), the Δ*hpt* Δ*ldh* Δ*pta* strain (lane 3), and the Δ*hpt* strain (lane 4). The expected size of the wild-type *pta* amplicon is 3.0 kb, while the Δ*pta* amplicon is 2.1 kb. The 1-kb DNA ladder is loaded in lane 1.

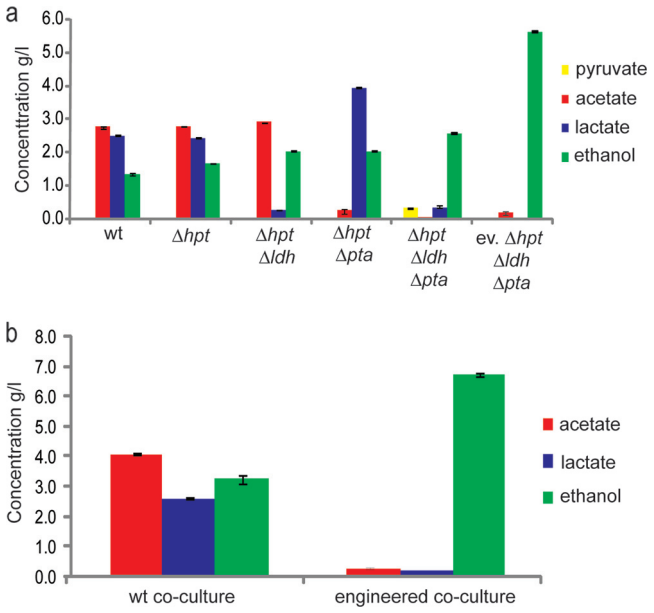


FIG. 3. Fermentation profiles of engineered *C. thermocellum* monocultures and coculture using both wild-type (wt) or engineered strains of *C. thermocellum* and *T. saccharolyticum*. (a) Batch fermentations were maintained at 55°C with initial pH of 7.0 and 19.5 g/liter Avicel. Data represent the averages and standard deviations of duplicate fermentations sampled at 72 h. An evolutionarily (ev) engineered version of the Δ*hpt* Δ*ldh* Δ*pta* strain was transferred in batch culture for 2,000 h. (b) Batch fermentations were maintained at 55°C with initial pH of 6.75 and 17.2 g/liter Avicel. Data represent the averages and standard deviations of duplicate fermentations sampled at 120 h. *C. thermocellum* ev. Δ*hpt* Δ*ldh* Δ*pta* and *T. saccharolyticum* ALK2 comprised the engineered coculture.

wild-type *pta* locus (3.0 kb) and that of the Δ*pta* strains (2.0 kb) (Fig. 2d). All deletions were confirmed by DNA sequencing.

**Fermentation analysis of *C. thermocellum* organic acid mutants.** Fermentation profiles support molecular evidence for the creation of Δ*hpt* Δ*ldh*, Δ*hpt* Δ*pta*, and Δ*hpt* Δ*ldh* Δ*pta* strains. Results in Fig. 3a show the product profile at the end of batch fermentations using 19.5 g/liter Avicel, a model microcrystalline cellulosic substrate. The Δ*hpt* strain, the parent of all genetically engineered strains in this study, produced acetate, lactate, and ethanol in a ratio of 1.7:1.5:1.0, similar to the 2.1:1.9:1.0 ratio produced by the wild type. The Δ*hpt* Δ*ldh* strain did not produce significant levels of lactate and had a 1.4:1.0 ratio of acetate to ethanol. Similarly, the Δ*hpt* Δ*pta* strain did not produced acetate and had a 1.9:1.0 ratio of lactate to ethanol. The Δ*hpt* Δ*ldh* Δ*pta* strain achieved ethanol selectivity of 40:1 relative to organic acids.

Acetate titer was not affected in the Δ*hpt* Δ*ldh* strain fermentation, indicating that flux through this node is not altered by deleting lactate dehydrogenase if there is capability to make ethanol, which increased by 23% (Fig. 3a). In contrast, the lactate titer increased 62% (Fig. 3a) in the Δ*hpt* Δ*pta* strain fermentation while the ethanol titer still only increased 23%, indicating a significant increase in flux from pyruvate to lactate if acetate production is prevented. Fermentation with the Δ*hpt* Δ*ldh* Δ*pta* strain, where cellular carbon flux is more constrained toward ethanol, showed the most dramatic effect, with a 56% increase in ethanol titer (Fig. 3a).

Pyruvate was detected in the Δ*hpt* Δ*ldh* Δ*pta* strain fermentation and could be an indication of a metabolic imbalance within the strain (Fig. 3a). Since metabolism and growth are related, a growth-based evolutionary engineering approach was taken to increase the efficiency of metabolism. The Δ*hpt* Δ*ldh* Δ*pta* strain was transferred in batch culture for 2,000 h and resulted in a strain that no longer secreted pyruvate and increased the ethanol titer ~4-fold relative to the wild type, achieving 5.61 g/liter (Fig. 3a) from consumption of 18.4 g/liter Avicel, corresponding to a yield of 0.27 g of ethanol/g of substrate.

**Fermentation analysis of an engineered *C. thermocellum* and *T. saccharolyticum* coculture.** *C. thermocellum* is classified as a mixed-acid fermenter; however, in laboratory fermentations a significant portion of the carbon balance remains open when only these end products are taken into account. Furthermore, additional fermentation products that close the carbon balance have not been routinely identified. For the Δ*hpt* strain, 73% of the carbon could be accounted for during Avicel fermentation (measurements used to calculate Avicel hydrolysis and carbon recovery for each strain are provided in Table SA2 in the supplemental material). The carbon recovery of the Δ*hpt* Δ*ldh* Δ*pta* strain dropped to 33% and through evolution was improved to 61%. To direct more cellulosic carbon to known end products, we cocultured *C. thermocellum* with *T. saccharolyticum*, a noncellulolytic mixed-acid fermenter which has two beneficial traits: it has previously been engineered for high ethanol yield, and all carbon can be accounted for following fermentation (24). Batch coculture fermentations containing 17.2 g/liter Avicel were conducted using wild-type or engineered versions of these two strains to make ethanol as the sole fermentation product. The fermentation profiles with respect to acetate, lactate, and ethanol are shown in Fig. 3b, and additional fermentation measurements addressing Avicel conversion and carbon recovery are found in Table SA3. During the course of both fermentations 95% to 96% of the Avicel was hydrolyzed, and the wild-type strain cocultures produced acetate, lactate, and ethanol with a 1.3:0.8:1.0 ratio; the engineered strain cocultures made predominantly ethanol, resulting in ~2-fold increase in titer at 6.7 g/liter (Fig. 3b). The wild-type coculture had a carbon recovery of 79%, which is a moderate improvement over the 72% recovery observed for wild-type *C. thermocellum*. However, the engineered coculture had a carbon recovery of 76%, a significant improvement over the 61% observed for the evolved Δ*hpt* Δ*ldh* Δ*pta* strain.

Figure 4 shows ethanol yield for all fermentations expressed as a percentage of the theoretical maximum based on Avicel hydrolysis (theoretical ethanol yields for each fermentation are reported in Tables SA2 and SA3 in the supplemental material). The Δ*hpt* strain is the reference for all engineered *C. thermocellum* strains and exhibited a small improvement over the wild type with respect to ethanol yield and achieved ~18% of theoretical maximum (Fig. 4). Compared to this strain, both the Δ*hpt* Δ*ldh* and Δ*hpt* Δ*pta* mutants did not drastically improve ethanol yield, while the Δ*hpt* Δ*ldh* Δ*pta* mutant improved yield by only ~33%, or ~27% of theoretical maximum. The evolved version of the Δ*hpt* Δ*ldh* Δ*pta* strain had a more significant improvement over the reference strain, increasing ethanol yield by ~70%, which is ~60% of theoretical maximum. Ethanol yield of the wild-type coculture was 35% of theoretical

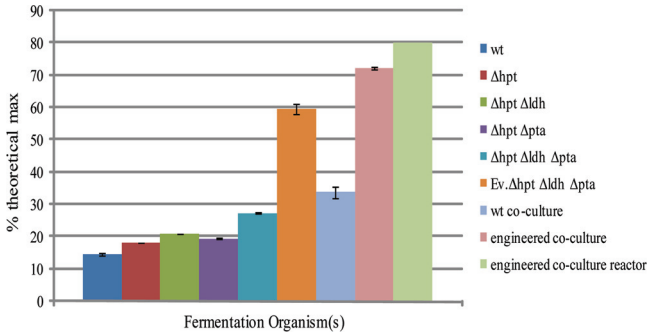


FIG. 4. Percent theoretical maximum (max) ethanol of fermentations performed in this study. Maximum theoretical ethanol yield was based on Avicel conversion and is listed in Tables SA2 and SA3 in the supplemental material. For a given strain or coculture, data represent averaged measured ethanol generated during fermentation divided by the theoretical maximum ethanol. The engineered coculture run in the reactor represents a single experiment.

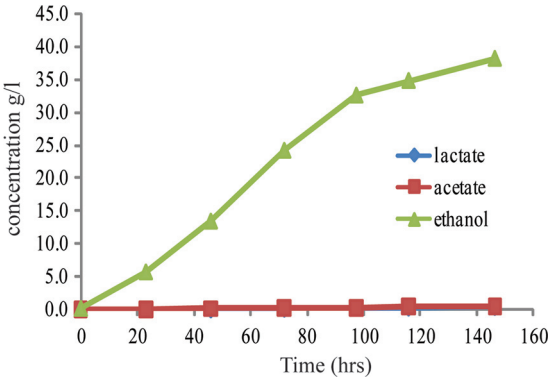


FIG. 5. High ethanol titer using the *C. thermocellum* evolutionarily (ev) engineered version of the Δ*hpt* Δ*ldh* Δ*pta* strain and in a *T. saccharolyticum* ALK2 coculture. A batch reactor with continuous stirring set at 300 rpm with an initial pH of 6.3 and 92.2 g/liter Avicel was maintained at 55°C. Data represent a single fermentation sampled over the course of 146 h.

maximum, an improvement of ~58% relative to wild-type *C. thermocellum*. This was more than doubled by using engineered strains in coculture fermentation, achieving ~75% of theoretical maximum.

In our hands the highest ethanol titer that could be achieved using the evolved Δ*hpt* Δ*ldh* Δ*pta* strain was 14 g/liter from 40 g/liter Avicel. To confirm that the limitation is not the cellulolytic capability of *C. thermocellum* but, rather, the metabolism of the organism, we tested higher substrate concentrations in the context of a coculture. A batch fermentation performed in a reactor with constant stirring and no pH control was used to test the ability of an engineered *C. thermocellum-T. saccharolyticum* coculture. Fermentation of 92.2 g/liter Avicel was carried out for 146 h, at which point pH became limiting, and ~90% of the Avicel had been hydrolyzed (see Table SA3 in the supplemental material). Throughout the fermentation, ethanol was the only fermentation product detected, reaching 38.1 g/liter (Fig. 5), which was ~80% of theoretical maximum (Fig. 4).

## DISCUSSION

A removable marker system for *C. thermocellum* was developed and used to delete genes responsible for organic acid formation. This development required information about host metabolism, which is available for a vast number of organisms in the form of annotated genome sequences that identify potential enzymatic reactions that might be targets for antimetabolites. We used this information to create counter-selections for use in *C. thermocellum* that are based on both endogenous (*hpt*) and exogenous (*tdk*) genes. Combining the markers allowed us to use replicating plasmids to insert and remove markers from the host chromosome, creating in-frame gene deletions in the process. Like *C. thermocellum*, many organisms have been isolated from nature because they exhibit interesting biology or industrial promise but cannot be studied through routine genetic manipulation. As we demonstrated with *C. thermocellum*, it is possible to overcome poor transformation efficiency and/or inadequate selections that hinder development of these organisms by exploiting information encoded within the host genome.

Low yield remains an issue for *C. thermocellum*; however, this was ameliorated through coculture with *T. saccharolyticum*. While the reason for the improved performance of the coculture is unclear, it may be that *C. thermocellum* secretes undetected metabolites that can be consumed by the coculture partner and turned into product. Since no soluble sugars were observed during coculture fermentation, it may be that the coculture partner rapidly consumes excess sugar and prevents overflow metabolism from occurring in *C. thermocellum*.

Evolutionary engineering was used to improve upon the fermentation performance of the Δ*hpt* Δ*ldh* Δ*pta* strain, but there are indications that additional metabolic tuning is needed. An important node that determines if acetyl-coenzyme A (CoA) is directed to acetate or ethanol is the mechanism by which the pool of reduced ferredoxin, formed during the conversion of pyruvate to acetyl-CoA by pyruvate-ferredoxin oxidoreductase (POR), is reoxidized. Since acetate is the primary fermentation product, it appears that the most efficient way to oxidize reduced ferredoxin in *C. thermocellum* is through $H_2$ evolution as opposed to the formation of NADH by ferredoxin:NADH oxidoreductase (FNOR) and subsequent redox balance by ethanol production. Thus, in the absence of sufficient FNOR activity, lack of Pta may cause redox imbalance and the accumulation of central metabolites leading to and including acetyl-CoA. In the Δ*hpt* Δ*pta* strain this bottleneck is partially alleviated by increasing the production of lactate. The Δ*hpt* Δ*pta* Δ*ldh* strain does not have the ability to make lactate in response to redox imbalance, and this may be the reason that pyruvate is secreted into the growth medium and that the overall product yield is lower. Exploration of the POR node and the identification of possible deficiencies, such as an FNOR-like activity capable of transferring electrons from reduced ferredoxin to $NAD^+$, may provide a means to enhance the performance of *C. thermocellum* ethanol fermentation.

The evolved Δ*hpt* Δ*ldh* Δ*pta* strain produced a 40:1 molar ratio of ethanol to organic acid while making 5.61 g/liter eth-

anol from 19.6 g/liter Avicel. This strain, in coculture with an engineered *T. saccharolyticum* ethanologen, fermented 92.2 g/liter Avicel into 38.1 g/liter ethanol with high conversion and a yield comparable to that of actively growing *S. cerevisiae*. This is the highest ethanol titer produced by a thermophilic, cellulolytic culture to date. The ability to improve fermentation of ethanol from cellulose by *C. thermocellum* in coculture with noncellulolytic thermophilic anaerobes has been established with numerous pentose-utilizing organisms including *T. saccharolyticum* (4, 7, 13). Recent advances described here and by Shaw et al. (24) have allowed us to eliminate organic acid production, which has been a major limitation to realizing the potential of these cocultures. With the development of the tools and strains described in this report, there are no insurmountable obstacles preventing the development of *C. thermocellum* as the cornerstone of a robust cellulolytic platform that is the basis for a CBP solution for cellulosic ethanol.

## ACKNOWLEDGMENTS

This research was supported by Mascoma Corporation and a grant from the BioEnergy Science Center, Oak Ridge National Laboratory, a U.S. Department of Energy (DOE) Bioenergy Research Center, supported by the Office of Biological and Environmental Research in the DOE Office of Science.

We thank A. Joe Shaw, Christopher D. Herring, and William R. Kenealy for useful discussion.

Several of the authors are employees of or hold a consulting position with Mascoma Corporation, which has a financial interest in the organisms described herein.

## REFERENCES

1. **Alper, H., and G. Stephanopoulos.** 2009. Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? Nat. Rev. Microbiol. **7:**715–723.
2. **Bayer, E. A., R. Lamed, B. A. White, and H. J. Flint.** 2008. From cellulosomes to cellulosomics. Chem. Rec. **8:**364–377.
3. **Benoit, L., C. Cailliez, E. Petitdemange, and J. Gitton.** 1992. Isolation of cellulolytic mesophilic clostridia from a municipal solid waste digestor. Microb. Ecol. **23:**117–125.
4. **Bryant, F. O., J. Wiegel, and L. G. Ljungdahl.** 1988. Purification and properties of primary and secondary alcohol dehydrogenases from *Thermoanaerobacter ethanolicus*. Appl. Environ. Microbiol. **54:**460–465.
5. **Czako, M., and L. Marton.** 1994. The herpes simplex virus thymidine kinase gene as a conditional negative-selection marker gene in *Arabidopsis thaliana*. Plant Physiol. **104:**1067–1071.
6. **Demain, A. L., M. Newcomb, and J. H. Wu.** 2005. Cellulase, clostridia, and ethanol. Microbiol. Mol. Biol. Rev. **69:**124–154.
7. **Duong, T.-V.C., E. A. Johnson, and A. L. Demain.** 1983. Thermophillic, anaerobic, cellulolytic bacteria. Topics Enzyme Ferment. Biotechnol. **7:**156–195.
8. **du Plessis, L., S. H. Rose, and W. H. van Zyl.** 2010. Exploring improved endoglucanase expression in *Saccharomyces cerevisiae* strains. Appl. Microbiol. Biotechnol. **86:**1503–1511.
9. **Gardiner, D. M., and B. J. Howlett.** 2004. Negative selection using thymidine kinase increases the efficiency of recovery of transformants with targeted genes in the filamentous fungus *Leptosphaeria maculans*. Curr. Genet. **45:**249–255.
10. **Hogsett, D. A. L.** 1995. Cellulose hydrolysis and fermentation by *Clostridium thermocellum* for the production of ethanol. Ph.D. thesis. Dartmouth College, Hanover, NH.
11. **Johnson, E. A., A. Madia, and A. L. Demain.** 1981. Chemically defined minimal medium for growth of the anaerobic cellulolytic thermophile *Clostridium thermocellum*. Appl. Environ. Microbiol. **41:**1060–1062.
12. **Khang, C. H., S. Y. Park, Y. H. Lee, and S. Kang.** 2005. A dual selection based, targeted gene replacement tool for *Magnaporthe grisea* and *Fusarium oxysporum*. Fungal Genet. Biol. **42:**483–492.
13. **Lamed, R., and J. G. Zeikus.** 1980. Ethanol production by thermophilic bacteria: relationship between fermentation product yields of and catabolic enzyme activities in *Clostridium thermocellum* and *Thermoanaerobium brockii*. J. Bacteriol. **144:**569–578.
14. **Lamed, R. J., J. H. Lobos, and T. M. Su.** 1988. Effect of stirring and hydrogen on fermentation products of *Clostridium thermocellum*. Appl. Environ. Microbiol. **54:**1216–1221.
15. **Lilly, M., H. P. Fierobe, W. H. van Zyl, and H. Volschenk.** 2009. Heterologous expression of a *Clostridium* minicellulosome in *Saccharomyces cerevisiae*. FEMS Yeast Res. **9:**1236–1249.
16. **Lynd, L. R., et al.** 2008. How biotech can transform biofuels. Nat. Biotechnol. **26:**169–172.
17. **Lynd, L. R., P. J. Weimer, W. H. van Zyl, and I. S. Pretorius.** 2002. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol. Mol. Biol. Rev. **66:**506–577.
18. **Maki, M., K. T. Leung, and W. Qin.** 2009. The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. Int. J. Biol. Sci. **5:**500–516.
19. **Olson, D. G., et al.** 2010. Deletion of the Cel48S cellulase from *Clostridium thermocellum*. Proc. Natl. Acad. Sci. U. S. A. **107:**17727–17732.
20. **Pritchett, M. A., J. K. Zhang, and W. W. Metcalf.** 2004. Development of a markerless genetic exchange method for *Methanosarcina acetivorans C2A* and its use in construction of new genetic tools for methanogenic archaea. Appl. Environ. Microbiol. **70:**1425–1433.
21. Reference deleted.
22. **Shanks, R. M., N. C. Caiazza, S. M. Hinsa, C. M. Toutain, and G. A. O'Toole.** 2006. *Saccharomyces cerevisiae*-based molecular tool kit for manipulation of genes from gram-negative bacteria. Appl. Environ. Microbiol. **72:**5027–5036.
23. **Shaw, A. J., D. A. Hogsett, and L. R. Lynd.** 2009. Identification of the [FeFe]-hydrogenase responsible for hydrogen generation in *Thermoanaerobacterium saccharolyticum* and demonstration of increased ethanol yield via hydrogenase knockout. J. Bacteriol. **191:**6457–6464.
24. **Shaw, A. J., et al.** 2008. Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. Proc. Natl. Acad. Sci. U. S. A. **105:**13769–13774.
25. **Stout, J. T., and C. T. Caskey.** 1985. HPRT: gene structure, expression, and mutation. Annu. Rev. Genet. **19:**127–148.
26. **Szybalski, W.** 1992. Use of the HPRT gene and the HAT selection technique in DNA-mediated transformation of mammalian cells: first steps toward developing hybridoma techniques and gene therapy. Bioessays **14:**495–500.
27. **Trinh, C. T., and F. Srienc.** 2009. Metabolic engineering of *Escherichia coli* for efficient conversion of glycerol to ethanol. Appl. Environ. Microbiol. **75:**6696–6705.
28. **Tripathi, S. A., et al.** 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.
29. **Tsai, S. L., J. Oh, S. Singh, R. Chen, and W. Chen.** 2009. Functional assembly of minicellulosomes on the *Saccharomyces cerevisiae* cell surface for cellulose hydrolysis and ethanol production. Appl. Environ. Microbiol. **75:**6087–6093.
30. **Tyurin, M. V., S. G. Desai, and L. R. Lynd.** 2004. Electrotransformation of *Clostridium thermocellum*. Appl. Environ. Microbiol. **70:**883–890.
31. Reference deleted.
32. **Yanase, S., et al.** 2010. Ethanol production from cellulosic materials using cellulase-expressing yeast. Biotechnol. J. **5:**449–455.
33. **Yomano, L. P., S. W. York, S. Zhou, K. T. Shanmugam, and L. O. Ingram.** 2008. Re-engineering *Escherichia coli* for ethanol production. Biotechnol. Lett. **30:**2097–2103.
34. **Zhang, Y., and L. R. Lynd.** 2003. Quantification of cell and cellulase mass concentrations during anaerobic cellulose fermentation: development of an enzyme-linked immunosorbent assay-based method with application to *Clostridium thermocellum* batch cultures. Anal. Chem. **75:**219–227.
35. **Zhang, Y. H., and L. R. Lynd.** 2004. Kinetics and relative importance of phosphorolytic and hydrolytic cleavage of cellodextrins and cellobiose in cell extracts of *Clostridium thermocellum*. Appl. Environ. Microbiol. **70:**1563–1569.
36. **Zhang, Y. H., and L. R. Lynd.** 2005. Regulation of cellulase synthesis in batch and continuous cultures of *Clostridium thermocellum*. J. Bacteriol. **187:**99–106.