AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

PHYSIOLOGY

Investigating the Central Metabolism of *Clostridium thermosuccinogenes*

Jeroen Girwar Koendjbiharie,[a] Kilian Wiersma,[b] Richard van Kranenburg[a,b]

[a]Corbion, Gorinchem, Netherlands
[b]Wageningen University, Laboratory of Microbiology, Wageningen, Netherlands

ABSTRACT *Clostridium thermosuccinogenes* is a thermophilic anaerobic bacterium able to convert various carbohydrates to succinate and acetate as main fermentation products. Genomes of the four publicly available strains have been sequenced, and the genome of the type strain has been closed. The annotated genomes were used to reconstruct the central metabolism, and enzyme assays were used to validate annotations and to determine cofactor specificity. The genes were identified for the pathways to all fermentation products, as well as for the Embden-Meyerhof-Parnas pathway and the pentose phosphate pathway. Notably, a candidate transaldolase was lacking, and transcriptomics during growth on glucose versus that on xylose did not provide any leads to potential transaldolase genes or alternative pathways connecting the C_5 with the C_3/C_6 metabolism. Enzyme assays showed xylulokinase to prefer GTP over ATP, which could be of importance for engineering xylose utilization in related thermophilic species of industrial relevance. Furthermore, the gene responsible for malate dehydrogenase was identified via heterologous expression in *Escherichia coli* and subsequent assays with the cell extract, which has proven to be a simple and powerful method for the basal characterization of thermophilic enzymes.

IMPORTANCE Running industrial fermentation processes at elevated temperatures has several advantages, including reduced cooling requirements, increased reaction rates and solubilities, and a possibility to perform simultaneous saccharification and fermentation of a pretreated biomass. Most studies with thermophiles so far have focused on bioethanol production. *Clostridium thermosuccinogenes* seems an attractive production organism for organic acids, succinic acid in particular, from lignocellulosic biomass-derived sugars. This study provides valuable insights into its central metabolism and GTP and PP_i cofactor utilization.

KEYWORDS *Clostridium thermosuccinogenes*, succinate, xylulokinase, malate dehydrogenase

The conversion of lignocellulosic biomass, which is typically a complex mixture of sugar polymers, into useful green chemicals, such as building blocks for polymers, is seen as an important step in the transition to a biobased economy. The United States Department of Energy has published several well-known papers dealing with the top 12 promising biobased building blocks (1, 2). Several organic acids and dicarboxylic acids, including succinic acid (SA), have been identified as such building block chemicals that can be produced from biomass. SA is currently used mainly as a food ingredient or as an additive and precursor for pharmaceuticals. However, it has a range of potential large-scale industrial applications, which are ultimately dependent on their ability to economically compete with petroleum-based alternatives. Current commercial processes exclusively use mesophilic organisms at temperatures around 30 to 37°C. The use of thermophiles for an industrial fermentation process running at 50 to 60°C

Received 21 February 2018 **Accepted** 15 April 2018

Accepted manuscript posted online 20 April 2018

Citation Koendjbiharie JG, Wiersma K, van Kranenburg R. 2018. Investigating the central metabolism of *Clostridium thermosuccinogenes*. Appl Environ Microbiol 84:e00363-18. https://doi.org/10.1128/AEM.00363-18.

Editor Haruyuki Atomi, Kyoto University

Address correspondence to Richard van Kranenburg, r.van.kranenburg@corbion.com.

TABLE 1 General features of genome sequences of *C. thermosuccinogenes*[a]

Attribute	DSM 5806 (Illumina)	DSM 5807^T (hybrid)	DSM 5808 (Illumina)	DSM 5809 (hybrid)
Genome size (bp)	4,519,012	4,731,216	4,509,994	4,665,658
GC content (%)	41.5	41.5	41.5	41.5
DNA scaffolds	266	1	240	8
No. of:				
Total genes	3,930	4,086	3,888	4,035
Protein-encoding genes	3,783	3,940	3,751	3,885
RNA genes	70	75	69	75
Pseudogenes	77	71	68	75
Hypothetical[b]	1,724 (43.9)	1,783 (45.3)	1,699 (43.7)	1,808 (44.8)

[a]It should be noted that DSM 5806 and DSM 5808 were assembled with Illumina data only, whereas the numbers for DSM 5807^T and DSM 5809 represent the hybrid assembly with Illumina and PacBio data. Therefore, the numbers of these different assembly methods cannot directly be compared to each other.
[b]The numbers in parentheses represent the percentage of hypothetical genes relative to total genes.

has several advantages. Such processes (i) require less cooling, (ii) generally have higher reaction rates and solubilities, and (iii) have the possibility to carry out simultaneous saccharification and fermentation of a pretreated biomass, making a more economic process (3, 4).

Clostridium thermosuccinogenes is the only known thermophile that naturally produces SA as one of its main fermentation products. It was first isolated in 1990 for its ability to grow on inulin at elevated temperatures, but like many *Clostridia*, it grows well on a range of C_5 and C_6 sugars (5). Although it belongs to a group of (hemi)cellulolytic organisms and is a close relative of the industrially relevant thermophilic *Clostridium thermocellum* and mesophilic *Clostridium cellulolyticum*, it is incapable of degrading (hemi)cellulose. A physiological characterization of the only four strains of the species that have been described (DSM 5806, DSM 5807^T, DSM 5808, and DSM 5809) includes an evaluation of the effect of pH and redox potential on the distribution of its fermentation products, namely, SA, acetic acid, formic acid, lactic acid, ethanol, and hydrogen (6–8). Furthermore, enzyme assays have been used to shed light on the fermentation pathways toward these products. A pathway from phosphoenolpyruvate (PEP) to succinate was proposed via (i) PEP carboxylase (PEPC), (ii) malate dehydrogenase (MDH), (iii) fumarate hydratase (FH), and (iv) fumarate reductase (FR). For close relatives of *C. thermosuccinogenes* and most natural SA producers, it is now well established that the conversion from PEP to oxaloacetate is done by PEP carboxykinase (PEPCK) rather than PEPC (9–11). A genome sequence should clarify whether there is a PEPC or PEPCK present. Likewise, ATP-linked phosphofructokinase (PFK) activity was detected, whereas for *C. thermocellum*, only PP_i-linked activity was detected (7, 11).

In this paper, we present an updated characterization of *C. thermosuccinogenes* metabolism by means of genomics, transcriptomics, and enzyme assays and provide evidence of a GTP-dependent xylulokinase (XK).

RESULTS

Genome sequences of *C. thermosuccinogenes*. Genomes of the four *C. thermosuccinogenes* strains were sequenced using Illumina HiSeq technology. From the initial assemblies, it was clear that DSM 5806, DSM 5807^T, and DSM 5808 were quite similar, while DSM 5809 was more different from the others (see below). In an attempt to reduce the number of contigs, DSM 5807^T was selected for PacBio sequencing as the type strain and representative of DSM 5806 and DSM 5808, together with DSM 5809, which appears to produce the most succinic acid (5, 6). The data from both Illumina and PacBio were combined in a hybrid assembly, which resulted in a closed genome for DSM 5807^T and an assembly of 8 scaffolds for DSM 5809. The general features of the four genomes are presented in Table 1.

As indicated, the four strains are very similar. Using the genome-to-genome distance calculator (GGDC) to estimate the DNA-DNA hybridization (DDH) (12), DSM 5806, DSM 5807^T, and DSM 5808 were found to have 100% hybridization with each other and 92.3% with DSM 5809, which falls well within the 70% and 79% thresholds used to delineate species and subspecies, respectively (13).

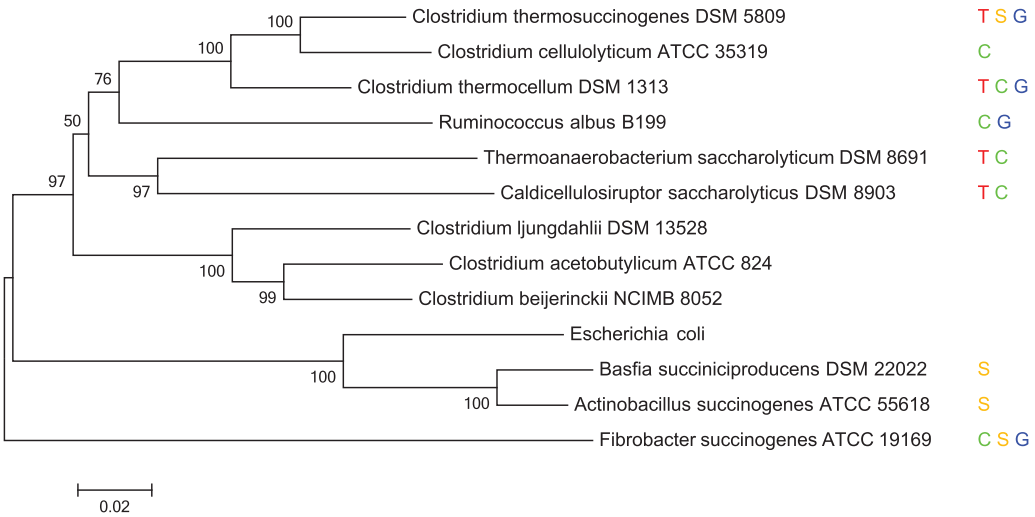

FIG 1 Neighbor-joining tree based on 16S rRNA sequences of a selection of industrially relevant clostridia and other species related to *C. thermosuccinogenes*, as well natural succinic acid producers and species with a known GTP-dependent glucokinase. T, thermophile; C, cellulose degrader; S, succinate producer; G, known to have a GTP-dependent glucokinase. The percentages next to branching points represent the results of bootstrapping with 1,000 replicates.

Orthologous and paralogous proteins in the four assembled genomes were defined using OrthoMCL to further quantify their similarity. Twenty-two orthologous groups (OGs) were found that had only one or two paralogs in DSM 5806 and DSM 5808, while these had up to 15 paralogs in DSM 5807^T and DSM 5809. These 22 OGs consisted almost exclusively of transposases and hypothetical proteins. The difference in the numbers of paralogs for the strains reflects the different assemblies used, as the short reads of Illumina do not enable a differentiation between repeated sequences without a larger scaffold sequence provided by PacBio sequence analysis. From the OrthoMCL analysis, it was also evident that DSM 5809 was more different from the other three strains; 369 OGs were uniquely absent from DSM 5809, whereas this was the case for only 1, 4, and 2 OGs for DSM 5806, DSM 5807^T, and DSM 5808, respectively. Conversely, 180, 154, 149, and 527 genes and/or OGs were unique for DSM 5806, DSM 5807^T, DSM 5808, and DSM 5809, respectively. A list of these unique genes together with the rest of the results from the OrthMCL analysis can be found in the supplemental material (see Table S1). The majority of the genes that are not shared between the four strains encode hypothetical proteins, and none encode apparent metabolic functions.

Many clostridia are being studied for their potential in biotechnology, including several close relatives of *C. thermosuccinogenes*: *C. thermocellum* and *C. cellulolyticum*. A 16S rRNA-based phylogenetic tree was constructed to place *C. thermosuccinogenes* in context with those biotechnologically relevant species, as well as other natural succinic acid producers and species with a known GTP-dependent glucokinase (GK) (Fig. 1).

Central metabolism. The central metabolism of *C. thermosuccinogenes* was reconstructed on the basis of the genome annotation of DSM 5809, which should be representative of the other three strains according to the OrthoMCL analysis (see above), although in some cases, they can have more or fewer isozymes for a certain reaction. The reconstruction was made by combining different annotations (NCBI pipeline, RAST, Prokka), which appeared to be significantly different in certain cases. Previous results from Sridhar et al. were also considered in the reconstruction (7). All genes for the Embden-Meyerhof-Parnas pathway and, apart from the transaldolase (TAL) gene, all genes for the pentose phosphate pathway (PPP) were annotated. Around the PEP-pyruvate-oxaloacetate node (9), four different pathways for the conversion from PEP to pyruvate seem to be present: pyruvate kinase (PYK); pyruvate, phosphate dikinase (PPdK); the "malate shunt;" and one via oxaloacetate decarboxylase (OAD). The latter two involve a GTP-dependent PEPCK rather than PEPC. As with most

anaerobes, the tricarboxylic acid (TCA) cycle is bifurcated, because no succinyl-coenzyme A (CoA) synthetase or succinyl-CoA ligase is present, which coincides with succinate being one of the end products. The routes to the other products (lactate, acetate, formate, ethanol, and hydrogen) are annotated as well. The formation of acetyl-CoA can go either via pyruvate:ferredoxin oxidoreductase or via pyruvate formate lyase. Sugar transport seems to occur mainly via ABC transporters, as no phosphotransferase systems were retrieved from the annotation, while many ABC transporters were found. The schematic reconstruction is shown in Fig. 2.

To help identify the missing TAL, a transcriptome sequencing analysis was performed on RNA isolated from cells growing on glucose versus on xylose. Table S2 shows the genes with the highest differential expression. As expected, genes responsible for the uptake of xylose and the subsequent conversion to xylulose-5-phosphate (X5P) showed the most differential expression. CDQ83_14310 and CDQ83_08695 are annotated by RAST as (xylose) transport proteins. CDQ83_14315 is the xylulokinase and CDQ83_14305 is almost certainly a xylose isomerase, rather than a fucose isomerase, and is also annotated as such by RAST. The rest of the differentially expressed genes are mostly hypothetical and are either very short peptides or still have a very low coverage. Since all carbon that enters the C_3/C_6 metabolism during growth on xylose has to be channeled through the nonoxidative PPP, it is not unlikely that the genes responsible for this conversion are upregulated as well and would include an unknown TAL or an alternative pathway. However, no leads for candidate genes were derived from the differentially expressed genes (Table S2).

Many reactions in the central metabolism have several isoforms annotated. The transcriptomics data can in some cases already indicate which of the isoforms are dominant during (exponential) growth on glucose and/or xylose. The transcriptomics data for all the annotated genes that potentially fulfil the roles of the reactions presented in Fig. 2 are listed in Table S3.

The annotation has no candidate MDH but has three candidate lactate dehydrogenases (LDHs), one of which (CDQ83_05915) is very ambiguously annotated. As it is difficult to differentiate between MDH and LDH on the basis of their sequences, it is likely that one of the genes annotated as LDH is really an MDH, which has not been annotated (14, 15). CDQ83_08825 has the highest homology with the isoform from *Clostridium thermocellum* that was found to be an MDH (16, 17) and seems therefore the most likely candidate. To test this, the two unambiguous isoforms were overexpressed in *Escherichia coli*, and the cell extract of *E. coli* was subsequently used to determine the MDH and LDH activities (Table 2). An SDS-PAGE analysis clearly showed the overproduction of a protein of the expected size in both strains (see Fig. S1). Background activities in the extract from *E. coli* with an empty vector control were negligible under the tested conditions. As expected, CDQ83_08825 is an MDH and shows activity with both NADH and NADPH. The other isoform, CDQ83_04860, is an LDH but also showed substantial NADH-dependent MDH activity.

Several genes are ambiguously annotated as potential oxaloacetate decarboxylases (OAD), including CDQ83_05940, which is part of a highly expressed operon containing three genes. Because none of those genes are predicted to be membrane proteins, which OAD is known to be (18, 19), and because they have orthologous genes in *C. thermocellum*, for which no OAD activity has currently been detected, it seems unlikely that they encode an OAD. Indeed, assays with a (crude) cell extract of *C. thermosuccinogenes* did not show any OAD activity; therefore, the OAD reaction is denoted with a dashed arrow in Fig. 2.

Cofactor specificities are difficult to derive from annotations. For *C. thermocellum*, it was demonstrated that it prefers GTP and PP_i for several glycolytic reactions instead of the "typical" ATP (11). Therefore, enzyme assays were carried out to evaluate cofactor specificities in *C. thermosuccinogenes*. As summarized in Table 3, GK activity depends on GTP and PFK activity on PP_i. Phosphoglycerate kinase activity is highest with ATP but is also detected with GTP. For XK, the highest activity was found with GTP, but ATP-dependent activity was also detected. The activity for malic enzyme (ME) was

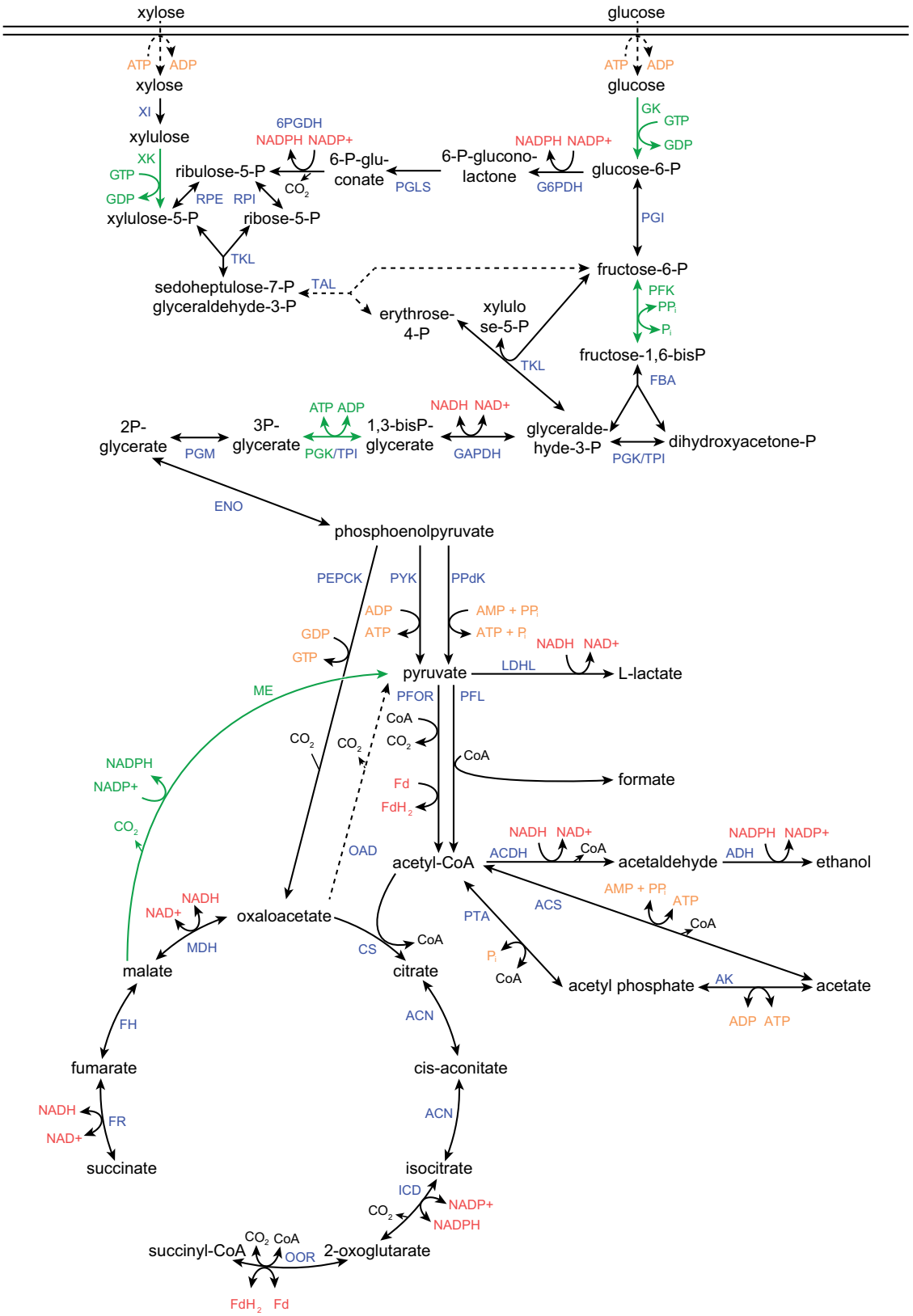

FIG 2 Reconstruction of the central metabolism of *C. thermosuccinogenes*. Green arrows denote reactions verified in enzyme assays with *C. thermosuccinogenes* cell extract. 6PGDH, 6-phosphogluconate dehydrogenase; ACDH, acetaldehyde dehydrogenase; ACN, aconitase; ACS, acetyl-CoA synthetase; ADH, alcohol dehydrogenase; AK, acetate kinase; CS, citrate synthase; ENO, enolase; FBA, fructose-bisphosphate aldolase; FH, fumarate hydratase; FR, fumarate reductase; G6PDH, glucose-6-phosphate dehydrogenase; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; GK, glucokinase; ICD, isocitrate dehydrogenase; LDHL, L-lactate dehydrogenase;

(Continued on next page)

TABLE 2 Enzyme assay with *E. coli* cell extract containing the two MDH candidates from *C. thermosuccinogenes*, which were initially annotated as LDHs

	Reaction velocity (μmol/mg cell extract protein/min)[a]			
	MDH[b] assay		LDH[c] assay	
Enzyme	NADH	NADPH	NADH	NADPH
CDQ83_08825	5.35 ± 0.04	6.27 ± 0.86	ND[d]	ND
CDQ83_04860	1.34 ± 0.42	ND	0.96 ± 0.13	ND

[a]Reaction velocities are given ± the standard deviations.
[b]MDH, malate dehydrogenase.
[c]LDH, lactate dehydrogenase.
[d]ND, not detected.

dependent on $NADP^+$. Except for XK, the difference in enzyme activities measured with extracts from cells grown on glucose or xylose appeared to be minimal.

DISCUSSION

Genome assembly. The hybrid assemblies of DSM 5807^T and DSM 5809 are approximately 200 kb and 150 to 200 genes larger than their SPAdes assemblies that are based on Illumina data. This difference presumably consists mostly of repeated sequences that cannot be assembled correctly based on the shorter reads from Illumina. This was also evident from the fact that a striking difference in the assemblies of DSM 5807^T and DSM 5809 compared to those of DSM 5806 and DSM 5808 was the much larger number of replicates of transposable elements.

With 4.7 Mb, the overall size of the genome is significantly larger than those of *C. thermocellum* and *C. cellulolyticum* of 3.6 Mb and 4.1 Mb, respectively.

Link between C_5 and C_3/C_6 metabolism. The absence of an annotated TAL has already been discussed for several of the cellulolytic clostridia (20–22), and according to Schellenberg et al. (22), 18% of clostridial genomes do not have a TAL annotated, of which at least a few, including *C. thermosuccinogenes*, can grow well on pentose sugars. From the transcriptomics data, no candidate genes were found, and none of the annotated genes from the nonoxidative PPP appeared to be differentially expressed during growth on xylose versus that on glucose, apart from the genes responsible for the formation of X5P (see Table S2 in the supplemental material). In many bacteria, including in *E. coli*, none of the genes of the nonoxidative pathway are differentially expressed during growth on xylose versus that on glucose (23). For *Clostridium termitidis*, an organism related to *C. thermosuccinogenes* that also lacks a transaldolase, the transketolase was found significantly upregulated during growth on xylose (21). However, we did not observe this in *C. thermosuccinogenes*. Assuming that this is indeed the case, the PPP enzymes are either already present in amounts high enough to channel all X5P to the C_3/C_6 metabolism or there is an alternative unknown pathway being expressed during growth on xylose. No indication for any such highly and differentially expressed pathway is apparent from our results. The presence of an alternative unknown pathway that is not differentially expressed during growth on xylose versus that on glucose would not appear in these results.

One alternative pathway proposed for *C. thermocellum* runs via sedoheptulose-1,7-bisphosphate formed from sedoheptulose-7-phosphate by a PFK that is subsequently cleaved to dihydroxyacetone-phosphate (DHAP) and erythrose-4-phosphate (E4P) by a fructose-bisphosphate aldolase, thereby replacing the missing TAL activity (20). This pathway has been demonstrated to exist in parasitic protists that also rely on PP_i-

FIG 2 Legend (Continued)
MDH, malate dehydrogenase; ME, malic enzyme; OAD, oxaloacetate; decarboxylase; OOR, 2-oxoglutarate,ferredoxin oxidoreductase; PEPCK, phosphoenolpyruvate carboxykinase; PFK, phosphofructokinase; PFL, pyruvate formate lyase; PFOR, pyruvate:ferredoxin oxidoreductase; PGI, phosphoglucose isomerase; PGK, phosphoglycerate kinase; PGLS, 6-phosphogluconolactonase; PPdK, pyruvate, phosphate dikinase; PTA, phosphate acetyltransferase; PYK, pyruvate kinase; RPE, ribulose-phosphate 3-epimerase; RPI, ribose-5-phosphate isomerase; TAL, transaldolase; TKL, transketolase; TPI, triosephosphate isomerase; XI, xylose isomerase; XK, xylulokinase.

TABLE 3 Enzyme assay with *C. thermosuccinogenes* cell extract to determine cofactor specificities for several glycolytic reactions

Reaction	Cofactor	Reaction velocity (μmol/mg cell extract protein/min)[a] in cell extracts grown on:	
		Glucose	Xylose
Glucokinase	ATP	ND[b]	ND
	GTP	0.90 ± 0.10	0.68 ± 0.12
	PP_i	ND	
Xylulokinase	ATP	0.01 ± 0.01	0.60 ± 0.01
	GTP	0.02 ± 0.00	0.83 ± 0.05
Malic enzyme	$NADP^+$	2.5 ± 0.3	2.3 ± 0.3
	NAD^+	ND	ND
6-Phosphofructokinase[c]	ATP	ND	
	GTP	ND	
	PP_i	2.8 ± 0.2	4.0 ± 0.2
Phosphoglycerate kinase[c]	ATP	2.8 ± 0.3	2.9 ± 0.2
	GTP	1.2 ± 0.1	1.1 ± 0.1
	PP_i		

[a]Reaction velocities are given ± the standard deviations.
[b]ND, not detected.
[c]Cell extract prepared with Tris-HCl buffer instead of potassium phosphate buffer.

dependent PFK (PPi-PFK) (24). As one of the hallmarks of PPi-PFK is its reversibility, in contrast to the ATP-dependent one (25), it is not unlikely that the PPi-PFK of *C. thermosuccinogenes* could also play a role in the nonoxidative PPP. PPi-PFK does appear to be roughly 40% more active in cell extract from cells grown on xylose, but this is insignificant as long as it has not been tested with sedoheptulose-7-phosphate. However, due to the limited availability of PPP intermediates that could be used for enzyme assays, it is not trivial to search for an unknown alternative pathway. Moreover, the recursive nature of the PPP and the high reverse fluxes, due to the low thermodynamic driving force of anaerobic metabolism in general, complicate stable-isotope-labeling studies, which would otherwise be a very powerful method.

PEP-pyruvate-oxaloacetate node. The PEP-pyruvate-oxaloacetate (PPO) node forms the junction between glycolysis and the TCA cycle and can compromise a large set of reactions (9). In *C. thermosuccinogenes*, genes are annotated for PEPCK, PPdK, PYK, OAD, MDH, and ME. Taken together, these reactions would enable four different pathways from PEP to pyruvate, each involving different cofactors: (i) PPdK: AMP, PPi → ATP + Pi; (ii) PYK: ADP → ATP; (iii) malate shunt: GDP, NADH → GTP, NADPH; and (iv) via PEPCK and OAD: GDP → GTP.

In contrast to what was found earlier by Sridhar et al. (7), *C. thermosuccinogenes* appears to rely on PEPCK rather than PEPC for the formation of oxaloacetate. One explanation for their conclusion could be the fact that they only tested activity with ADP, whereas it is a GTP-dependent PEPCK. Furthermore, it is probable that the activity they measured, using MDH as a reporter enzyme, was in fact that of PYK, via the nonspecific activity of MDH with pyruvate.

From the transcriptomics, it seems that during exponential growth on glucose and xylose, the malate shunt and OAD are expressed the highest (Table S3). However, the annotation of OAD is ambiguous, and activity could not be detected in enzyme assays. In *C. thermocellum*, no OAD activity has been detected either and is said to be absent (26), which is the most likely explanation. Furthermore, *C. thermocellum* also does not contain a PYK, and both the malate shunt and PPdK are shown to contribute substantially to pyruvate formation (26). Much effort has been put into engineering the PPO node in *C. thermocellum*, mostly to decrease the transhydrogenase effect of the malate shunt in order to increase the ethanol yield, which requires NADH (11, 16, 26, 27). For

C. thermosuccinogenes, the PPO node will likely also be an important target for improving succinate yield, since FH and FR directly compete with ME for malate.

CDQ83_08825, annotated as an LDH, was shown to be responsible for the formation of malate from oxaloacetate, as was already shown for its homolog in *C. thermocellum*, which also utilized both NADH and NADPH, albeit with a 12-fold lower catalytic activity for the latter (16). In *C. thermocellum*, the then putative MDH gene is located next to the ME gene, which made it the obvious MDH candidate to test. However, in *C. thermosuccinogenes*, the two genes are not adjacent to each other.

The other gene annotated as an LDH (CDQ83_04860) did indeed show LDH activity. However, it also appeared to exhibit significant NADH-dependent MDH activity. The relatively low LDH activity compared to MDH activity could be due to suboptimal assay conditions resulting, for example, from the *E. coli* cell extract or an inadequate fructose 1,6-bisphosphate concentration, which is typically required for LDH activity. Nevertheless, the heterologous expression of proteins in *E. coli* without further purification, save for an optional heating step, has proven to be a simple and powerful method for the basic characterization of enzymes with a thermophilic origin.

GTP-dependent xylulokinase. Our study provides evidence for an XK that prefers GTP over ATP. Although significant, the difference between the two nucleotides is not as big as for GK, where no ATP-dependent activity was detected at all. For cells grown on glucose, the XK activity is negligible compared to that from cells grown on xylose, which corresponds to the large difference in expression observed in the transcriptomics results. In only very few studies, GTP has been tested as a cofactor for XK, and if any GTP activity is observed, it is much lower than ATP-dependent activity (28–32). Nevertheless, GTP-dependent XKs could be widespread, similar to GTP-dependent GKs, as it is increasingly apparent that a "typical" glycolysis does not exist (33). Furthermore, it is possible that other sugar kinases in *C. thermosuccinogenes*, such as fructokinase and galactokinase, are also GTP dependent, as well as other kinases such as acetate kinase.

C. thermocellum is not able to grow on xylose, as it lacks the genes for xylose isomerase and XK. By expressing those missing genes from *Thermoanaerobacterium saccharolyticum*, *C. thermocellum* was previously engineered to grow on xylose (34), with the goal to increase its efficiency of ethanol production from hemicellulosic biomass. However, it is not known whether *T. saccharolyticum* XK uses GTP or ATP, which could be of importance. Besides being from a closer relative, the XK from *C. thermosuccinogenes* might also be an interesting candidate to test in *C. thermocellum* to potentially further improve its ability to grow on xylose.

There is no clear explanation for the reliance of the central energy metabolism on GTP. The fact that the GTP-dependent GK appears throughout a range of distantly related bacteria, e.g., *C. thermocellum* and *Fibrobacter succinogenes*, indicates that it is not simply an artifact of evolution but instead is driven by a certain underlying mechanism. Too few have been characterized to form a meaningful theory, but it could be related to cellulolytic (rumen) bacteria, since *F. succinogenes*, *Ruminococcus albus*, *C. thermocellum*, and *C. thermosuccinogenes*, the only organisms known to use a GTP-dependent GK (11, 35, 36), can all be associated with this group. The underlying mechanism can only be speculated upon. Perhaps GTP enables the existence of an additional energy charge, next to ATP, similar to NADH and NADPH having different oxidation states, enabling them to fulfil different roles in the metabolism. It could also be a primitive regulatory mechanism due to the direct link of GTP with anabolism, via protein synthesis.

Conclusions. We have sequenced and annotated the genomes of four strains of the thermophilic succinate producer *Clostridium thermosuccinogenes* and, with this, reconstructed its central metabolism. All enzymes for glycolysis and the fermentation pathways to its main products, including succinate, were identified, with the exception of the transaldolase in the PPP. A transcriptomics study for growth on glucose versus that on xylose did not hint at any transaldolase candidate genes or alternative pathways. Furthermore, we showed that *C. thermosuccinogenes*, similar to its close relative *C.*

thermocellum, uses GTP and PP_i for several of the glycolytic reactions. Xylulokinase, which is not present in *C. thermocellum*, was found to be GTP dependent as well and could therefore potentially aid in engineering xylose utilization in *C. thermocellum*. The malate dehydrogenase and the lactate dehydrogenase genes were identified via the heterologous expression of the two candidate genes in *E. coli*. Unpurified cell-free *E. coli* extracts were used for the assays, and this was found to be an efficient means of characterizing thermophilic enzymes.

MATERIALS AND METHODS

Anaerobic cultivation. *C. thermosuccinogenes* DSM 5806, DSM 5807^T, DSM 5808, and DSM 5809 were acquired from DSMZ (Braunschweig, Germany). The strains were routinely cultivated anaerobically in 120-ml serum bottles containing 50 ml bicarbonate-buffered liquid medium and an N_2/CO_2 (80%/20%) atmosphere. The cultures were incubated at 60°C.

Glycerol stocks for storage at −80°C were prepared by adding 2 ml of exponentially growing culture (optical density at 600 nm [OD_{600}] of 0.2 to 0.6) into previously prepared anaerobic vials containing 2 ml of 50% glycerol in phosphate-buffered saline (pH 7.3) and 0.5 mg/liter resazurin. The vials were reduced with a few drops of titanium citrate (100 mM) directly before the addition of the culture.

Overnight precultures were typically grown with 2 g/liter of substrate (i.e., glucose or xylose), as they were found to have a shorter lag phase upon transfer than that of overnight cultures grown in 5 g/liter substrate, presumably because there is less to no acidification.

Medium composition and preparation. *C. thermosuccinogenes* was grown in adapted CP medium (37), which contained per liter 0.408 g KH_2PO_4, 0.534 g $Na_2HPO_4{\cdot}2H_2O$, 0.3 g NH_4Cl, 0.3 g NaCl, 0.1 g $MgCl_2{\cdot}6H_2O$, 0.11 g $CaCl_2{\cdot}2H_2O$, 4.0 g $NaHCO_3$, 0.1 g Na_2SO_4, 1.0 g L-cysteine, 1.0 g yeast extract (BD Bacto), 0.5 mg resazurin, 1 ml vitamin solution, 1 ml trace elements solution I, and 1 ml trace elements solution II. The medium was autoclaved in serum bottles under 80:20 N_2/CO_2 atmosphere with ~70 kPa overpressure, containing a final volume of 50 ml medium. A solution containing $NaHCO_3$ and L-cysteine was autoclaved separately and added later as well as a solution containing $CaCl_2{\cdot}2H_2O$, to which the vitamin solution was added after it was autoclaved. The substrate (glucose or xylose) was also autoclaved separately and added later to a final concentration of 2 g/liter or 5 g/liter.

The vitamin solution, which was 1,000× concentrated, contained per liter 20 mg biotin, 20 mg folic acid, 100 mg pyridoxine-HCl, 50 mg thiamine-HCl, 50 mg riboflavin, 50 mg nicotinic acid, 50 mg Ca-D-pantothenate, 1 mg vitamin B_{12}, 50 mg 4-aminobenzoid acid, and 50 mg lipoic acid.

Trace elements solution I, which was 1,000× concentrated, contained per liter 50 mM HCl, 61.8 mg H_3BO_4, 99.0 mg $MnCl_2{\cdot}4H_2O$, 1.49 g $FeCl_2{\cdot}4H_2O$, 119 mg $CoCl_2{\cdot}6H_2O$, 23.8 mg $NiCl_2{\cdot}6H_2O$, 68.2 mg $ZnCl_2$, and 17.0 mg $CuCl_2{\cdot}2H_2O$.

Trace elements solution II, which was 1,000× concentrated, contained per liter 10 mM NaOH, 17.3 mg Na_2SeO_3, 33.0 mg $Na_2WO_4{\cdot}2H_2O$, and 24.2 mg $Na_2MoO_4{\cdot}2H_2O$.

Genome sequencing and annotation. Ten to twenty milliliters of exponentially growing cells was harvested for DNA extraction with the Gram positive DNA purification kit (Epicentre, Madison, Wisconsin) according to the manufacturer's instructions.

Library preparation and sequencing was carried out by BaseClear (Leiden, The Netherlands) both for Illumina and PacBio. Illumina sequencing was performed with the HiSeq2500 system, using paired-end chemistry and run lengths of 125 bp. FASTQ sequence files were generated using the Illumina Casava pipeline version 1.8.3. The initial quality assessment was based on data passing the Illumina Chastity filtering. Subsequently, the reads containing adapters and/or the PhiX control signal were removed using an in-house filtering protocol. The second quality assessment was based on the remaining reads using the FASTQC quality control tool version 0.10.0 (38). This resulted in 4,159,110, 5,143,445, 4,483,080, and 3,992,113 reads for DSM 5806, DSM 5807, DSM 5808, and DSM 5809, respectively. The data collected from the PacBio RS instrument were processed and filtered using the SMRT analysis software suite. The continuous long read data were filtered by read length (>35), subread length (>35), and read quality (>0.75). This resulted in 216,616 and 91,998 reads for DSM 5807 and DSM 5809, respectively, with average read lengths of 3,577 and 5,794, respectively, and maximum read lengths of 41,645 and 36,236, respectively.

Illumina sequence data from the four strains were assembled using SPAdes genome assembler (39). For DSM 5807 and DSM 5809, the additional sequence data from PacBio sequencing were used in combination with the Illumina data for a hybrid assembly. The hybrid assembly was carried out by BaseClear (Leiden, The Netherlands), using ABySS assembler version 1.5.1, SSPACE-LongRead scaffolder version 1.0 (40), and GapFiller version 1.10 (41).

The final assemblies of the four strains, i.e., SPAdes assemblies for DSM 5806 and DSM 580 and hybrid assemblies for DSM 5807 and DSM 5809, have been submitted to the NCBI database and were annotated by their in-house annotation pipeline (42). Additionally, the different assemblies have also been annotated with RAST and with Prokka, aiding the manual reconstruction of the central metabolism (43, 44).

***In silico* DNA-DNA hybridization.** The genome-to-genome distance calculator (GGDC) version 2.1 from the DSMZ website was used to calculate the DNA-DNA hybridization of the four *C. thermosuccinogenes* strains (12). For all four strains, the SPAdes assembly of Illumina HiSeq sequence data was used as the input. The results of formula 2 were used, as is recommended, but results from formula 1 and formula 3 led to identical conclusions.

OrthoMCL analysis. The OrthoMCL algorithm was used to identify orthologous proteins in the four *C. thermosuccinogenes* genomes, using the default settings (45). Protein FASTA files were used, derived from the genomes that were annotated by the NCBI pipeline, meaning that for DSM 5806 and DSM 5808, these were based the SPAdes assemblies, whereas these were based on the hybrid assembly for DSM 5807^T and DSM 5809. The output of the program was used to compile a list with unique genes, i.e., a list of genes that did not appear in any of the orthologous (and paralogous) groups, including all the RNA genes, as these are not part of the OrthoMCL analysis. Additionally, lists were made with orthologous groups that were present in all but one of the strains. The output of the OrthoMCL analysis can be found in Table S1 in the supplemental material.

Construction of phylogenetic tree. For the construction of the phylogenetic tree, 16S RNA sequences were aligned using ClustalW with the standard settings in MEGA6 (46). Aligned sequences were trimmed to a length of 1,324 bp and were used for the construction of a phylogenetic tree by using the neighbor-joining methods in MEGA6. A bootstrap test with 1,000 replicates was performed.

Transcriptomics. Ten milliliters of exponentially growing cells at an OD_{600} of ~0.3 was harvested for RNA extraction by centrifugation at 4,800 × *g* for 15 min at 4°C. The pellet was suspended in 0.5 ml ice-cold Tris-EDTA (TE) buffer (pH 8.0). The samples were divided into two 2-ml screw-cap tubes containing 0.5 g zirconium beads, 30 μl 10% SDS, 30 μl 3 M sodium acetate (pH 5.2), and 500 μl water-saturated phenol, chloroform, and isoamyl alcohol at a ratio of 25:24:1 (pH 4.5 to 5) (Roti-Aqua-P/C/I; Carl Roth, Karlsruhe, Germany). Cells were disrupted in a Precellys 24 tissue homogenizer (Bertin Instruments, Montigny-le-Bretonneux, France) at speed 6 for 40 s. After a centrifugation for 5 min at 9,300 × *g* at 4°C, the water phase (top) was transferred to a new tube containing 400 μl chloroform. After a centrifugation for 3 min at maximum speed, the water phase was again transferred to a new tube and mixed with the lysis buffer from the High Pure RNA isolation kit from Roche. From there on, the protocol of the kit was followed. The integrity of the RNA was checked using Experion RNA StdSens chips from Bio-Rad.

Library preparation and sequencing was performed by BaseClear (Leiden, The Netherlands), using an Illumina HiSeq platform, yielding 50-bp single reads. The samples from cells growing on glucose and on xylose resulted in 25 and 67 million reads, respectively. Transcript reads were aligned on the coding sequences from the NCBI annotation using the Burrows-Wheeler aligner (BWA; version 0.7.12-r1039). From this, the reads per kilobase per million mapped reads (RPKM) were calculated for every coding sequence.

Preparation of cell extract of *C. thermosuccinogenes*. Cell extracts were made using 150 ml (3 × 50 ml) of exponentially growing cultures at an OD_{600} of 0.3 to 0.4, according to a protocol adapted from Zhou et al. (11). The cells were harvested by centrifugation at 4,800 × *g* for 10 min at 4°C and were washed twice with 50 mM Tris-HCl buffer (pH 8.0) containing 5 mM freshly added dithioerythritol (DTT). The cells were finally suspended in 5 ml of either the wash buffer or 50 mM potassium phosphate buffer (pH 7.0) containing 5 mM DTT. The cell suspensions were homogenized in a French press at ~120 kPa. The lysate was centrifuged for 10 min at maximum speed in microcentrifuge tubes, and the supernatants were used as the cell extracts. For the oxaloacetate decarboxylase assay, the crude extract that was not centrifuged was also used to test if any activity was present in the solid fraction. The total protein concentration in the cell extracts was determined using the Bradford assay, with bovine serum albumin as a standard. Protein concentrations in all cases were >1 mg/ml. All steps were performed aerobically.

Heterologous expression of malate dehydrogenase candidates in *E. coli*. The two malate dehydrogenase gene candidates from *C. thermosuccinogenes* investigated in this study were heterologously overexpressed in *E. coli*. CDQ83_08825 was amplified via PCR using the forward primer TACTTC CAATCCAATGCAGTAAAATCCAAATCAAAAGTTGCAATAATC and the reverse primer TTATCCACTTCCAATG TTATATATCCTTCACCTGATCGATTATAGCCTTTAC. CDQ83_04860 was amplified via PCR using the forward primer TACTTCCAATCCAATGCAATGCATGAAATTACACCAAAAAAGATC and the reverse primer TTATCCA CTTCCAATGCTACAATTTAAGTTTGCCGGC. Via ligase-independent cloning, the cloned enzymes were inserted in a backbone derived from pET-28b(+) (Novagen, Madison, Wisconsin) that was generated using the primers ATTGGATTGGAAGTACAGGTTTTCATGGTGATGGTGATGGTGAGAAGAACCCATGGTATA TCTCCTTCTTAAAG and ATTGGAAGTGGATAACGGATCCGAATTCGAGCGCCGTCGACAAGCTTGCGG. The ligase-independent cloning protocol is described elsewhere and takes advantage of the exonuclease activity of T4 polymerase to create compatible ends (47). In the final construct, the enzyme has an N-terminal His tag with a TEV protease site in between. Constructs containing CDQ83_08825 and CDQ83_04860 were transformed to *E. coli* DH5α and subsequently verified via sequencing.

CDQ83_08825 and CDQ83_04860 were heterologously overexpressed in *E. coli* Rosetta (Novagen), which is an *E. coli* BL21 derivative containing a plasmid encoding tRNAs of rare codons and a chloramphenicol resistance marker. Cells carrying the expression plasmids were grown in LB containing 50 μg/ml kanamycin and 20 μg/ml chloramphenicol to an OD_{600} of 0.6 to 0.8 and subsequently placed on ice for 20 min. Fifty milliliters of the culture was used for making *E. coli* cell extracts. Heterologous gene expression was then induced by the addition of 0.2 mM IPTG (isopropyl-β-D-thiogalactopyranoside), and the cultures were grown for an additional 3 to 4 h at 37°C. Then, the cell extract of *E. coli* was prepared in a manner identical to that for *C. thermosuccinogenes* cell extract, except that no DTT was used in the buffers. Additionally, the *E. coli* cell extract was also subjected to a heating step of 30 min at 60°C, followed by a centrifugation step to remove precipitated proteins. This heating step was included to decrease the background activity of the *E. coli* extract, although the assays were found to work very well even without this last step.

Enzyme assays. Activities in all assays were determined either indirectly (via one or more auxiliary enzymes) or directly by measuring the change in absorbance at 340 nm, which corresponds to $NAD(P)^+$ reduction or NAD(P)H oxidation. A Shimadzu U-2010 spectrophotometer in combination with a ther-

moelectric cell holder was used for the measurements that were performed at 55°C. Crystal cuvettes containing 1 ml of the reaction mixture were used with a 1.0-cm path length. Activities are expressed in micromoles of product per minute per mg of cell extract protein. The enzymes and biochemicals were obtained from Sigma. Tris-HCl buffer used was set at pH 8.0 at room temperature, which corresponds to ~pH 7.0 at 55°C. All enzyme assays contained 50 mM Tris-HCl. Ten to sixty microliters of cell extract was used in the assays, and in all cases, at least three different concentrations were tested to verify that the extract was the limiting factor in the assay. Water was added to a final volume of 1 ml. In the case of glucokinase and phosphofructokinase, significant background activity was present without the addition of a phosphoryl donor, which was subtracted from the final values.

The glucokinase (EC 2.7.1.2) assay was adapted from Zhou et al. (11) and contained 5 mM $MgCl_2$, 60 mM KCl, 2 mM glucose, 0.15 mM $NADP^+$, and 2 U/ml glucose-6-phosphate dehydrogenase; 2 mM ATP, GTP, or PP_i was added to start the reaction.

The xylulokinase (EC 2.7.1.17) assay was adapted from Dills et al. (29) and contained 5 mM $MgCl_2$, 2 mM xylulose, 2 mM phosphoenolpyruvate, 0.15 mM NADH, 4 U/ml pyruvate kinase, and 4 U/ml lactate dehydrogenase; 2 mM ATP, GTP, or PP_i was added to start the reaction.

The phosphofructokinase (EC 2.7.1.11/EC 2.7.1.90) assay was adapted from Zhou et al. (11) and contained 5 mM $MgCl_2$, 1 mM fructose-6-phosphate, 0.15 mM NADH, 4 U/ml aldolase, 4 U/ml triose-phosphate isomerase, and 4 U/ml α-glycerophosphate dehydrogenase; 2 mM ATP, GTP, or PP_i was added to start the reaction.

The phosphoglycerate kinase (EC 2.7.2.3) assay was adapted from Zhou et al. (11) and contained 5 mM $MgCl_2$, 2 mM EDTA, 2 mM 3-phosphoglycerate, 0.15 mM NADH, and 2 U/ml glyceraldehyde-3-phosphate dehydrogenase; 2 mM ATP, GTP, or PP_i was added to start the reaction.

The malic enzyme (EC 1.1.1.40) assay was adapted from Zhou et al. (11) and contained 5 mM $MgCl_2$, 5 mM NH_4Cl, 5 mM dithiothreitol, and 0.15 mM $NADP^+$; 2 mM malic acid was added to start the reaction.

The malate dehydrogenase (EC 1.1.1.37) assay was adapted from Taillefer et al. (16) and contained 0.3 mM NADH; 10 mM oxaloacetate was added to start the reaction.

The lactate dehydrogenase (EC 1.1.1.27) assay was adapted from Taillefer et al. (16) and contained 0.005 mM fructose 1,6-bisphosphate and 0.15 mM NADH; 10 mM pyruvate was added to start the reaction.

The oxaloacetate decarboxylase (EC 4.1.1.3) assay was adapted from Olson et al. (26) and contained 0 to 2 mM $MgCl_2$, 0 to 2 mM NaCl, and 1 to 1.6 mM oxaloacetic acid. The (crude) cell extract was added to start the reaction after the rate of spontaneous oxaloacetic acid degradation was determined. The reaction was monitored at 265 nm, which is the absorbance peak for oxaloacetic acid.

Accession number(s). The NCBI database accession numbers for DSM 5806, DSM 5807, DSM 5808, and DSM 5809 are NIOJ00000000, CP021850, NIOK00000000, and NIOI00000000, respectively.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.00363-18.

SUPPLEMENTAL FILE 1, XLSX file, 0.5 MB.
SUPPLEMENTAL FILE 2, PDF file, 0.1 MB.
SUPPLEMENTAL FILE 3, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

This research was funded by the European Union Marie Skłodowska-Curie Innovative Training Networks (ITN), contract number 642068.

We thank Bastienne Vriesendorp for her help with the bioinformatics analyses.

REFERENCES

1. Werpy T, Petersen G, Aden A, Bozell J, Holladay J, White J, Manheim A, Eliot D, Lasure L, Jones S. 2004. Top value added chemicals from biomass: volume 1–results of screening for potential candidates from sugars and synthesis, gas. U.S. Department of Energy, Washington, DC.
2. Bozell JJ, Petersen GR. 2010. Technology development for the production of biobased products from biorefinery carbohydrates—the US Department of Energy's "top 10" revisited. Green Chem 12:539–554. https://doi.org/10.1039/b922014c.
3. Kádár Z, Szengyel Z, Réczey K. 2004. Simultaneous saccharification and fermentation (SSF) of industrial wastes for the production of ethanol. Ind Crops Prod 20:103–110. https://doi.org/10.1016/j.indcrop.2003.12.015.
4. Abdel-Banat BMA, Hoshida H, Ano A, Nonklang S, Akada R. 2010. High-temperature fermentation: how can processes for ethanol production at high temperatures become superior to the traditional process using mesophilic yeast? Appl Microbiol Biotechnol 85:861–867. https://doi.org/10.1007/s00253-009-2248-5.
5. Drent WJ, Lahpor GA, Wiegant WM, Gottschal JC. 1991. Fermentation of inulin by *Clostridium thermosuccinogenes* sp. nov., a thermophilic anaerobic bacterium isolated from various habitats. Appl Environ Microbiol 57:455–462.
6. Sridhar J, Eiteman MA. 1999. Influence of redox potential on product distribution in *Clostridium thermosuccinogenes*. Appl Biochem Biotechnol 82:91–101. https://doi.org/10.1385/ABAB:82:2:91.
7. Sridhar J, Eiteman MA, Wiegel JW. 2000. Elucidation of enzymes in fermentation pathways used by *Clostridium thermosuccinogenes* growing on inulin. Appl Environ Microbiol 66:246–251. https://doi.org/10.1128/AEM.66.1.246-251.2000.
8. Sridhar J, Eiteman MA. 2001. Metabolic flux analysis of *Clostridium thermosuccinogenes*: effects of pH and culture redox potential. Appl Biochem Biotechnol 94:51–69. https://doi.org/10.1385/ABAB:94:1:51.
9. Sauer U, Eikmanns BJ. 2005. The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. FEMS Microbiol Rev 29:765–794. https://doi.org/10.1016/j.femsre.2004.11.002.
10. Lee SJ, Song H, Lee SY. 2006. Genome-based metabolic engineering

of *Mannheimia succiniciproducens* for succinic acid production. Appl Environ Microbiol 72:1939–1948. https://doi.org/10.1128/AEM.72.3.1939-1948.2006.
11. Zhou J, Olson DG, Argyros DA, Deng Y, van Gulik WM, van Dijken JP, Lynd LR. 2013. Atypical glycolysis in *Clostridium thermocellum*. Appl Environ Microbiol 79:3000–3008. https://doi.org/10.1128/AEM.04037-12.
12. Auch AF, von Jan M, Klenk H-P, Göker M. 2010. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci 2:117–134. https://doi.org/10.4056/sigs.531120.
13. Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, Rohde C, Rohde M, Fartmann B, Goodwin LA, Chertkov O, Reddy T, Pati A, Ivanova NN, Markowitz V, Kyrpides NC, Woyke T, Göker M, Klenk H-P. 2014. Complete genome sequence of DSM 30083T, the type strain (U5/41T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. Stand Genomic Sci 9:2. https://doi.org/10.1186/1944-3277-9-2.
14. Wilks HM, Hart KW, Feeney R, Dunn CR, Muirhead H, Chia WN, Barstow DA, Atkinson T, Clarke AR, Holbrook JJ. 1988. A specific, highly active malate dehydrogenase by redesign of a lactate dehydrogenase framework. Science 242:1541–1545. https://doi.org/10.1126/science.3201242.
15. Yin Y, Kirsch JF. 2007. Identification of functional paralog shift mutations: conversion of *Escherichia coli* malate dehydrogenase to a lactate dehydrogenase. Proc Natl Acad Sci U S A 104:17353–17357. https://doi.org/10.1073/pnas.0708265104.
16. Taillefer M, Rydzak T, Levin DB, Oresnik IJ, Sparling R. 2015. Reassessment of the transhydrogenase/malate shunt pathway in *Clostridium thermocellum* ATCC 27405 through kinetic characterization of malic enzyme and malate dehydrogenase. Appl Environ Microbiol 81:2423–2432. https://doi.org/10.1128/AEM.03360-14.
17. Özkan M, Yllmaz EI, Lynd LR, Özcengiz G. 2004. Cloning and expression of the *Clostridium thermocellum* L-lactate dehydrogenase gene in *Escherichia coli* and enzyme characterization. Can J Microbiol 50:845–851. https://doi.org/10.1139/w04-071.
18. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580. https://doi.org/10.1006/jmbi.2000.4315.
19. Dimroth P, Jockel P, Schmid M. 2001. Coupling mechanism of the oxaloacetate decarboxylase Na(+) pump. Biochim Biophys Acta 1505:1–14. https://doi.org/10.1016/S0005-2728(00)00272-3.
20. Rydzak T, McQueen PD, Krokhin OV, Spicer V, Ezzati P, Dwivedi RC, Shamshurin D, Levin DB, Wilkins JA, Sparling R. 2012. Proteomic analysis of *Clostridium thermocellum* core metabolism: relative protein expression profiles and growth phase-dependent changes in protein expression. BMC Microbiol 12:214. https://doi.org/10.1186/1471-2180-12-214.
21. Munir RI, Spicer V, Krokhin OV, Shamshurin D, Zhang X, Taillefer M, Blunt W, Cicek N, Sparling R, Levin DB. 2016. Transcriptomic and proteomic analyses of core metabolism in *Clostridium termitidis* CT1112 during growth on α-cellulose, xylan, cellobiose and xylose. BMC Microbiol 16:91. https://doi.org/10.1186/s12866-016-0711-x.
22. Schellenberg JJ, Verbeke TJ, McQueen P, Krokhin OV, Zhang X, Alvare G, Fristensky B, Thallinger GG, Henrissat B, Wilkins JA, Levin DB, Sparling R. 2014. Enhanced whole genome sequence and annotation of *Clostridium stercorarium* DSM8532T using RNA-seq transcriptomics and high-throughput proteomics. BMC Genomics 15:567. https://doi.org/10.1186/1471-2164-15-567.
23. Gonzalez R, Tao H, Shanmugam KT, York SW, Ingram LO. 2002. Global gene expression differences associated with changes in glycolytic flux and growth rate in *Escherichia coli* during the fermentation of glucose and xylose. Biotechnol Prog 18:6–20.
24. Susskind BM, Warren LG, Reeves RE. 1982. A pathway for the interconversion of hexose and pentose in the parasitic amoeba *Entamoeba histolytica*. Biochem J 204:191–199. https://doi.org/10.1042/bj2040191.
25. Mertens E. 1991. Pyrophosphate-dependent phosphofructokinase, an anaerobic glycolytic enzyme? FEBS Lett 285:1–5. https://doi.org/10.1016/0014-5793(91)80711-B.
26. Olson DG, Hörl M, Fuhrer T, Cui J, Zhou J, Maloney MI, Amador-Noguez D, Tian L, Sauer U, Lynd LR. 2017. Glycolysis without pyruvate kinase in *Clostridium thermocellum*. Metab Eng 39:169–180. https://doi.org/10.1016/j.ymben.2016.11.011.
27. Deng Y, Olson DG, Zhou J, Herring CD, Joe Shaw A, Lynd LR. 2013. Redirecting carbon flux through exogenous pyruvate kinase to achieve high ethanol yields in *Clostridium thermocellum*. Metab Eng 15:151–158. https://doi.org/10.1016/j.ymben.2012.11.006.
28. Bhuyan BK, Simpson FJ. 1962. Some properties of the D-ylulokinase of *Aerobacter aerogenes*. Can J Microbiol 8:737–745. https://doi.org/10.1139/m62-095.
29. Dills WL, Parsons PD, Westgate CL, Komplin NJA. 1994. Assay, purification, and properties of bovine liver D-xylulokinase. Protein Expr Purif 5:259–265. https://doi.org/10.1006/prep.1994.1039.
30. Matheron C, Delort A-M, Gaudet G, Forano E. 1997. Re-investigation of glucose metabolism in *Fibrobacter succinogenes*, using NMR spectroscopy and enzymatic assays. Evidence for pentose phosphates phosphoketolase and pyruvate formate lyase activities. Biochim Biophys Acta 1355:50–60. https://doi.org/10.1016/S0167-4889(96)00118-8.
31. Damen B, Pollet A, Dornez E, Broekaert WF, Van Haesendonck I, Trogh I, Arnaut F, Delcour JA, Courtin CM. 2012. Xylanase-mediated *in situ* production of arabinoxylan oligosaccharides with prebiotic potential in whole meal breads and breads enriched with arabinoxylan rich materials. Food Chem 131:111–118. https://doi.org/10.1016/j.foodchem.2011.08.043.
32. Martín del Campo JS, Chun Y, Kim J-E, Patiño R, Zhang Y-HP. 2013. Discovery and characterization of a novel ATP/polyphosphate xylulokinase from a hyperthermophilic bacterium *Thermotoga maritima*. J Ind Microbiol Biotechnol 40:661–669. https://doi.org/10.1007/s10295-013-1265-7.
33. Taillefer M, Sparling R. 2016. Glycolysis as the central core of fermentation. Adv Biochem Eng Biotechnol 156:55–78. https://doi.org/10.1007/10_2015_5003.
34. Argyos A, Barrett T, Caiazza N, Hogsett D. June 2012. Genetically modified *Clostridium thermocellum* engineered to ferment xylose. World Intellectual Property Organization patent WO2012088467 (A2).
35. Glass TL, Sherwood JS. 1994. Phosphorylation of glucose by a guanosine-5′-triphosphate (GTP)-dependent glucokinase in *Fibrobacter succinogenes* subsp. *succinogenes* S85. Arch Microbiol 162:180–186. https://doi.org/10.1007/BF00314472.
36. Lou J, Dawson KA, Strobel HJ. 1997. Cellobiose and cellodextrin metabolism by the ruminal bacterium *Ruminococcus albus*. Curr Microbiol 35:221–227. https://doi.org/10.1007/s002849900242.
37. Plugge CM. 2005. Anoxic media design, preparation, and considerations. Methods Enzymol 397:3–16. https://doi.org/10.1016/S0076-6879(05)97001-8.
38. Andrews S. 2010. FastQC: a quality control tool for high-throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
39. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
40. Boetzer M, Pirovano W. 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinformatics 15:211. https://doi.org/10.1186/1471-2105-15-211.
41. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. Genome Biol 13:R56. https://doi.org/10.1186/gb-2012-13-6-r56.
42. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
43. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.
44. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
45. Li L. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178–2189. https://doi.org/10.1101/gr.1224503.
46. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725–2729. https://doi.org/10.1093/molbev/mst197.
47. Aslanidis C, de Jong PJ. 1990. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res 18:6069–6074. https://doi.org/10.1093/nar/18.20.6069.