0273-2289/1559–0291 (Online)/07/136–140/663–674/$30.00

Construction and Evaluation of a *Clostridium thermocellum* ATCC 27405 Whole-Genome Oligonucleotide Microarray

Steven D. Brown,[1] Babu Raman,[2] Catherine K. McKeown,[2] Shubha P. Kale,[3] Zhili He,[1,4] and Jonathan R. Mielenz*,[2]

[1]Microbial Ecology and Physiology Group; [2]Bioconversion Science and Technology Group, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, E-mail: mielenzjr@ornl.gov; [3]Department of Biology, Xavier University of Louisiana, New Orleans, LA; and [4]Current address: Institute of Environmental Genomics, Department of Botany and Microbiology, University of Oklahoma, Norman, OK

Abstract

Clostridium thermocellum is an anaerobic, thermophilic bacterium that can directly convert cellulosic substrates into ethanol. Microarray technology is a powerful tool to gain insights into cellular processes by examining gene expression under various physiological states. Oligonucleotide microarray probes were designed for 96.7% of the 3163 *C. thermocellum* ATCC 27405 candidate protein-encoding genes and then a partial-genome microarray containing 70 *C. thermocellum* specific probes was constructed and evaluated. We detected a signal-to-noise ratio of three with as little as 1.0 ng of genomic DNA and only low signals from negative control probes (nonclostridial DNA), indicating the probes were sensitive and specific. In order to further test the specificity of the array we amplified and hybridized 10 *C. thermocellum* polymerase chain reaction products that represented different genes and found gene specific hybridization in each case. We also constructed a whole-genome microarray and prepared total cellular RNA from the same point in early-logarithmic growth phase from two technical replicates during cellobiose fermentation. The reliability of the microarray data was assessed by cohybridization of labeled complementary DNA from the cellobiose fermentation samples and the pattern of hybridization revealed a linear correlation. These results taken together suggest that our oligonucleotide probe set can be used for sensitive and specific *C. thermocellum* transcriptomic studies in the future.

Index Entries: Biomass; cellulose; ethanol; fermentation; transcriptomics.

*Author to whom all correspondence and reprint requests should be addressed.

Introduction

In 2004, 3.4 billion gal of ethanol was blended into gasoline, which was approx 2% of gasoline sold in the United States by volume or 1.3% (2.5×10^{17} J) of its energy content *(1)*. Transportation ethanol is derived primarily from corn; however, there is need to develop an emerging industry to produce ethanol from lignocellulosic biomass to meet expected future demand *(2)*. Because of the complexities of biomass, extraction and hydrolysis of the cellulose requires thermochemical pretreatment of the biomass followed by the addition of enzymes needed to hydrolyze these polymers to simple sugars that can be fermented to ethanol by an added fermentative microorganism *(3)*. However, a game-changing technology is being developed for a process to simultaneously convert the cellulosic component of biomass to end products with a *single processing step* that consolidates cellulase enzyme production, cellulose hydrolysis, and fermentation *(4)*. Particularly important is that no added cellulase enzymes are needed, thus avoiding the added cellulose production costs, recently been reported in the range of 10–20 ¢/gal of ethanol produced *(5)*. Central to this consolidated bioprocessing approach is a thermophilic (high-temperature) bacterium called *Clostridium thermocellum*, which has the critical ability to produce its own cellulases that permit it to very rapidly hydrolyze cellulose, using a structure called the cellulosome, for growth and energy. In fact, *C. thermocellum* exhibits the highest rate of cellulose hydrolysis known and as a result the protein chemistry of this process has been extensively studied for 20 yr *(6)*.

There has been significant for cellulosome structural biology, and cellulose fermentation *(6)*, however, little is known regarding key enzyme expression levels that might be either bottlenecks or key catalytic steps that will serve as targets for further metabolic engineering of this organism to maximize ethanol yields from biomass. Therefore, this article outlines initial progress aimed at investigating the intrinsic biology of this unique organism, especially toward gene expression during cellobiose and cellulose fermentation.

The application of microarray technology to study gene expression at the level of whole transcriptome has been widely used for many years *(7,8)* and more recently transcriptomic studies have been conducted in ethanologenic bacteria and yeast *(9–14)*. The availability of the 3.8 Mb *C. thermocellum* genome sequence (http://genome.ornl.gov/microbial/cthe) predicted to encode 3163 candidate protein-encoding genes permitted the development of a whole-genome microarray, thus allowing the application of microarray technology to investigate patterns of gene expression during fermentation of cellulose to ethanol in *C. thermocellum*. As a first step, in this study, we present the design, fabrication, and assessment of a whole-genome microarray for *C. thermocellum* ATCC 27405.

Materials and Methods

Bacterial Strains, Culture Conditions, and Chemicals

C. thermocellum strain ATCC 27405 was a kind gift from Prof. Herb Strobel, University of Kentucky, Lexington, KY. A 1 : 25 dilution of a fresh overnight culture (16 h, optical density [OD_{600}] approx 0.9) of *C. thermocellum* was used to inoculate 2 L of MTC medium containing 5.0 g/L cellobiose and 1.0 g/L yeast extract in a Braun BioStat B fermentor (Sartorius BBI Systems Inc., Bethlehem, PA), essentially as described previously *(15)*, except that *C. thermocellum* was cultured at 58°C and pH 7.0 (controlled through addition of 3 *N* NaOH) with an agitation of 250 rpm. Reagent grade chemicals were obtained from Sigma (St. Louis, MO) unless indicated otherwise. Cell free culture supernatants were analyzed by high-performance liquid chromatography (Waters Corp, Milford, MA) to measure cellobiose, acetate, lactate, and ethanol concentrations.

Whole-Genome DNA Microarray Construction and Design of Polymerase Chain Reaction Primers

DNA sequences for the 3163 *C. thermocellum* ATCC 27405 predicted protein-encoding genes were obtained from The Joint Genome Institute (http://genome.ornl.gov/microbial/cthe/17nov03_obsolete/) using sequence assembled in November 2003. Oligonucleotide probes that represented the whole genome of *C. thermocellum* were designed using the CommOligo software *(16,17)* and were commercially synthesized without modification (MWG Biotech, High Point, NC) in 96-well plates. The concentration of the probes was adjusted to 100 pmol/μL, transferred to 384-well printing plates in a final concentration of 50% dimethyl sulfoxide using a BioMek FX liquid handling robot (Beckman-Coulter, Fullerton, CA) and then spotted onto UltraGAPS glass slides (Corning Life Sciences, Corning, NY) using a BioRobotics Microgrid II microarrayer (Genomic Solutions, Ann Arbor, MI) in a dust-free clean room maintained at 21°C and 50% relative humidity. Spotted DNA was stabilized on slides by ultraviolet crosslinking using an Ultraviolet 1800 Stratalinker (Stratagene, La Jolla, CA) according to slide manufacturer's instructions (Corning Life Sciences). A partial genome microarray that contained 16 replicates for 72 *C. thermocellum* and control probes was constructed initially, and subsequently, whole-genome microarrays were fabricated, which contained two replicates per probe on each slide. The Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/ primer3_www.cgi) was used to design primers to amplify gene specific polymerase chain reaction (PCR) products for hybridizations.

Nucleic Acid Isolation and Preparation of Labeled complementary DNA Targets

Early-exponential phase (OD_{600}, 0.25) batch cultures were used for RNA isolation. Total cellular RNA was isolated using a lysozyme

coupled with TRIzol reagent (Invitrogen, Carlsbad, CA) treatment as described previously *(18)*. Precipitated RNA was further treated with RNase-free DNase I (Qiagen, Valencia, CA) to digest any residual chromosomal DNA, and subsequently purified with RNeasy Mini kit according to the manufacturer's instructions (Qiagen). *C. thermocellum* genomic DNA was isolated using the DNeasy Tissue kit (Qiagen). The concentration and purity of the extracted nucleic acids was determined at OD_{260} and OD_{280} with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE). Purified RNA was used as the template to generate complementary DNA (cDNA) copies labeled with either Cy3-dUTP or Cy5-dUTP (Amersham Biosciences, Piscataway, NJ) and in a duplicate set of reactions the fluorescent Cy dyes were reversed for the different technical replicates to analyze dye-specific variations in hybridization signal intensity. Genomic DNA and PCR products were labeled with Cy-3dUTP and Cy-5dUTP, respectively, using a Bioprime Labeling kit (Invitrogen) with random-primers and then purified using a Qiaquick PCR kit (Qiagen) *(19)*. The labeled and purified cDNA were then dried using the SDP1010 SpeedVac System (ThermoSavant, Holbrook, NY). The sequences for the oligonucleotides used to amplify the probe-specific PCR products are shown in Table 1.

Microarray Hybridization, Scanning, Image Quantification, and Data Analysis

Preliminary microarray quality assessments were made by staining the microarrays with a 1 : 1000 dilution of Syto61 dye (Invitrogen) pure containing 10.0 µg/mL of bovine serum albumin (New England Biolabs, Ipswich, MA) for 20 min at room temperature, followed by two 0.1X SSC (Ambion, Austin, TX) washes. Hybridization and washing conditions for oligonucleotide microarrays have been described elsewhere *(20)*. Microarray images were scanned using a ScanArray Express (PerkinElmer) scanner, and spot signal, quality, and background fluorescent intensities were quantified using ImaGene version 6.0 (Biodiscovery, Marina Del Rey, CA). Signal-to-noise ratio were calculated as described previously *(21)* in Microsoft Excel, and GeneSpring 7.0 (Agilent Technologies, Palo Alto, CA) was used to transform microarray data with the locally weighted scatterplot smoothing method of normalization and to remove poor/empty spots.

Results and Discussion

C. thermocellum *Cellobiose Fermentations*

Initially, we investigated the growth of *C. thermocellum* during fermentation of cellobiose. Cells were grown in amended MTC medium and growth was monitored for approx 12 h. Two independent experiments

Table 1
Oligonucleotides Used to Amplify Gene-Specific PCR Products

Number[a]	Gene	Forward primer (5′–3′)	Reverse primer (5′–3′)	Gene product
1	*Ct3187*	GCATAGGAACATCCCTGTGTG	CATTGTCCACAGGAAGCAAG	Phosphoglycerate kinase
2	*Ct0462*	GCTGTCAATTCCACTGCAAA	CGCAATCGGCATATACAAAG	Phosphotransacetylase
3	*Ct0463*	TCACAAGCTTGCCATACAGG	CGGAGCCAGTTCAACACAAT	Acetate kinase
4	*Ct3150*	GGACTTTTTCTGTGGCAAGG	ACATTCCCCGTTTGTACAGG	Iron-containing alcohol dehydrogenase
5	*Ct3185*	TGTACAACAAACTGCCTTGCTC	TGAGTTGCAGTTGTAGCGTGT	Glyceraldehyde-3-phosphate dehydrogenase
6	*Ct3012*	CGGGAGGAGAAGGTACCAG	TTCAGAAGTTCAATAATATGCTCCA	Nucleotidyl transferase
7	*Ct1135*	GTGGAGCTGCACACATATCG	GCACCCGCAGTTTTAACATC	L-lactate dehydrogenase
8	*Ct2924*	ATTATTGGCGAACACGGTGA	AATCTGCTCCTCGCACTGAT	L-lactate dehydrogenase
9	*Ct3589*	TCGTTCTGCCTGAGAAACCT	CCCCTTCGGCAACTATAACA	6-Phosphofructokinase
10	*Ct3735*	ACAGGGAAGAATTTkCGAGA	TGGAATGAGTGGGAAAGCAT	Glucose-6-phosphate isomerase

[a]Corresponds to labeled PCR targets in Fig. 3 and primer pairs that will be used for Q-PCR for genes involved in cellulose fermentation identified in Fig. 6.

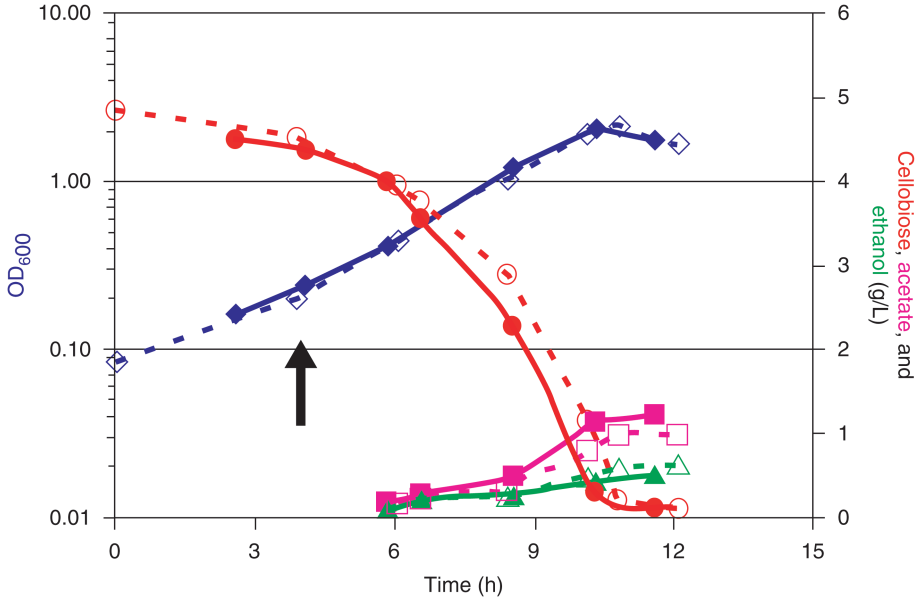

Fig. 1. *C. thermocellum* cellobiose fermentations. OD at 600 nm (OD_{600}) (blue lines), cellobiose (red lines), acetate (pink lines), and ethanol (green lines) concentration (g/L) is plotted against fermentation time (h). Open and closed symbols correspond to data from two independent fermentations. Arrow indicates sample harvest time for total RNA isolation.

based on culture turbidity measurements and high-performance liquid chromatography analysis of substrate (cellobiose) consumption and byproduct (acetate and ethanol) formation, indicated that batch cultures of *C. thermocellum* had typical bacterial growth cycle kinetics (Fig. 1). Total cellular RNA was extracted from two technical replicates at the same point in early-logarithmic growth phase during single cellobiose fermentation (Fig. 1).

C. thermocellum *Microarray Probe Design*

Unique 70-mer oligonucleotide probes were designed for 94.2% of the 3163 *C. thermocellum* candidate protein-encoding genes by the CommOligo software *(16)*. A further 10 probes were designed for groups of highly similar genes, so that 96.7% of the putative coding sequences were represented and only 104 coding sequences remained unrepresented on the whole-genome microarray. The *C. thermocellum* genome sequence contains a large number of putative genes whose products are predicted to encode proteins with transposon-related functions *(22)*. Wherein sequences were highly similar, such as for genes encoding proteins with transposon-related functions, individual gene probes could not be designed. Group probes 1–10 represented putative hypothetical, transposase, transposase (IS116/IS110/IS902), phage/plasmid primase P4, hypothetical, transposase (mutator type),

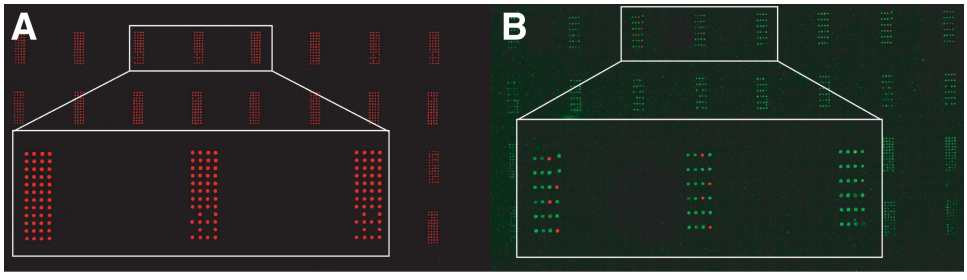

Fig. 2. Specificity of *C. thermocellum* microarray probes. *C. thermocellum* partial-genome microarray stained with nonspecific Syto61 dye for DNA **(A)** and hybridized with gDNA labeled with Cy3 dye (green), which hybridizes to all *C. thermocellum* probes and *Ct3187* PCR product labeled with Cy5 dye (red) that hybridizes to replicates of *Ct3187* probe **(B)**.

transposase (IS3/IS911), transposase (mutator type, transposase (IS30 family), and transposase proteins, respectively. The percentage of *C. thermocellum* probes designed for the whole genome is similar to several recently constructed whole-genome microarrays for *Desulfovibrio vulgaris* (Hildenborough) and *Shewanella oneidensis* MR-1, which had 98.6 and 94.3% probe coverage of the genomes, respectively *(20,23)*. The group probe design feature of CommOligo enabled probes to be designed for highly homologous sequences, thus extending the fraction of gene expression comparisons that will be able to be made in the future.

Microarray Probe Specificity and Sensitivity

Initially, a subset of the whole-genome microarray probes was tested for sensitivity and specificity by producing partial-genome microarrays that contained 70 *C. thermocellum* probes representing the range of functional diversity in the predicted gene products. The DNA stain Syto61 was used to assess the quality of the partial-genome microarrays and to confirm signal was observed from negative control probes (Fig. 2A). PCR products designed to hybridize to individual probes that represented specific genes were labeled with Cy-5 and hybridized to the partial array to confirm signal specificity of the designed oligonucleotide probes (Fig. 2B). In each case the 10-labeled PCR products gave signal-to-noise ratios of more than three, or positive for probe-target hybridization interaction *(24)*, whereas only low values and SNRs below three were observed from negative control probes (Fig. 3).

The hybridization of different amounts of *C. thermocellum* genomic DNA labeled with Cy-dye to *C. thermocellum* partial microarrays showed that as little as 1.0 ng of genomic DNA could be labeled and give signal-to-noise ratios more than three (Fig. 4). A signal-to-noise ratio more than three is a general criterion considered as the minimum probe signal necessary that can be quantified accurately *(24)*. Our results were in keeping with other studies *(25,26)* and were indicative of the *C. thermocellum* microarray probes

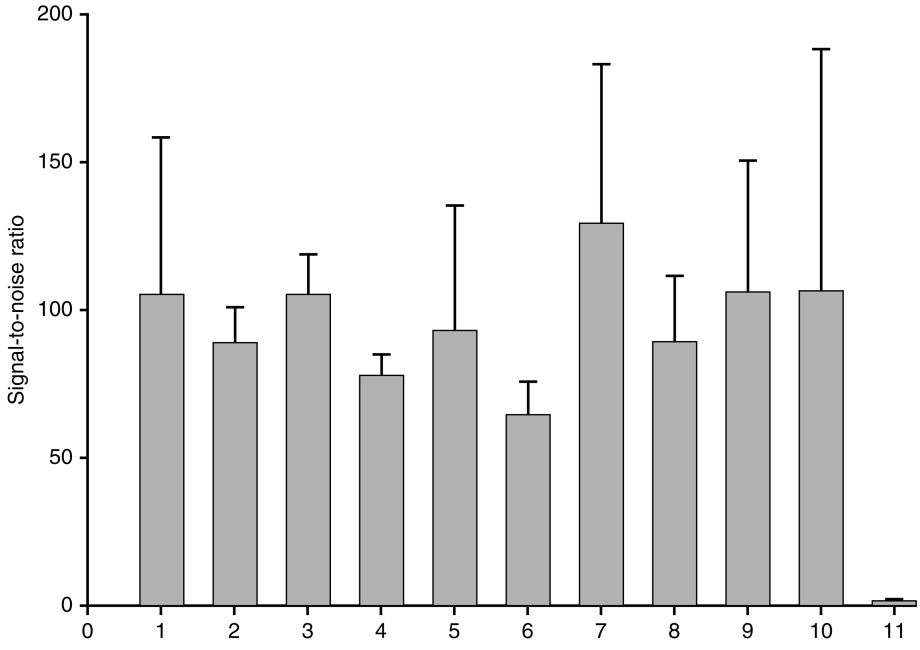

Fig. 3. Hybridization of PCR products for specific *C. thermocellum* genes to partial-genome microarray. Signal-to-noise ratios for gene-specific probes: 1, *Ct3187*; 2, *Ct0462*; 3, *Ct0463*; 4, *Ct3150*; 5, *Ct3185*; 6, *Ct3012*; 7, *Ct1135*; 8, *Ct2924*; 9, *Ct3589*; 10, *Ct3735*; and 11, remaining *C. thermocellum* probes. Primer sequences used to amplify PCR products and probe descriptions are given in Table 1.

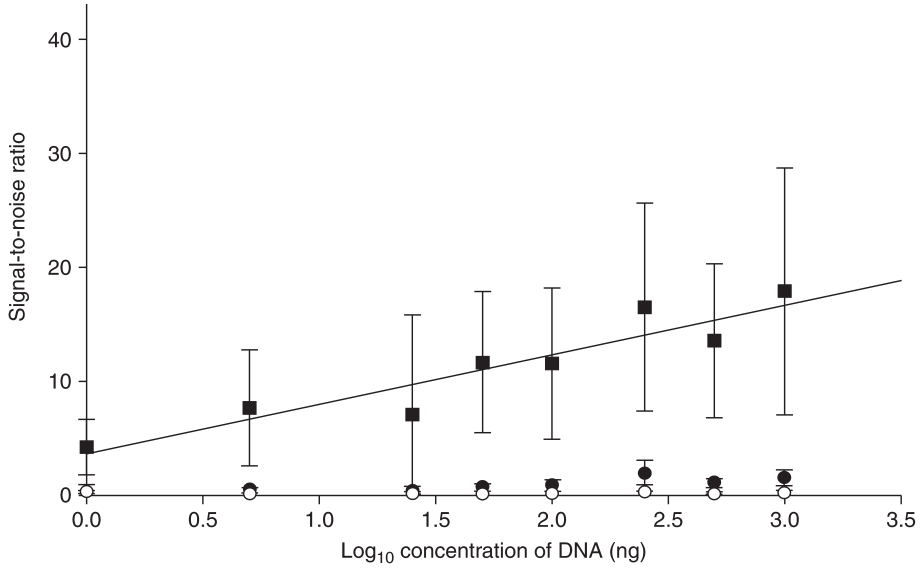

Fig. 4. Hybridization of *C. thermocellum* genomic DNA to *C. thermocellum* partial microarray. Average of all *C. thermocellum* probes, ■ and ▬ regression plot; negative control 1, ●; negative control 2, ○.

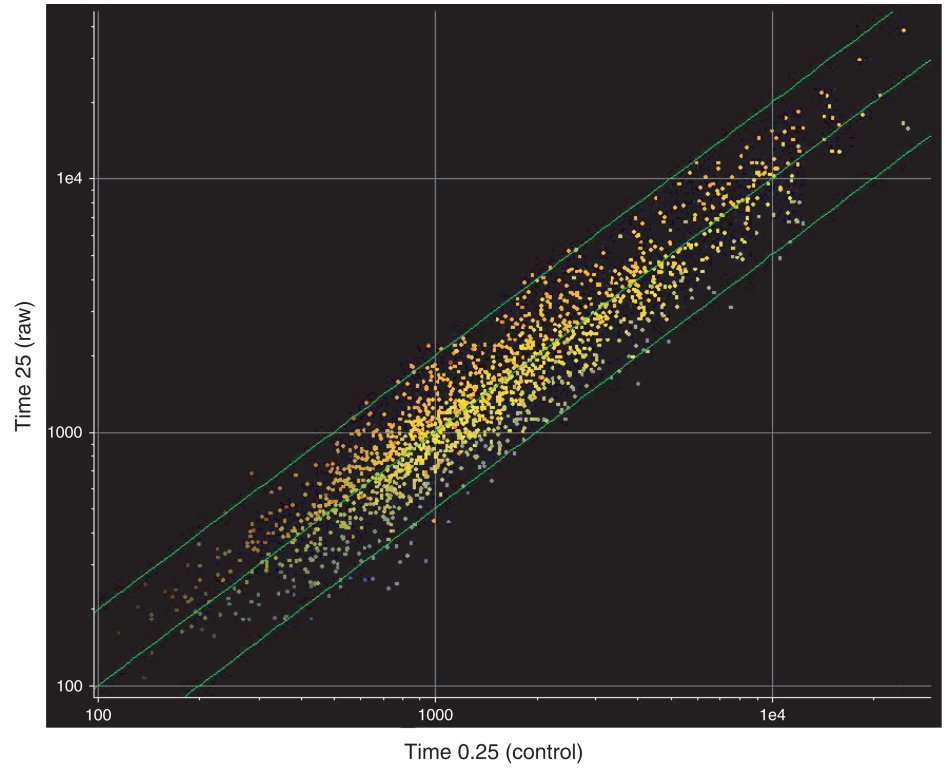

Fig. 5. Scatterplot of whole-genome microarray data. The upper and lower green lines represent twofold levels of differential expression and the middle green line represents no change in expression.

being sensitive as well as specific. In order to further assess the reliability of the microarray data we constructed a whole-genome microarray and cohybridized cDNA prepared from total cellular RNA extracted from two technical replicates from the same *C. thermocellum* cellobiose fermentation in early-logarithmic growth phase. The pattern of hybridization revealed a linear relationship between the samples and 97.4% of the genes fell within a twofold threshold after data normalization and poor/empty spot removal by GeneSpring (Fig. 5). This value may be further improved on in future studies, as one potential source of variation was average incorporation of the Cy-dye, which was less than 1 pmol/μL of purified target DNA. In this experiment only three of the 1708 genes were slightly outside the bounds of a threefold change in expression value between samples. Overall, the results described above suggest our oligonucleotide probe set can be used for sensitive and specific *C. thermocellum* transcriptomic studies in the future.

Conclusion

To our knowledge this is the first whole-genome microarray for a thermophile capable of converting cellulose to ethanol in a consolidated

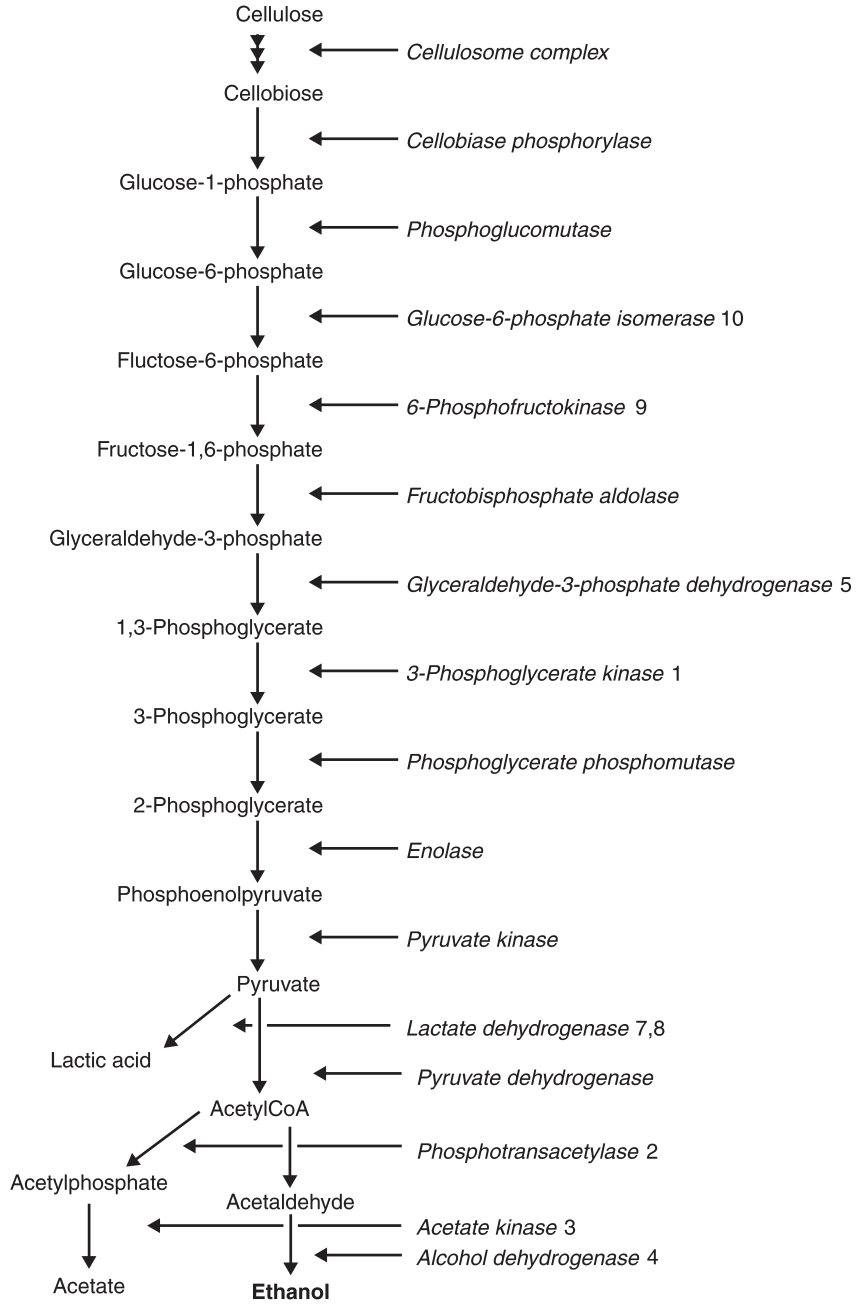

Fig. 6. Metabolic pathway for conversion of cellulose to ethanol, lactic acid, and acetic acid. The numbered genes correspond to PCR products in Table 1 and their location in this metabolic pathway.

process. The PCR products shown in Table 1 were chosen not only to verify the quality of the *C. thermocellum* microarray, but also for future quantitative-PCR primer analysis of key genes in the pathway from cellulose to the final metabolic end products of ethanol, lactic acid, and acetic acid (Fig. 6).

Future transcriptomics using the whole-genome microarray will provide insight into the expression genes whose products are thought to be important for cellulose fermentation and also potentially provide insight into the roles of other genes whose importance cannot be inferred by *a priori* assumptions. Interestingly, such experiments will answer questions whether both genomic copies of lactate dehydrogenase are expressed (Table 1, genes 7 and 8) or if transcription occurs from only one locus during fermentation of either cellobiose and/or cellulose. Information on these and other genes contributing to or detracting from the pathway for ethanol production will be invaluable as the complexities of *C. thermocellum* metabolic circuits are unraveled. In conclusion, the application of transcriptomic profiling, along with other molecular biology tools and physiological studies offer the opportunity to better understand fundamental biology of ethanologenic bacteria and the potential for improved production of ethanol.

Acknowledgments

Research sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory (ORNL), managed by UT-Battelle, LLC for the US Department of Energy under Contract No. DE-AC05-00OR22725. S.P.K was supported by the DOE Faculty Sabbatical Program sponsored by the DOE Office of Science. We thank Liyou Wu for assistance with microarray printing.

Note Added in Proof

As this paper was going to press, the *C. thermocellum* ATCC 27405 genome was closed and finished. The finished genome sequence can be found at http://genome.jgi.psf.org/finished_microbes/cloth/cloth.home.html.

References

1. Davis, S. C. and Diegel, S. W. (2004), In: *Transportation Energy Data Book: Technical Report No. ORNL-6973*. Oak Ridge National Laboratory, Oak Ridge, TN.
2. http://www.ethanolrfa.org/industry/locations/. Accessed March 19, 2007.
3. Mielenz, J. R. (2001), *Curr. Opin. Microbiol.* **4,** 324–329.
4. Lynd, L. R., van Zyl, W. H., McBride, J. E., and Laser, M. (2005), *Curr. Opin. Biotechnol.* **16,** 577–583.
5. Greer, D. (2005), *Biocycle* **46,** 61–65.
6. Lynd, L. R., Weimer, P. J., van Zyl, W. H., and Pretorius, I. S. (2002), *Microbiol. Mol. Biol. Rev.* **66,** 506–577.
7. DeRisi, J., Penland, L., Brown, P. O., et al. (1996), *Nat. Genet.* **14,** 457–460.
8. Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O., and Davis, R. W. (1996), *Proc. Natl. Acad. Sci. USA* **93,** 10,614–10,619.
9. Sonderegger, M., Jeppsson, M., Hahn-Hagerdal, B., and Sauer, U. (2004), *Appl. Environ. Microbiol.* **70,** 2307–2317.
10. Tomas, C. A., Welker, N. E., and Papoutsakis, E. T. (2003), *Appl. Environ. Microbiol.* **69,** 4951–4965.
11. Sedlak, M., Edenberg, H. J., and Ho, N. W. Y. (2003), *Enzyme Microbial Technol.* **33,** 19–28.

12. Jewett, M. C., Oliveira, A. P., Patil, K. R., and Nielsen, J. (2005), *Biotech. Bioprocess Eng.* **10,** 385–399.
13. Wahlbom, C. F., Otero, R. R. C., van Zyl, W. H., Hahn-Hägerdal, B., and Jonsson, L. J. (2003), *Appl. Environ. Microbiol.* **69,** 740–746.
14. Salusjärvi, L., Pitkänen, J. -P., Aristidou, A., Ruohonen, L., and Penttilä, M. (2006), *Appl. Biochem. Biotechnol.* **128,** 237–274.
15. Zhang, Y. -H. P. and Lynd, L. R. (2005), *J. Bacteriol.* **187,** 99–106.
16. Li, X., He, Z., and Zhou, J. (2005), *Nucleic Acids Res.* **33,** 6114–6123.
17. Bozdech, Z., Zhu, J., Joachimiak, M. P., Cohen, F. E., Pulliam, B., and DeRisi, J. L. (2003), *Genome Biol.* **4,** R9.
18. Tomas, C. A., Alsaker, K. V., Bonarius, H. P. J., et al. (2003), *J. Bacteriol.* **185,** 4539–4547.
19. Pollack, J. R., Perou, C. M., Alizadeh, A. A., et al. (1999), *Nat. Genet.* **23,** 41–46.
20. Chhabra, S. R., He, Q., Huang, K. H., et al. (2006), *J. Bacteriol.* **188,** 1817–1828.
21. Tiquia, S. M., Chong, S. C., Fields, M. W., and Zhou, J. (2004), In: *Molecular Microbial Ecology Manual*, Kowalchuk, G. A., de Bruijn, F. J., Head, I. M., Lakkennans, A. D., and van Elsas, J. D. (eds.), Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 1743–1763.
22. http://genome.ornl.gov/microbial/cthe/.
23. Gao, H. C., Wang, Y., Liu, X. D., et al. (2004), *J. Bacteriol.* **186,** 7796–7803.
24. Verdnik, D., Handran, S., and Pickett, S. (2002), In: *DNA image analysis: Nuts & bolts*, Kamberova, G. (ed.), DNA Press, Salem, MA, pp. 89–98.
25. Tiquia, S. M., Wu, L., C. Chong, S. C., et al. (2004), *BioTechniques* **36,** 664–675.
26. He, Z. L., Wu, L. Y., Li, X. Y., Fields, M. W., and Zhou, J. Z. (2005), *Appl. Environ. Microbiol.* **71,** 3753–3760.