Applied and Environmental Microbiology, Nov. 2011, p. 7998–8008 Vol. 77, No. 22
0099-2240/11/$12.00 doi:10.1128/AEM.05677-11

Correlation of Genomic and Physiological Traits of *Thermoanaerobacter* Species with Biofuel Yields▽†

Christopher L. Hemme,[1] Matthew W. Fields,[2] Qiang He,[3,4] Ye Deng,[1] Lu Lin,[1,5] Qichao Tu,[1] Housna Mouttaki,[1,6] Aifen Zhou,[1] Xueyang Feng,[7] Zheng Zuo,[7] Bradley D. Ramsay,[2] Zhili He,[1] Liyou Wu,[1] Joy Van Nostrand,[1] Jian Xu,[5] Yinjie J. Tang,[6] Juergen Wiegel,[8] Tommy J. Phelps,[9] and Jizhong Zhou[1,10]*

Institute for Environmental Genomics, University of Oklahoma, Norman, Oklahoma[1]; Department of Microbiology, Montana State University, Bozeman, Montana[2]; Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, Tennessee[3]; Center for Environmental Biotechnology, University of Tennessee, Knoxville, Tennessee[4]; Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, China[5]; Institute of Groundwater Ecology, Helmholtz Zentrum München, German Research Center for Environment and Health, Munich, Germany[6]; Department of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, Missouri[7]; Department of Microbiology, University of Georgia, Athens, Georgia[8]; Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee[9]; and Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California[10]

Received 18 May 2011/Accepted 15 September 2011

Thermophilic anaerobic noncellulolytic *Thermoanaerobacter* species are of great biotechnological importance in cellulosic ethanol production due to their ability to produce high ethanol yields by simultaneous fermentation of hexose and pentose. Understanding the genome structure of these species is critical to improving and implementing these bacteria for possible biotechnological use in consolidated bioprocessing schemes (CBP) for cellulosic ethanol production. Here we describe a comparative genome analysis of two ethanologenic bacteria, *Thermoanaerobacter* sp. X514 and *Thermoanaerobacter pseudethanolicus* 39E. Compared to 39E, X514 has several unique key characteristics important to cellulosic biotechnology, including additional alcohol dehydrogenases and xylose transporters, modifications to pentose metabolism, and a complete vitamin B_{12} biosynthesis pathway. Experimental results from growth, metabolic flux, and microarray gene expression analyses support genome sequencing-based predictions which help to explain the distinct differences in ethanol production between these strains. The availability of whole-genome sequence and comparative genomic analyses will aid in engineering and optimizing *Thermoanaerobacter* strains for viable CBP strategies.

Global energy demands will increase significantly in coming decades (16), and renewable energy sources such as biofuels have been proposed to help reduce dependence upon fossil energy (8, 23, 27, 49). Current efforts focus on biofuel production from renewable lignocellulosic feedstock (e.g., switchgrass), which constitutes ~50% of the world's biomass (2, 8, 23, 27, 49). Although intensive research and development have been performed on the effective utilization of lignocellulose, problems associated with practical use of this material have not been resolved fully (13). When enzymatic hydrolysis is adopted for cellulosic ethanol production, different levels of process integration can be visualized: (i) separate (or sequential) hydrolysis and fermentation (SHF), where the enzymes (cellulases) are used separately from fermentation tanks; (ii) simultaneous saccharification and fermentation (SSF), which consolidates enzymatic hydrolysis with fermentation of hexose or pentose; (iii) simultaneous saccharification and cofermentation (SSCF), which further combines the fermentation of hexose and pentose together; and (iv) consolidated bioprocessing (CBP), where all required enzymes and ethanol are produced in a single reactor (1, 13, 24, 29, 46, 51). While these production schemes represent increasing levels of simplification through process consolidation, consolidation of multiple steps often results in a loss of process efficiency. Thus, improving the efficiency of individual steps, such as cellulose hydrolysis and ethanol fermentation, remains an important task for the development of economically feasible cellulosic bioethanol.

Recent efforts have focused on metabolic engineering and (more recently) synthetic biology to produce strains or consortia capable of producing biofuels. Efforts to engineer ethanologenic strains optimized for specific process conditions require critical knowledge of the genetic and metabolic determinants of the divergent characteristics of ethanol production, which are best illustrated by cultures of thermophilic ethanologenic *Thermoanaerobacter* species and *Clostridium thermocellum*. *C. thermocellum* is a thermophilic, ethanol-producing bacterium capable of cellulose and hemicellulose decomposition into pentose and hexose, but it ferments hexose to ethanol with

* Corresponding author. Mailing address: Institute for Environmental Genomics, Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019. Phone: (405) 325-6073. Fax: (405) 325-3442. E-mail: jzhou@ou.edu.

† Supplemental material for this article may be found at http://aem.asm.org/.

▽ Published ahead of print on 23 September 2011.

only very low yields (12, 52). In contrast, *Thermoanaerobacter* species efficiently convert both pentose and hexose to ethanol at high yields but cannot degrade cellulose. Previous studies in our laboratory have shown that stable cellulolytic cocultures of *Thermoanaerobacter* and *C. thermocellum* can be established for more efficient simultaneous cellulose degradation and ethanol production (J. Zhou, unpublished results). Such cocultures are a possible solution for consortium-based CBP schemes (8, 18, 28). Interestingly, ethanol yields in these cocultures vary greatly depending on which *Thermoanaerobacter* species is present. Physiological analyses have shown that *Thermoanaerobacter* sp. X514 (33) produces more ethanol in cocultures with selected *C. thermocellum* strains than the corresponding *Thermoanaerobacter pseudethanolicus* 39E (30) cocultures, despite a very close phylogenetic relationship between the two strains (9, 18). To provide insight into the molecular mechanisms underlying differences in sugar utilization and ethanol production, physiological, comparative, and functional genomic analyses were conducted with X514 and 39E. The results indicate key differences in the mechanisms involved in vitamin B_{12} biosynthesis, xylose transport and metabolism, and ethanol production, which may explain the phenotypic differences in ethanol yields observed between these two strains. These results will aid efforts to engineer *Thermoanaerobacter* strains for efficient cellulosic biofuel production in coculture-based CBP schemes.

MATERIALS AND METHODS

Whole-genome sequencing and analysis. (i) Organism source and culturing. Strain 39E (ATCC 33223) was obtained from the American Type Culture Collection (ATCC). A laboratory culture of X514 was used for sequencing, and the culture has been deposited in the ATCC as strain ATCC BAA-938. *Thermoanaerobacter* strains were cultured as described previously, using DCB-1 medium (33).

(ii) Sequencing and annotation. 39E and X514 were sequenced and annotated as previously described (15). All general aspects of library construction and sequencing performed at the Joint Genome Institute (JGI) can be found at http://www.jgi.doe.gov/. Additional manual curation was conducted at the University of Oklahoma, based on literature searches and comparative sequence analysis.

(iii) Sequence analysis. The majority of sequence analysis, including metabolic pathway analysis and ortholog detection, was conducted using online tools at the Joint Genome Institute Integrated Microbial Genomes database (26; http://img.jgi.doe.gov/cgi-bin/w/main.cgi) or was done manually using *ad hoc* Perl scripts. Putative lateral gene transfer events were detected by SIGI-HMM analysis (47).

Growth analysis. (i) Cellulose measurement. Cellulose degradation was measured as previously described (45). Briefly, 1-ml samples were harvested at the stationary phase of growth by centrifugation at 5,000 × *g* for 15 min. The pellet was suspended in distilled water and heated at 100°C for 30 min to achieve cell lysis. The residual cellulose was washed with distilled water two times by centrifugation at 5,000 × *g* for 15 min and then hydrolyzed into soluble sugars with 65% H_2SO_4. The hydrolysis samples were diluted 200 times, and soluble sugars were assayed by using the phenol-sulfuric acid method, with glucose as a standard.

(ii) Phenol-sulfuric acid method for measurement of total sugars. Glucose (25, 50, 75, 100, and 150 mg/liter) was used as the standard. Samples (200 μl) were diluted to the appropriate concentrations and mixed with 200 μl 5% phenol and 1 ml of 98% H_2SO_4. The mixtures were incubated at room temperature for 30 min, and the absorbances of the standards and samples were measured at 490 nm.

(iii) Ethanol production measured by HPLC. One-milliliter samples were harvested at the stationary phase of growth by centrifugation at 5,000 × *g* for 15 min. The supernatant was used to analyze the concentration of ethanol by use of a high-performance liquid chromatography (HPLC) apparatus (Agilent Technologies, CA) operating at 55°C. The mobile phase consisted of 0.025% sulfuric acid at a flow rate of 0.6 ml/min. Ethanol yields with different concentrations of exogenous vitamin B_{12} were compared between the two strains, using the one-tailed paired *t* test (34).

Metabolic flux analysis. (i) Biomass preparation. *Thermoanaerobacter* sp. X514 and 39E were grown anaerobically at 60°C without shaking in DCB-1 medium (33). The strains were initially grown in a 50-ml standard culture medium with an unlabeled carbon source (xylose). At the mid-log phase of growth, a 3% inoculum was added to a 50-ml culture containing 2 g/liter of [1-^{13}C]xylose. At the mid-log phase of growth of this culture, a 3% inoculum from the first ^{13}C-labeled culture medium was used to inoculate a 50-ml subculture (with the same labeled carbon source) in order to reduce the effect of unlabeled carbon from the initial stock.

Biomass was harvested at the late log phase of growth by centrifugation at 8,000 × *g* for 15 min at 10°C. The concentrations of glucose, xylose, acetate, ethanol, and lactate were analyzed with a high-performance liquid chromatography apparatus (Agilent Technologies, CA) equipped with a variable-wavelength (190 to 600 nm) detector (VWD) (with UV absorption measured at 245 nm) and an ion-exclusion column (Aminex HPX-87H; 300 mm × 7.8 mm; Bio-Rad Laboratories, CA) operating at 55°C. The mobile phase consisted of 0.025% sulfuric acid at a flow rate of 0.6 ml/min.

(ii) Flux estimation. Isotopic analysis of proteogenic amino acids was performed (42, 43). Biomass was hydrolyzed in 6 M HCl at 100°C. The amino acid solution was dried and then derivatized in tetrahydrofuran and *N*-(*tert*-butyl dimethylsilyl)-*N*-methyl-trifluoroacetamide at 70°C for 1 h. A gas chromatograph (GC) (model 7890A; Hewlett-Packard, Agilent Technologies, CA) equipped with a DB5-MS column (J&W Scientific, Folsom, CA) and a mass spectrometer (MS) (model 5975C; Agilent Technologies, CA) was used to analyze amino acid labeling profiles. The $[M - 57]^+$ MS group (containing unfragmented amino acids) was used for flux calculations (48). Central metabolic pathway maps of X514 and 39E included the glycolysis pathway, the pentose phosphate pathway, the Entner-Doudoroff (ED) pathway, the citric acid cycle, and the anaplerotic pathway. The fluxes to extracellular metabolites (ethanol, acetate, and lactate) were measured using HPLC. The standard deviations of fluxes were based on errors in concentration measurements of extracellular metabolites. The fluxes toward biomass synthesis were estimated based on dry biomass weight and biomass compositions (assumed to be the same as those of *Escherichia coli*). A ^{13}C flux analysis model (42) was developed using MATLAB 7.0 (The Mathworks, Natick, MA). Metabolic fluxes were constrained by mass balance as follows: $Sv = 0$, where S was the stoichiometry matrix. Fluxes (v) were calculated using isotopomer information from amino acids. The following grid search algorithm was applied to find a set of flux distributions that minimized the difference (ε) between the experimentally observed and predicted isotopomer labeling patterns in five key amino acids (see Table S4 in the supplemental material):

$$\varepsilon(v_n) = \sqrt{\sum_{i=1}^{k}\left(\frac{M_i - N_i(v_n)}{\delta_i}\right)^2},$$

where v_n were the unknown fluxes to be optimized in the program, M_i were the measured MS data, N_i were the corresponding model-simulated MS data, and δ_i were the corresponding standard deviations in the GC-MS data.

Global transcriptomic analysis. (i) Oligonucleotide probe design and microarray construction. Microarrays for 39E and X514 were constructed with 70-mer nucleotide probes. The microarray for 39E contains 2,334 nucleotide probes, covering 2,261 of 2,362 annotated gene sequences. The microarray for X514 contains 2,315 probes, covering 2,322 of 2,475 annotated gene sequences. Gene-specific, inclusive and exclusive group-specific oligonucleotide probes were designed by a new version of the computer program CommOligo (22), based on the following criteria: (i) for gene-specific probes, ≤85% sequence identity, a ≤20-base continuous stretch, and ≥40 kcal/mol of free energy; and (ii) group-specific probes had to meet the above requirements for nontarget groups and also had to have ≥96% sequence identity, a ≥55-base continuous stretch, and ≤90 kcal/mol of free energy within the group.

All designed oligonucleotides were synthesized commercially without modification by MWG Biotech Inc. (High Point, NC). The concentration of oligonucleotides was adjusted to 100 pmol/μl. Oligonucleotide probes were prepared in 50% (vol/vol) dimethyl sulfoxide (Sigma) and spotted onto UltraGAPS glass slides by use of a BioRobotics Microgrid II microarrayer (Genomic Solutions, Ann Arbor, MI). Each oligonucleotide probe had two replicates on a single slide. In addition, six different concentrations (11, 22, 45, 90, 180, and 360 ng/μl) of genomic DNA were spotted (eight replicates for each of the six concentrations on a single slide) as additional positive controls. After the oligonucleotide probes were printed, they were fixed onto the slides by UV cross-linking (600 mJ of energy).

(ii) Culture growth for microarray analysis. X514 and 39E were grown at 60°C in defined anaerobic DCB-1 medium as described above. Cysteine-HCl (0.005 g/liter) was added as the reductant after the medium had been boiled and cooled to room temperature and the pH value adjusted to 7.0. Xylose at different concentrations (1 mM, 10 mM, 50 mM, and 200 mM) was subsequently added to the autoclaved medium from filter-sterilized anaerobic stock solutions. When X514 and 39E cultures were grown to the mid-log phase, cell samples of each culture were harvested by centrifugation (7,000 × *g*) for 10 min at 4°C. Cell pellets were then immediately frozen in liquid N_2 and stored at −80°C. All experiments were conducted in triplicate.

(iii) Total RNA extraction, purification, and labeling. Total cellular RNAs from 39E and X514 were isolated using TRIzol reagent (Invitrogen) according to the manufacturer's protocol. RNA samples were purified by use of an RNeasy minikit (Qiagen), and RNA samples were treated with on-column RNase-free DNase I (Qiagen) to digest genomic DNA. RNA samples were quantified by Nanodrop ND-1000 spectrophotometry at wavelengths of 260 and 280 nm. To generate labeled cDNA targets with reverse transcriptase, 10 μg purified total RNA was used in each reaction mix, using a previously described protocol (44). Briefly, 10 μg random primers (3.0 μg/μl; Invitrogen) was used for priming, and the fluorophore Cy5-dUTP was used for labeling. After labeling, cDNA was purified with a Qiagen PCR purification kit and concentrated with an SPD1010 Speedvac apparatus. The efficiency of labeling was measured by Nanodrop ND-1000 spectrophotometry.

(iv) Genomic DNA labeling. To facilitate data comparison, genomic DNAs from X514 and 39E were labeled separately with Cy3-dUTP, which was used as the control. Each 50-μl labeling reaction mixture contained 500 ng genomic DNA, 20 μl 2.5× random primers (Invitrogen), a 5 mM concentration (each) of dATP, dGTP, and dCTP, 2.5 mM dTTP, 2 μl Klenow polymerase, and 0.5 mM Cy3-dUTP. The labeling reaction proceeded for 3 h at 37°C, and the labeled DNA probe was purified by use of a Qiagen PCR purification kit and concentrated in an SPD1010 Speedvac apparatus (5).

(v) Microarray hybridization and scanning. To hybridize microarray glass slides, the Cy5-dUTP-labeled cDNA targets were mixed with the Cy3-dUTP-labeled genomic DNA and dissolved in a hybridization solution that contained 50% formamide, 5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 10% SDS, and 0.1 mg/ml herring sperm DNA. The labeled sample was incubated for 98°C for 3 min to denature the probe and then kept at 65°C and applied to microarray slides. Hybridization, washing, and drying were all carried out by a Tecan HS4800 Pro hybridization station (U.S. Tecan, Durham, NC). The slides were scanned for the fluorescence intensities of both the Cy5 and Cy3 fluorophores by a ScanArray Express microarray analysis system (Perkin Elmer, Boston, MA).

(vi) Image quantification and data analysis. To determine the signal intensity of fluorescence for each spot, 16-bit scanned TIFF images were analyzed by ImaGene 6.1 software (Biodiscovery Inc., El Segundo, CA) to quantify spot signal, spot quality, and background fluorescence intensities. Data preprocessing and analysis included the following major steps. (i) Empty spots (with a signal-to-noise ratio [SNR] of <2.0), poor spots, and negative spots were flagged according to the instructions of the software and were removed in the subsequent analysis. (ii) The net signal of each spot was calculated by subtracting the background signal and adding a pseudosignal of 100 to obtain a positive value. If the resulting net signal was less than 50, a value of 50 was used. (iii) The total signal intensity for all microarrays (slides) in this experiment was normalized with the genomic DNA signal (Cy3 signal), and a normalization factor was calculated for each slide. (iv) Both Cy3 and Cy5 signal intensities of each spot were adjusted by multiplying by the normalization factor calculated above, and the resultant Cy5 signal of each probe presents the normalized signal intensity for each gene. (v) A ratio (R) of Cy5/Cy3 signals was calculated for each spot on each array. (vi) For comparative genomic analysis, the control and treatment conditions were defined so that the ratios R_c and R_t were defined for each gene under control and treatment conditions, respectively. The $\log_2(R_t/R_c)$ value was then calculated for each gene. (vii) Finally, to assess the significance of gene expression with the treatment, a Z score was calculated by using the following equation: $Z = \log_2(R_t/R_c)/\sqrt{0.25 + \sum \text{variance}}$, where 0.25 is a pseudovariance term. Typically, a Z score cutoff of >2.0 was used for significant changes.

Accession numbers. The sequences and annotations of the complete *T. pseudethanolicus* 39E and *Thermoanaerobacter* sp. X514 genomes are available under GenBank accession numbers CP000924 and CP000923, respectively. (*Caldanaerobacter subterraneus* subsp. *tengcongensis* MB4 was previously deposited under the name *Thermoanaerobacter tengcongensis* MB4, under accession number AE008691, but is no longer considered the prototypical *Thermoanaerobacter* genome.) All sequences are also available at the JGI-IMG website (26; http://img.jgi.doe.gov/cgi-bin/w/main.cgi). Microarray results are available through the NCBI GEO site under series accession numbers GSE31521 and GSE31522.

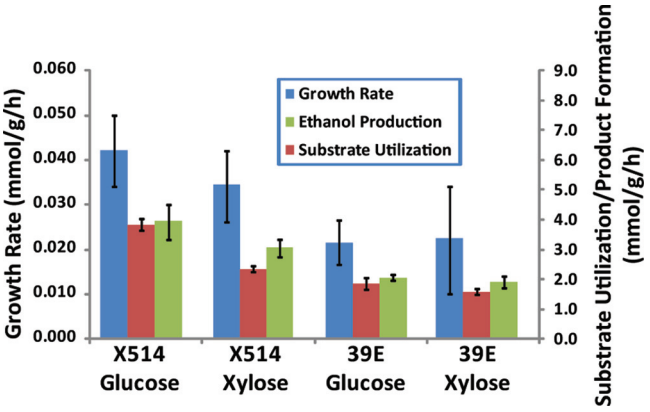

FIG. 1. Physiological comparison of X514 and 39E grown on glucose and xylose and of corresponding ethanol yields. Error bars represent errors for three replicates. The samples from which these data were derived were used for flux analyses as described in the main text.

RESULTS

Strain X514 produces more ethanol from xylose than does strain 39E. *T. pseudethanolicus* 39E was originally isolated from Octopus Spring in Yellowstone Park (30), a relatively nutritionally stable environment, whereas X514 was isolated from the presumably nutrient-poor deep subsurface (33) (see Table S1 in the supplemental material). Though the growth rates on xylose were not significantly different (0.022 ± 0.012 mmol/g/h for 39E and 0.034 ± 0.008 mmol/g/h for X514), X514 exhibited higher rates of substrate utilization (1.5 ± 0.1 mmol/g/h for 39E and 2.3 ± 0.1 mmol/g/h for X514) and higher ethanol yields (1.9 ± 0.2 mmol/g/h for 39E and 3.0 ± 0.3 mmol/g/h for X514) than 39E when the strains were grown on xylose (Fig. 1; see Fig. S1). Multiple combinations of cocultures using different *Thermoanaerobacter* and *C. thermocellum* strains showed that the combination of X514 and *C. thermocellum* LQRI gave the best results for cellulose degradation rates and ethanol yields. Average nucleotide identity (ANI) (19) values suggest that X514 and 39E are likely members of the same species (97.6% ANI; a 97% ANI cutoff is used to define species), and thus they are expected to show similar but not identical genome structures and metabolisms. Thus, the genomic differences exhibited by the two strains could potentially correlate with the phenotypes observed in mono- and coculture. The completed genomes of the two strains were thus compared to identify genetic differences which could manifest as the observed phenotypic differences in substrate utilization and product formation (15).

Energy metabolism and cofactor biosynthesis. Cellular energetics in *Thermoanaerobacter* appear to be based on the establishment of transmembrane Na^+ gradients, as observed in multiple *Clostridia* species (Fig. 2) (37, 40). *Thermoanaerobacter* species employ a vacuolar V-type ATPase which likely operates as a sodium efflux pump during fermentation, as well as ferredoxin-linked Na^+-translocating Rnf complexes (Fig. 2, reactions 1 and 2). Predicted Na^+-translocating decarboxylase enzymes (Fig. 2, reactions 3 and 4) and Na^+/H^+ antiporters

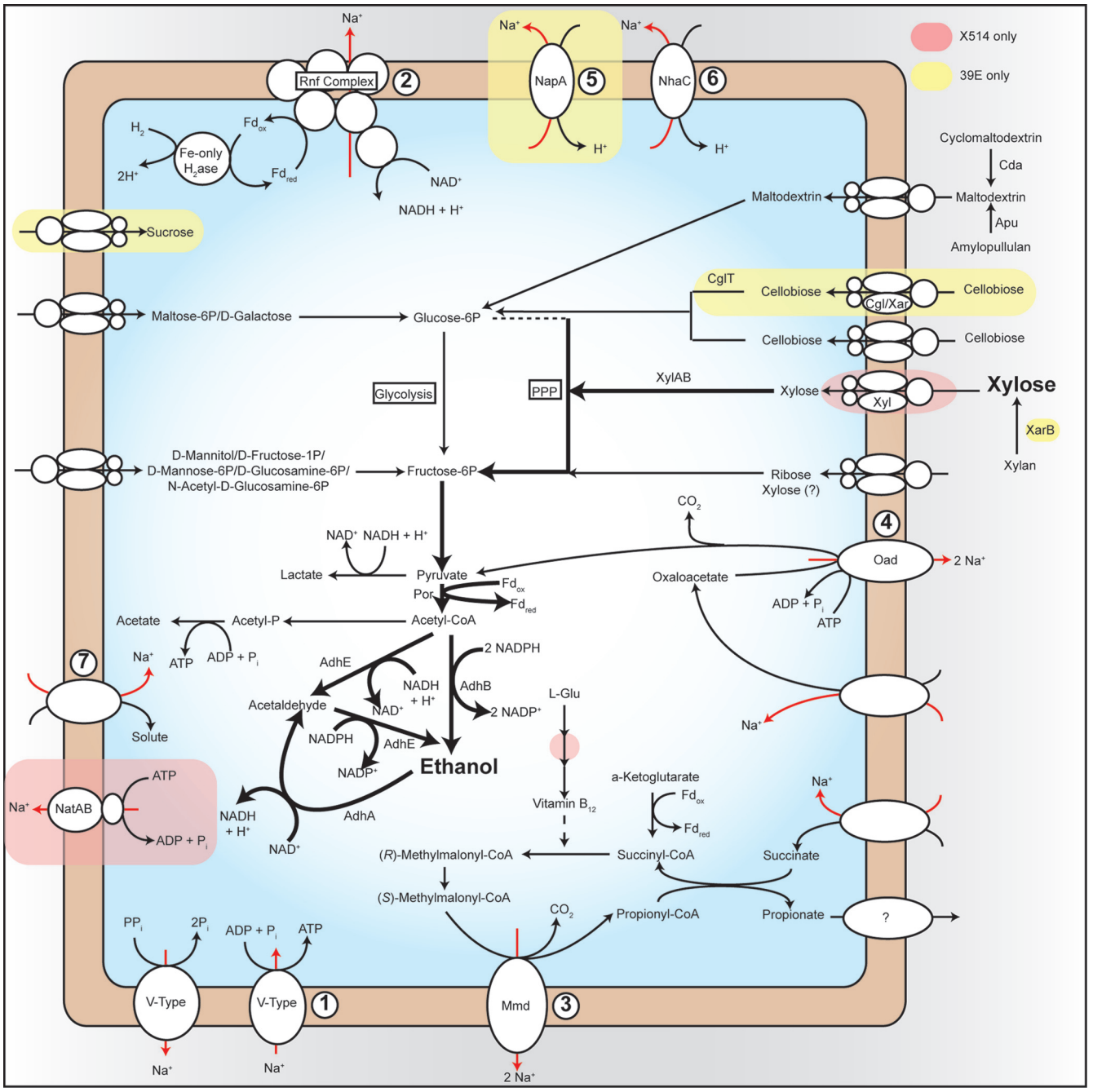

FIG. 2. Central carbon and energy metabolism of *Thermoanaerobacter* strains. Key pathways and enzymes related to ethanol production are indicated. Enzymes/pathways unique to a specific lineage are colored as follows: pink, X514 only; yellow, 39E only. The dotted line in the pentose phosphate pathway indicates an incomplete pathway, while the dotted line from vitamin B_{12} to methylmalonyl-CoA mutase indicates the requirement of this cofactor for enzyme activity.

(Fig. 2, reactions 5 and 6) likely serve to maintain H^+ and Na^+ gradients, while Na^+/solute symporters may allow for nonspecific influx of solutes, including sugars (Fig. 2, reaction 7). While numerous *Clostridia* species employ Na^+ energetics, one particular aspect of *Thermoanaerobacter* Na^+-based energetics was predicted to have a significant, albeit indirect, effect on ethanol yields.

The key to energetics in *Thermoanaerobacter* appears to be Na^+-translocating decarboxylation reactions, in particular those with methylmalonyl-coenzyme A (methylmalonyl-CoA) decarboxylase (Fig. 2, reaction 3). A key enzyme in methylmalonyl-CoA metabolism is methylmalonyl-CoA mutase, which requires vitamin B_{12} as a cofactor and feeds methylmalonyl-CoA to methylmalonyl-CoA decarboxylase. A total of 21 genes encoding putative vitamin B_{12}-dependent enzymes were identified in the two *Thermoanaerobacter* genomes, in-

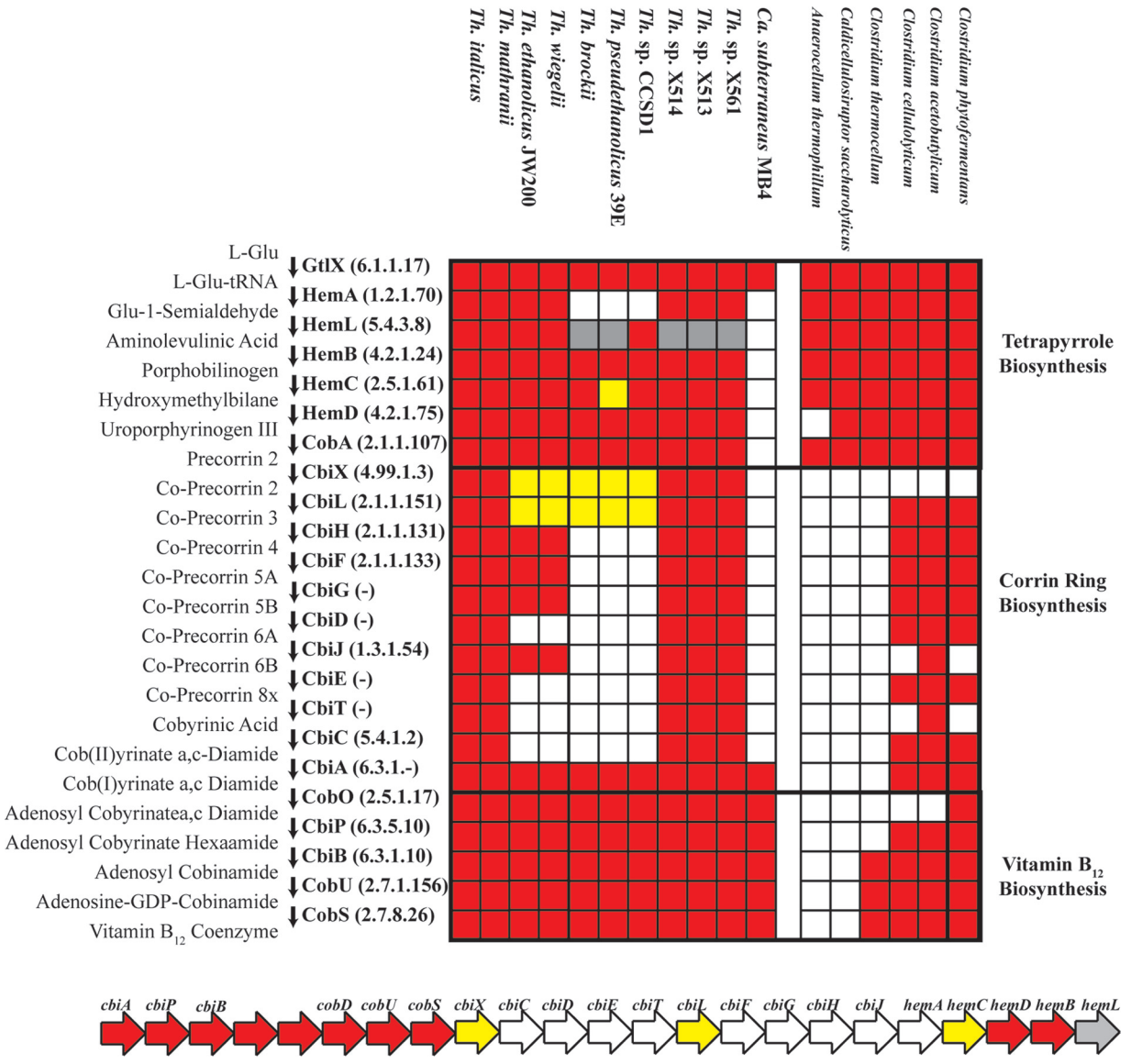

FIG. 3. Comparison of vitamin B_{12} biosynthesis genes in *Thermoanaerobacter* species. The grid is arranged with columns representing *Thermoanaerobacter* and outlier *Clostridia* species and rows representing protein-encoding genes of the vitamin B_{12} biosynthesis pathway. Each square on the grid thus represents a given vitamin B_{12} biosynthesis gene for a given species and is colored depending on whether the gene is present (red), present but degraded or truncated (yellow), presumed absent (white), or presumed to be part of a gene fusion (gray; in this case representing a bifunctional siroheme synthase/glutamate-1-semialdehyde aminotransferase). Below the grid is the complete vitamin B_{12} biosynthesis operon structure of X514, with genes colored as in the grid to indicate that the orthologous 39E gene is present, absent, or degraded. *Caldanaerobacter subterraneus* subsp. *tengcongensis* sp. MB4 was originally classified as *Thermoanaerobacter tengcongensis* MB4 when the genome was sequenced, but the strain has since been reclassified.

cluding methylmalonyl-CoA mutase, methionine synthase, and homocysteine *S*-methyltransferase genes (see Table S2 in the supplemental material). X514 carries a complete vitamin B_{12} biosynthesis operon, whereas 39E has lost genes encoding the anaerobic corrin ring biosynthesis portion of the pathway (*cbiCDT* and *cbiFGHJ-hemA*) (Fig. 3). Extension of this analysis to all currently available *Thermoanaerobacter* genomes suggests that the *de novo* vitamin B_{12} pathway is likely ancestral to the *Thermoanaerobacter* genus and that the anaerobic corrin ring biosynthesis portion of the pathway has been lost in the 39E/*Thermoanaerobacter brockii*/CCSD1 lineage (Fig. 3). This loss appears to have occurred in two steps, with the *cbiCDET* cluster first lost in the *Thermoanaerobacter ethanolicus*/*pseudethanolicus* lineage and the *cbiFGHJ-hemA* genes lost in the *Thermoanaerobacter pseudethanolicus* lineage (Fig. 3). Consistent with this hypothesis, *in silico* analysis of the sirohydrochlorin cobalt chelatase (*cbiX*) genes of the *Thermoanaerobacter* strains showed that in X514, the *cbiX* gene is adjacent to the precorrin-8X methylmutase (*cbiC*) gene, which is the first gene in the cluster missing from 39E (Fig. 3). In 39E, the *cbiX* gene lacks approximately 120 nucleotides corresponding to the N terminus of the expressed full-length protein, resulting in the

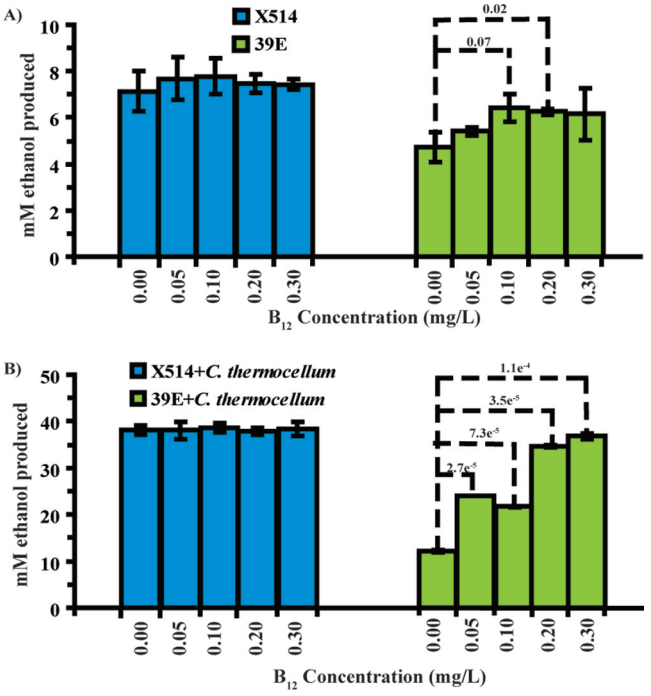

FIG. 4. Effects of exogenous vitamin B_{12} on ethanol yields for *Thermoanaerobacter* strains (X514 and 39E) grown on 50 mM xylose (A) and *Thermoanaerobacter-C. thermocellum* cocultures grown on 1% cellulose (B). Vitamin B_{12} concentrations in the media are indicated on the *x* axis (mg/liter). The data represent the means for three replicates, with error bars presenting standard deviations. Significant differences in mean ethanol yields were calculated using the one-tailed paired *t* test, and *P* values of <0.1 are shown above the dotted line.

gene being located adjacent to the truncated *cbiL* gene (Fig. 3). Interestingly, some strains show a putative gene fusion encoding siroheme synthase and glutamate-1-semialdehyde aminotransferase, whereas in other strains these functions appear to be encoded by different genes (Fig. 3). Thus, disruption of the Na^+ gradient in strain 39E by removing vitamin B_{12} from the growth medium would be expected to impair growth and ethanol yields, while X514 would be unaffected. Furthermore, it has been observed previously that increased ethanol yields from cellulose fermentation can be achieved in some strains of *C. thermocellum* by the addition of exogenous vitamin B_{12} (35). Thus, the ability to synthesize vitamin B_{12} *de novo* could potentially enhance the ability of X514 to maintain energy metabolism under nutrient-poor conditions and may provide a needed cofactor in cellulolytic cocultures with *C. thermocellum* which would enhance cellulose fermentation.

To test this hypothesis, X514 and 39E were grown individually on xylose and in coculture with *C. thermocellum* LQRI on cellulose, with differing concentrations of exogenous vitamin B_{12} added to the medium (Fig. 4). In *Thermoanaerobacter* monocultures grown on 50 mM xylose, ethanol yields in X514 were largely independent of vitamin B_{12} concentrations even when no vitamin B_{12} was present in the medium, while yields decreased significantly (~60% of maximum yield) in 39E when no exogenous vitamin B_{12} was present (Fig. 4A). This effect was much more pronounced when the strains were grown in coculture with *C. thermocellum* on 1% cellulose (Fig. 4B). Ethanol yields were strikingly lower (~25% of maximum yield) in the 39E coculture with *C. thermocellum* on 1% cellulose when no exogenous vitamin B_{12} was added, and yields approached those of the X514 coculture only when 2- to 3-fold more vitamin B_{12} was added to the growth medium (Fig. 4B). In contrast, ethanol yields remained unchanged in X514-*C. thermocellum* cocultures, even when no exogenous vitamin B_{12} was added to the growth medium (Fig. 4B). These results suggest that X514 synthesizes vitamin B_{12} *de novo*, as predicted, and that this ability has a significant effect on ethanol yields in both mono- and coculture.

Central carbon metabolism. Differences in energetics between the strains provide a clear, albeit indirect, link between differing genotypes and ethanol yields. However, other genotypic differences appear to provide a more direct correlation to ethanol yield phenotypes. An identical core set of Embden-Meyerhof-Parnas (EMP) pathway genes were identified in both strains (Fig. 2). 39E carries additional ED pathway genes, though neither strain appears to employ the modified ED pathway observed in many *Clostridia* species (6). Multiple pathways for metabolizing hexose and pentose sugars (glucose, fructose, xylose, maltose, ribose, and galactose) are present, and xylose metabolism appears to occur exclusively via the traditional xylose isomerase-xylulokinase-pentose phosphate pathway. Both strains also carry genes annotated as encoding pyruvate decarboxylase subunit B, a key enzyme in ethanol production in *Zymomonas* strains (3, 38), but sequence homology suggests that this enzyme is more likely an oxaloacetate decarboxylase and thus not expected to be involved directly in ethanol formation. In addition, both strains utilize multiple complex carbohydrates, including amylopectin, cyclodextrin, cellodextrin (cellobiose, cellotriose, cellotetraose, and cellopentaose), pectin, and pullulan, while only 39E is capable of metabolizing sucrose and degrading xylan (see below). Thus, the phenotypic difference in ethanol yields between the strains does not appear to be due to observable differences in carbon metabolism pathways between the strains.

A total of 46, 45, and 48 presumptive carbohydrate-active enzymes (7) were identified in 39E, X514, and MB4, respectively (see Table S3 in the supplemental material). While many of the relevant genes correspond to characterized carbohydrate metabolism enzymes (e.g., amylopullulanase, etc.), several remain uncharacterized. Whole-transcriptome analysis of each of the strains showed that many of these genes were actively expressed when cells were grown with 50 mM xylose (see Table S3). SIGI-HMM analysis (47) suggests that X514 has acquired (via lateral gene transfer) a lineage-specific, ~1-kb gene cluster encoding additional copies of xylulokinase, transketolase, transaldolase, alcohol dehydrogenase (ADH), aldehyde dehydrogenase, acetate kinase, and bacterial microcomponent shell components that could potentially facilitate carbon metabolism in X514 by alleviating metabolic bottlenecks between xylose and ethanol. However, microarray analysis suggests that these genes are not actively expressed when cells are grown on xylose and thus likely have no significant effect on ethanol production yields in X514.

A total of 7 and 9 alcohol dehydrogenase genes were identified in 39E and X514, respectively (Table 1). Both strains encode an NADPH-dependent secondary alcohol dehydrogenase (*adhB*) predicted to catalyze the terminal step in ethanol

TABLE 1. Alcohol dehydrogenase genes carried by strains 39E and X514[a]

adh gene product	39E		X514	
	Locus tag	GeneID	Locus tag	GeneID
Iron-containing ADH			Teth514_0145	167038817
	Teth39_1979	167038363	Teth514_0241	167038910
			Teth514_0564	167039223
Bifunctional secondary ADH/aldehyde dehydrogenase (*adhE*)	Teth39_0206	167036636	Teth514_0627	167039286
Bifunctional secondary ADH/acetyl-CoA thioesterase (*adhB*)	Teth39_0218	167036648	Teth514_0653	167039312
Primary ADH (*adhA*)	Teth39_0220	167036650	Teth514_0654	167039313
Short-chain ADH	Teth39_2190	167038570	Teth514_1808	167040441
ADH with GroES domain	Teth39_0878	167037292	Teth514_1882	167040513
Iron-containing ADH			Teth514_1935	167040564
	Teth39_1597	167037999		

[a] Orthologous genes are presented in rows.

production in *Thermoanaerobacteraceae* species (also found in other *Clostridia* species, such as *Thermosinus carboxidivorans*, *Clostridium botulinum*, and *Ethanoligenens harbinense*) and an NADPH-dependent primary alcohol dehydrogenase (*adhA*) involved in ethanol catabolism (3a, 4). In both strains, these genes are located adjacent to each other (in opposite orientation) and adjacent to a predicted $NADP^+$ reductase gene (Teth514_0652). Both strains also encode a bifunctional secondary alcohol dehydrogenase/aldehyde dehydrogenase (*adhE*) (Fig. 2; Table 1) implicated in ethanol production in *Thermoanaerobacter* (32) and in butanol production in *Clostridium acetobutylicum* (11). This complement of ADH enzymes is novel compared to that of most ethanol-producing bacteria, which typically employ a primary ADH (*adhA*) as the terminal step in ethanol production (4). X514 also carries three additional lineage-specific iron-only *adh* genes compared to 39E, which carries only one additional lineage-specific *adh* gene (Table 1). Microarray analysis of 39E and X514 grown on xylose showed that the three major *adh* genes (*adhA*, *adhB*, and *adhE*) were expressed in both strains at different xylose concentrations (see Fig. S2 in the supplemental material), a result consistent with previous analysis of the expression of these genes (4). *adhB* in particular appears to be expressed highly in both strains at 10 mM xylose, suggesting that AdhB is the key terminal enzyme involved in ethanol production (see Fig. S2). Interestingly, the *adhA* gene encoding the catabolic primary ADH is expressed much more highly in 39E than in X514, and the conserved ADH encoded by Teth514_0241/Teth39_1979 is expressed more highly in 39E under the same growth conditions (see Fig. S2). AdhB has been proposed as a biotechnological target for enhancing ethanol yields in *Thermoanaerobacter* strains, and the observation of differential expression of *adh* genes in response to xylose is intriguing, but no definitive evidence linking differences in *adh* expression to strain-specific ethanol yields was observed.

Flux analyses of carbon metabolism pathways. Genomic differences alone were insufficient to link the observed physiological phenotypes to genotypic traits. To extend the analyses, ^{13}C-assisted flux analysis was conducted for both strains to quantify the relative and absolute flux rates from xylose to ethanol for the two strains and to complement a previous metabolomic analysis of X514 (Fig. 5; see Table S4 in the supplemental material) (10). The measured extracellular fluxes (i.e., ethanol production, growth rates, and the consumption rates of glucose and xylose) in X514 were >50% higher than those in 39E (Fig. 1). Using glucose as the carbon source (see below) revealed that the oxidative pentose phosphate pathway (a common pathway associated with cellular NADPH production) in X514 exhibited very low activity for glucose metabolism (relative flux, <5% of total flux) (Fig. 3), consistent with the lack of a 6-phosphogluconolactonase (EC 3.1.1.31) gene in the *Thermoanaerobacter* genomes (Fig. 2 and 5). The flux through the oxidative pentose phosphate pathway was moderately higher in 39E (relative flux, ~12% of total flux), suggesting that X514 might be more dependent on alternate pathways for NADPH production. For example, based on sequence annotation, X514 could use pyruvate:ferredoxin oxidoreductase (Teth514_0781) to produce reduced ferredoxin and a ferredoxin-$NADP^+$ reductase (Teth514_0652) to reduce $NADP^+$.

As expected from physiological characterization, a substantially higher carbon overflow (2- to 3-fold for glucose culture and 5- to 10-fold for xylose culture) was channeled to ethanol than to acetate and lactate for both strains (e.g., 1 mol glucose can generate about 1 mol ethanol by X514) (Fig. 5). Furthermore, the intracellular carbon flux distributions after normalization by carbon substrate (i.e., glucose or xylose) consumption rates were not significantly different between these two strains (Fig. 5; see Table S4 in the supplemental material), though absolute growth rates and metabolite production in X514 were much higher than those in 39E under both glucose and xylose growth conditions. Such observations suggest that the two phylogenetically closely related strains share similar regulation of central metabolic pathways, but unique genotypes of X514 likely significantly enhance ethanol synthesis rates.

Carbon transporters. Since observed differences in carbon metabolism pathways were insufficient to explain increased absolute flux rates in X514, the complement of sugar transporters (for xylose in particular) was examined. *Thermoanaerobacter* strains carry a variety of genes for carbon transporters, encompassing primarily ABC-type (120 and 134 genes in 39E and X514, respectively) and phosphotransferase system (PTS) (28 and 29 genes in 39E and X514, respectively) transporters (Fig. 2). These transporters are specific for hexoses (glucose, galactose, fructose, and mannose), pentoses (xylose and ribose), disaccharides (sucrose, lactose, cellobiose, and maltose), oligosaccharides (maltodextrin), and unspecified sugar and polysaccharide transporters (Fig. 2). A major difference in

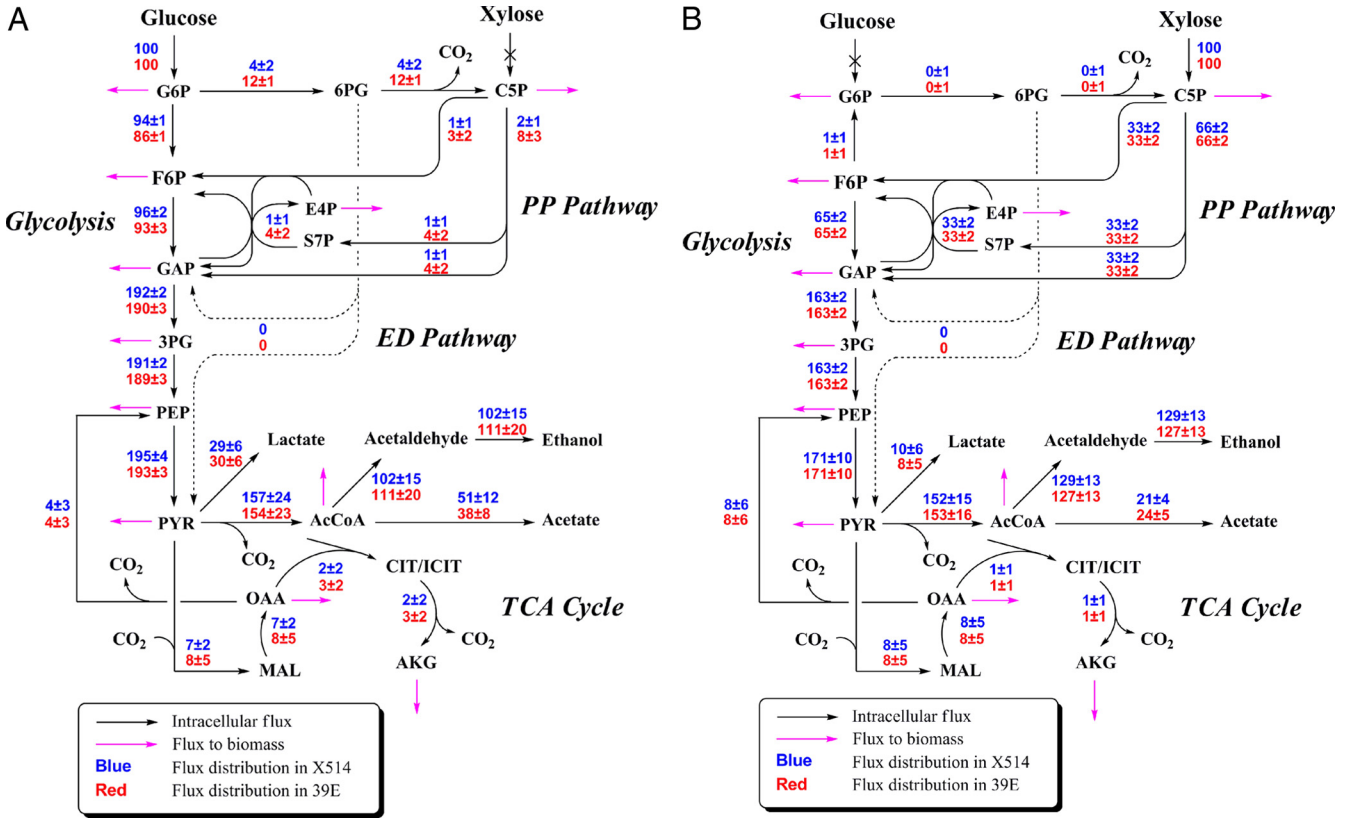

FIG. 5. Relative flux distributions for glucose (A) and xylose (B) metabolism. The actual carbon utilization mole rates were normalized to 100. The flux distributions describe the utilization of "100-mol" carbon sources by each pathway. The numbers in blue show the fluxes of X514, and the numbers in red show the fluxes of 39E. The fluxes to biomass are shown with pink arrows. Abbreviations: AcCoA, acetyl-coenzyme A; CIT, citrate; E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; G6P, glucose-6-phosphate; 6PG, 6-phosphogluconate; ICIT, isocitrate; MAL, malate; OAA, oxaloacetate; AKG, 2-oxoglutarate; PEP, phosphoenolpyruvate; 3PG, 3-phosphoglycerate; C5P, ribose-5-phosphate (or ribulose-5-phosphate) and xylulose-5-phosphate; S7P, sedoheptulose-7-phosphate; SUC, succinate; GAP, 3-P-glyceraldehyde; PYR, pyruvate.

carbon transport between these two strains is the complement of putative xylose transport genes in X514 (Fig. 6).

Genes encoding a bifunctional xylanosidase/arabinosidase (*xarB*), associated cellobiose transporters (*xarG*), and β-glucosidase (*cglT*) have previously been cloned and characterized for *Thermoanaerobacter ethanolicus* JW200 (25). The genomes of 39E and X514 were compared to genomes of eight additional *Thermoanaerobacter* species to determine if the variability of this region extends across multiple *Thermoanaerobacter* species and if this variability can be linked to observed physiological phenotypes (Fig. 6). The JW200 *xarBG-cglFT* gene cluster is conserved in 39E and is associated with genes of a ribose transport operon, with xylose metabolism (*xylAB*), and with mobile elements (Fig. 6). In contrast, X514 utilizes a xylose-specific XylFGH system but lacks the *xarB* xylanosidase gene (Fig. 6). *Thermoanaerobacter italicus* and *Thermoanaerobacter mathranii*, which represent a distinct lineage within *Thermoanaerobacter*, carry both operons as well as multiple lineage-specific xylanase genes (Fig. 6). Transcriptomic analysis revealed that all of the genes of the *xyl* operon in X514 were highly expressed when cells were grown with xylose at a low concentration (10 mM) compared to a high concentration (200 mM), but the *cgl-xar* operon in 39E was largely unresponsive to xylose, consistent with its presumed function as a cellobiose transport operon (see Fig. S4 and Table S5 in the supplemental material). The presence of a dedicated xylose transport system may explain why X514 can grow on xylose concentrations as low as 1 mM while 39E cannot, and it may also explain the observed increase in absolute flux rates from xylose to ethanol in X514 and raises the question of how xylose is taken up by 39E. A conserved ribose-specific ABC transport operon is located adjacent to the *xyl* and *xar* operons and could potentially display xylose uptake activity as well. However, transcriptomic analysis does not suggest that this operon is upregulated in response to xylose in 39E (see Table S5). A fructose/mannose/sorbose-specific ABC transport operon showed >2-fold increased expression in 39E cells grown on 10 mM xylose versus 200 mM xylose, suggesting that this operon could potentially have xylose uptake activity (see Table S5). Alternatively, 39E may employ less specific uptake systems, possibly via Na^+ gradient-mediated solute uptake.

DISCUSSION

The study of *Thermoanaerobacter* species has long been driven by interest in the ability to efficiently produce ethanol at high yields from both glucose and xylose (21, 32, 50) and in the potential applications in CBP, a concept offering the most

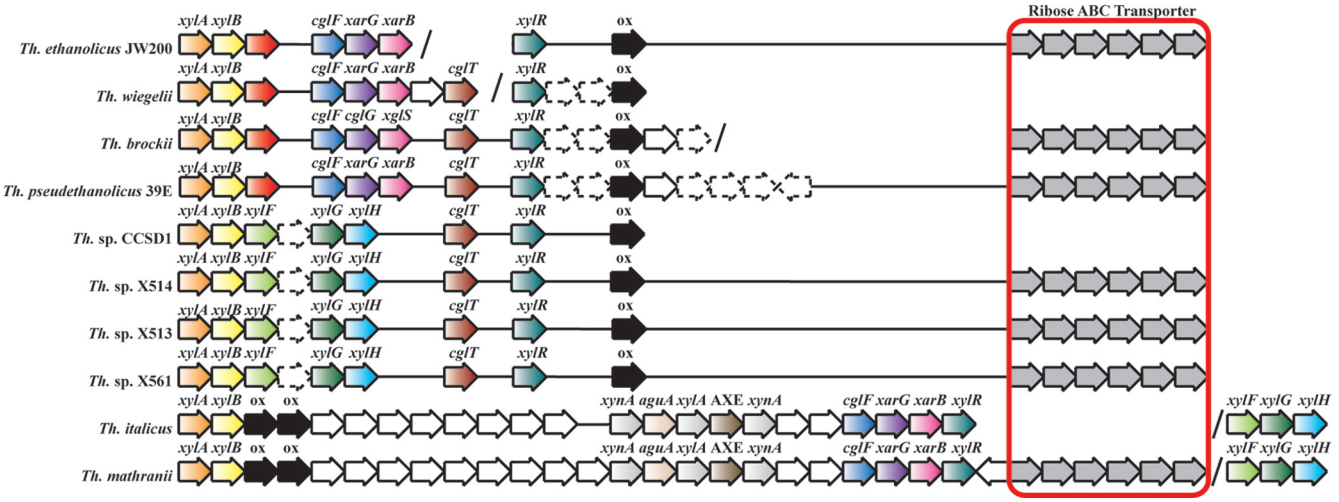

FIG. 6. Xylan degradation and xylose transport genes of *Thermoanaerobacter* species. For a given strain, genes are considered to be contiguous except where indicated by a slash (indicating a contig boundary or large genomic region). Orthologous genes are colored similarly and aligned vertically as much as possible. Open reading frames with dotted outlines indicate mobile elements such as transposons and integrase genes. The six-gene cluster colored gray represents a conserved ribokinase/ribose ABC transporter gene cluster. "ox" refers to a conserved oxidoreductase gene of unknown function. White arrows indicate hypothetical genes or miscellaneous genes not obviously relevant to xylose metabolism or transport.

promise in addressing central challenges to cellulosic bioethanol production (23). Several features in particular make *Thermoanaerobacter* species attractive for integration into CBP schemes, including (i) the ability to efficiently ferment both hexose and pentose sugars (particularly xylose) (21) to ethanol in their natural state, in contrast to non-pentose-fermenting strains such as *Zymomonas* (38); (ii) relatively high ethanol yields generated from sugar fermentation (14); (iii) simplification of nutrient requirements due to the capability for *de novo* cofactor (i.e., vitamin B_{12}) biosynthesis; (iv) a high growth temperature which is better suited for industrial processing of ethanol and minimization of microbial contamination than those for mesophilic strains such as engineered *E. coli*; (v) a high substrate affinity coupled with nearly complete substrate utilization; (vi) ease of growth in microbial consortia; and (vii) a unique terminal ethanol production pathway involving novel lineage-specific alcohol dehydrogenase enzymes which may positively affect fermentation balance and ethanol yields (3a, 4, 32). Although *Thermoanaerobacter* species have great biotechnological potential for cellulosic ethanol production, their genetics, biochemistry, and physiology are only partially understood. The sequencing of multiple *Thermoanaerobacter* genomes coupled to experimental analyses allowed for the first comprehensive attempt to link genotypic differences to phenotypic effects on substrate usage and product yields. This information in turn is vital to future attempts to engineer these strains for integration into CBP processes.

The use of *Thermoanaerobacter* species for ethanol production presents several major challenges. First, ethanol yields from *Thermoanaerobacter*, while high, do not currently meet desired industrial standards. Accumulation of alternative fermentation end products, such as lactate and acetate, reduces the efficiency of substrate conversion to ethanol, though *Thermoanaerobacter* strains are naturally more efficient at maintaining a high ratio of ethanol formed to other fermentation products than other strains commonly employed in industrial fermentation processes. Genetic engineering of strains by blocking metabolic pathways of acetate and lactate formation is a common mechanism for obtaining higher ethanol yields (17, 39, 41), though such efforts have proven difficult with *Thermoanaerobacter* species due to the lack of reliable genetic systems. Second, *Thermoanaerobacter* species efficiently convert sugars to ethanol at relatively low substrate concentrations (<10 g/liter), considerably lower than the desired, industry-scale substrate levels (>50 g carbohydrate/liter). Understanding how substrate concentrations affect ethanol yields and how they are regulated at the molecular level is important for the optimization of biotechnological processes. Third, pretreatment of lignocellulosic materials typically results in the formation of inhibitory compounds, including weak acids, furans, and phenolic compounds (31). Knowledge on how various inhibitors affect ethanol production at the molecular level is also critical to developing resistant strains for more efficient conversion of lignocellulosic materials to ethanol. Fourth, certain strains require growth factors (e.g., vitamins) susceptible to high temperatures prevalent under thermophilic conditions, which could hinder the efficiency of high-density thermophilic fermentation. Finally, rapid cell lysis has been observed in *Thermoanaerobacter* cultures reaching high densities, making it difficult to maintain the high process efficiencies required by industry-scale fermentation operations. The availability of whole-genome sequences of X514 and 39E provides the basis for further genetic and metabolic engineering for desired biotechnological applications.

Strains X514 and 39E are phylogenetically closely related, but significant differences in physiology have been observed with respect to ethanol production. For example, X514 produces higher yields of ethanol from xylose, and the stable coculture of *C. thermocellum* LQRI and X514 produces substantially more ethanol than the corresponding *C. thermocel-*

lum LQRI-39E coculture. The whole-genome sequence comparison and experimental analyses in this study have revealed key genotypic differences between these two strains which may explain these phenotypes. In contrast to 39E, X514 encodes a xylose-specific transporter, additional lineage-specific alcohol dehydrogenase enzymes, and a complete vitamin B_{12} *de novo* biosynthesis pathway, which is critical for maintenance of Na^+ gradient-based energetics. Additional genes encoding xylose metabolism and pentose phosphate enzymes in X514 are present but do not appear to significantly affect carbon flux in X514. While relative carbon flux from xylose appears to be the same in each strain, the observed greater absolute flux in X514 and the ability of X514 to grow at lower xylose concentrations than those for growth of 39E may be explained by the employment of xylose-specific Xyl transporters in X514. These transporters may allow for more efficient uptake and utilization of xylose in this strain than in 39E, which must presumably utilize a more generic xylose uptake system (i.e., Na^+ gradient-linked solute transporters, etc.). Furthermore, the capacity for *de novo* synthesis of vitamin B_{12} by X514 appears to be a key factor in maintaining high ethanol yields in cellulosic cocultures compared to those with 39E. Finally, transcriptomic analysis shows that alcohol dehydrogenase genes are expressed differentially in both strains under different growth conditions but that the AdhB enzyme is likely the key ADH enzyme involved in ethanol production and is thus a logical candidate for future experimental analysis and engineering efforts. Recent studies of the closely related *Thermoanaerobacteraceae* species *Thermoanaerobacterium saccharolyticum* have shown that metabolic and genetic engineering can be used successfully to enhance ethanol yields from sugar, to yields of 37 g/liter in the case of *T. saccharolyticum* (39). While similar results would be expected from engineered *Thermoanaerobacter* species, such efforts have been hampered by the relative genetic recalcitrance of strains. Further systematic in-depth functional genomic studies are needed to fully understand how these genotypic differences contribute to the observed phenotypic differences in sugar utilization and ethanol production. Such information will be essential for the metabolic engineering of *Thermoanaerobacter* strains for high ethanol production. In addition, the unique genes observed in *Thermoanaerobacter* (e.g., alcohol dehydrogenase and xylose transporter genes) could also be useful for genetically engineering non-xylose-utilizing microorganisms for xylose metabolism and/or improved ethanol yields (17, 20, 36, 39, 40, 41).

In conclusion, *Thermoanaerobacter* species are important in biotechnological applications of biofuel production. However, implementation of *Thermoanaerobacter* species into CBP schemes requires a detailed, systems-level understanding of the genetics, biochemistry, physiology, and ecology of this group of bacteria. The comparative whole-genome sequence analyses of X514 and 39E presented here provide valuable insights into the nature of carbon metabolism, energy metabolism, carbon transporters, ethanol production, and cofactor biosynthesis in these *Thermoanaerobacter* strains. Such information is important not only for implementing *Thermoanaerobacter* strains into industrially viable CBP schemes through further developing and/or optimizing more efficient strains but also for potentially engineering other, non-xylose-utilizing and/or ethanol-producing microorganisms.

ACKNOWLEDGMENTS

We thank D. Klingman at Oak Ridge National Laboratory for preparation of genomic DNA for sequencing.

This material is based on work supported by the Oklahoma Bioenergy Center (J.Z.), the U.S. Department of Energy Joint BioEnergy Institute (JBEI) (J.Z.), the Chinese Academy of Sciences (J.X.), and the National Science Foundation EPSCoR program under grant EPS-0814361 (J.Z.). The genome sequencing work was performed under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental Research Program and by the University of California, Lawrence Berkeley National Laboratory, under contract DE-AC02-05CH11231, the Lawrence Livermore National Laboratory, under contract DE-AC52-07NA27344, and the Los Alamos National Laboratory, under contract DE-AC02-06NA25396, as previously described (15).

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

J.Z. and M.W.F. developed the original concept. J.Z. and T.J.P. isolated strain X514. C.L.H. conducted the comparative computational analysis, designed experiments, and analyzed results. Y.D. and Q.T. aided in computational analyses. Z.H., L.W., and J.V.N. oversaw microarray design and construction. L.L. and H.M. conducted growth analysis of strains on xylose, and L.L. and A.Z. conducted microarray analysis of strains on xylose. M.W.F. and B.D.R. conducted growth analysis of strains on cellodextrins and other sugars. Y.J.T., X.F., and Z.Z. conducted flux analysis. J.Z. and C.L.H. prepared the manuscript, with contributions from Y.J.T., M.W.F., Q.H., Z.H., J.X., T.J.P., and J.W.

The authors declare that they have no competing interests.

REFERENCES

1. **Ballesteros, I., et al.** 1991. Selection of thermotolerant yeasts for simultaneous saccharification and fermentation (SSF) of cellulose to ethanol. Appl. Biochem. Biotechnol. **28–29:**307–316.
2. **Bayer, E. A., R. Lamed, and M. E. Himmel.** 2007. The potential of cellulases and cellulosomes for cellulosic waste management. Curr. Opin. Biotechnol. **18:**237–245.
3. **Bräu, B., and H. Sahm.** 1986. Cloning and expression of the structural gene for pyruvate decarboxylase of *Zymomonas mobilis* in *Escherichia coli*. Archives of gene encoding the *Thermoanaerobacter ethanolicus* 39E secondary-alcohol dehydrogenase and biochemical characterization of the enzyme. Biochem. J. **316:**115–122.

3a. **Burdette, D. S., C. Vielle, and J. G. Zeikus.** 1996. Cloning and expression of the gene encoding the *Thermoanaerobacter ethanolicus* 39E secondary-alcohol dehydrogenase and biochemical characterization of the enzyme. Biochem. J. **316:**115–122.

4. **Burdette, D., and J. G. Zeikus.** 1994. Purification of acetaldehyde dehydrogenase and alcohol dehydrogenases from *Thermoanaerobacter ethanolicus* 39E and characterization of the secondary-alcohol dehydrogenase as a bifunctional alcohol dehydrogenase–acetyl-CoA reductive thioesterase. Biochem. J. **302:**163–170.
5. **Chhabra, S. R., et al.** 2006. Global analysis of heat shock response in *Desulfovibrio vulgaris* Hildenborough. J. Bacteriol. **188:**1817–1828.
6. **Conway, T.** 1992. The Entner-Doudoroff pathway: history, physiology and molecular biology. FEMS Microbiol. Lett. **103:**1–28.
7. **Coutinho, P. M., and B. Henrissat.** 1999. Carbohydrate-active enzymes: an integrated database approach, p. 3–12. *In* H. J. Gilbert, G. Davies, B. Henrissat, and B. Svensson (ed.), Recent advances in carbohydrate bioengineering. The Royal Society of Chemistry, Cambridge, United Kingdom.
8. **Demain, A. L., M. Newcomb, and J. H. D. Wu.** 2005. Cellulase, clostridia, and ethanol. Microbiol. Mol. Biol. Rev. **69:**124–154.
9. **Fang, Z.** 2010. Enhanced role of the co-culture of thermophilic anaerobic bacteria on cellulosic ethanol. Huan Jing Ke Xue **4:**1059–1065.
10. **Feng, X., et al.** 2009. Characterization of the central metabolic pathways in *Thermoanaerobacter* sp. X514 via isotopomer-assisted metabolite analysis. Appl. Environ. Microbiol. **75:**5001–5008.
11. **Fontaine, L., et al.** 2002. Molecular characterization and transcriptional analysis of *adhE2*, the gene encoding the NADH-dependent aldehyde/alcohol dehydrogenase responsible for butanol production in alcohologenic cultures of *Clostridium acetobutylicum* ATCC 824. J. Bacteriol. **184:**821–830.
12. **Freier, D., C. P. Mothershed, and J. Wiegel.** 1988. Characterization of *Clostridium thermocellum* JW20. Appl. Environ. Microbiol. **54:**204–211.

13. **Hamelinck, C. N., G. V. Hooijdonk, and A. P. C. Faaij.** 2005. Ethanol from lignocellulosic biomass: techno-economic performance in short-, middle- and long-term. Biomass Bioenergy **28:**384–410.
14. **He, Q., P. M. Lokken, S. Chen, and J. Zhou.** 2009. Characterization of the impact of acetate and lactate on ethanolic fermentation by *Thermoanaerobacter ethanolicus*. Bioresour. Technol. **100:**5955–5965.
15. **Hemme, C. L., et al.** 2010. Sequencing of multiple *Clostridia* genomes related to biomass conversion and biofuels production. J. Bacteriol. **192:**6494–6496.
16. **Houghton, J., S. Wheatherwax, and J. Ferrell.** 2006. Breaking the biological barriers to cellulosic ethanol: a joint research agenda, p. 206. U.S. Department of Energy, Washington, DC.
17. **Ingram, L. O., et al.** 1998. Metabolic engineering of bacteria for ethanol production. Biotechnol. Bioeng. **58:**204–214.
18. **Jiang, H., Z. He, Z. Fang, and J. Zhou.** 2008. Abstr. 108th Gen. Meet. Am. Soc. Microbiol., Boston, MA, abstr. O–009.
19. **Konstantinidis, K. T., and J. M. Tiedje.** 2005. Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. U. S. A. **102:** 2567–2572.
20. **Kuyper, M., et al.** 2005. Metabolic engineering of a xylose-isomerase-expressing *Saccharomyces cerevisiae* strain for rapid anaerobic xylose fermentation. FEMS Yeast Res. **5:**399–409.
21. **Lacis, L. S., and H. G. Lawford.** 1988. Ethanol production from xylose by *Thermoanaerobacter ethanolicus* in batch and continuous culture. Arch. Microbiol. **150:**48–55.
22. **Li, X., Z. He, and J. Zhou.** 2005. Selection of optimal oligonucleotide probes for microarrays using multiple criteria, global alignment and parameter estimation. Nucleic Acids Res. **33:**6114–6123.
23. **Lynd, L. R., et al.** 2008. How biotech can transform biofuels. Nat. Biotechnol. **26:**169–172.
24. **Lynd, L. R., W. H. van Zyl, J. E. McBride, and M. Laser.** 2005. Consolidated bioprocessing of cellulosic biomass: an update. Curr. Opin. Biotechnol. **16:** 577–583.
25. **Mai, V., J. Wiegel, and W. W. Lorenz.** 2000. Cloning, sequencing, and characterization of the bifunctional xylosidase-arabinosidase from the anaerobic thermophile *Thermoanaerobacter ethanolicus*. Gene **247:**137–143.
26. **Markowitz, V. M., et al.** 2006. The integrated microbial genomes (IMG) system. Nucleic Acids Res. **34:**D344–D348.
27. **Mielenz, J. R.** 2001. Ethanol production from biomass: technology and commercialization status. Curr. Opin. Microbiol. **4:**324–329.
28. **Ng, T. K., A. Ben-Bassat, and J. G. Zeikus.** 1981. Ethanol production by thermophilic bacteria: fermentation of cellulosic substrates by cocultures of *Clostridium thermocellum* and *Clostridium thermohydrosulfuricum*. Appl. Environ. Microbiol. **41:**1337–1343.
29. **Olofsson, K., M. Bertilsson, and G. Liden.** 2008. A short review on SSF—an interesting process option for ethanol production from lignocellulosic feedstocks. Biotechnol. Biofuels **1:**7.
30. **Onyenwoke, R. U., V. V. Kevbrin, A. M. Lysenko, and J. Wiegel.** 2007. *Thermoanaerobacter pseudethanolicus* sp. nov., a thermophilic heterotrophic anaerobe from Yellowstone National Park. Int. J. Syst. Evol. Microbiol. **57:**2191–2193.
31. **Palmqvist, E., and B. Hahn-Hagerdal.** 2000. Fermentation of lignocellulosic hydrolysates. II. Inhibitors and mechanisms of inhibition. Bioresour. Technol. **74:**25–33.
32. **Peng, H., G. Wu, and W. Shao.** 2008. The aldehyde/alcohol dehydrogenase (AdhE) in relation to the ethanol formation in *Thermoanaerobacter ethanolicus* JW200. Anaerobe **14:**125–127.
33. **Roh, Y., et al.** 2002. Isolation and characterization of metal-reducing *Thermoanaerobacter* strains from deep subsurface environments of the Piceance Basin, Colorado. Appl. Environ. Microbiol. **68:**6013–6020.
34. **Rosner, B.** 2005. Fundamentals of biostatistics, 6th ed. Duxbury Press, Boston, MA.
35. **Sato, K., et al.** 1992. Effect of yeast extract and vitamin B_{12} on ethanol production from cellulose by *Clostridium thermocellum* I-1-B. Appl. Environ. Microbiol. **58:**734–736.
36. **Sauer, U.** 2001. Evolutionary engineering of industrially important microbial phenotypes. Adv. Biochem. Eng. Biotechnol. **73:**129–169.
37. **Seedorf, H., et al.** 2008. The genome of *Clostridium kluyveri*, a strict anaerobe with unique metabolic features. Proc. Natl. Acad. Sci. U. S. A. **105:**2128–2133.
38. **Seo, J.-S., et al.** 2005. The genome sequence of the ethanologenic bacterium *Zymomonas mobilis* ZM4. Nat. Biotechnol. **23:**63–68.
39. **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.
40. **Speelmans, G., B. Poolman, T. Abee, and W. N. Konings.** 1994. The F- or V-type Na(+)-ATPase of the thermophilic bacterium *Clostridium fervidus*. J. Bacteriol. **176:**5160–5162.
41. **Stephanopoulos, G.** 2007. Challenges in engineering microbes for biofuels production. Science **315:**801–804.
42. **Tang, Y. J., J. S. Hwang, D. Wemmer, and J. D. Keasling.** 2007. The *Shewanella oneidensis* MR-1 fluxome under various oxygen conditions. Appl. Environ. Microbiol. **73:**718–729.
43. **Tang, Y. J., et al.** 2009. Analysis of metabolic pathways and fluxes in a newly discovered thermophilic and ethanol-tolerant *Geobacillus* strain. Biotechnol. Bioeng. **102:**1377–1386.
44. **Thompson, D. K., et al.** 2002. Transcriptional and proteomic analysis of a ferric uptake regulator (Fur) mutant of *Shewanella oneidensis*: possible involvement of Fur in energy metabolism, transcriptional regulation, and oxidative stress. Appl. Environ. Microbiol. **68:**881–892.
45. **Updegraff, D. M.** 1969. Semimicro determination of cellulose in biological materials. Anal. Biochem. **32:**420–424.
46. **van Zyl, W., L. Lynd, R. den Haan, and J. McBride.** 2007. Consolidated bioprocessing for bioethanol production using *Saccharomyces cerevisiae*. Adv. Biochem. Eng. Biotechnol. **108:**205–235.
47. **Waack, S., et al.** 2006. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. BMC Bioinformatics **7:**142.
48. **Wahl, S. A., M. Dauner, and W. Wiechert.** 2004. New tools for mass isotopomer data evaluation in ^{13}C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. Biotechnol. Bioeng. **85:** 259–268.
49. **Wiegel, J.** 1982. Ethanol from cellulose. Experientia (Basel) **38:**151–156.
50. **Wiegel, J.** 1980. Formation of ethanol by bacteria. A pledge for the use of extreme thermophilic anaerobic bacteria in industrial ethanol fermentation processes. Cell. Mol. Life Sci. **36:**1434–1446.
51. **Wingren, A., M. Galbe, and G. Zacchi.** 2003. Techno-economic evaluation of producing ethanol from softwood: comparison of SSF and SHF and identification of bottlenecks. Biotechnol. Prog. **19:**1109–1117.
52. **Zhang, Y.-H. P., and L. R. Lynd.** 2005. Cellulose utilization by *Clostridium thermocellum*: bioenergetics and hydrolysis product assimilation. Proc. Natl. Acad. Sci. U. S. A. **102:**7321–7325.