



# The Complete Genome Sequence of Hyperthermophile *Dictyoglomus turgidum* DSM 6724™ Reveals a Specialized Carbohydrate Fermentor

Phillip J. Brumm<sup>1,2\*</sup>, Krishne Gowda<sup>2,3</sup>, Frank T. Robb<sup>4</sup> and David A. Mead<sup>2,5</sup>

<sup>1</sup> C5-6 Technologies LLC, Fitchburg, WI, USA, <sup>2</sup> DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI, USA, <sup>3</sup> Lucigen Corporation, Middleton, WI, USA, <sup>4</sup> Department of Microbiology and Immunology, Institute of Marine and Environmental Technology, University of Maryland, Baltimore, MD, USA, <sup>5</sup> Varigen Biosciences Corporation, Madison, WI, USA

Here we report the complete genome sequence of the chemoorganotrophic, extremely thermophilic bacterium, *Dictyoglomus turgidum*, which is a Gram negative, strictly anaerobic bacterium. *D. turgidum* and *D. thermophilum* together form the *Dictyoglomi* phylum. The two *Dictyoglomus* genomes are highly syntenic, and both are distantly related to *Caldicellulosiruptor* spp. *D. turgidum* is able to grow on a wide variety of polysaccharide substrates due to significant genomic commitment to glycosyl hydrolases, 16 of which were cloned and expressed in our study. The GH5, GH10, and GH42 enzymes characterized in this study suggest that *D. turgidum* can utilize most plant-based polysaccharides except crystalline cellulose. The DNA polymerase I enzyme was also expressed and characterized. The pure enzyme showed improved amplification of long PCR targets compared to Taq polymerase. The genome contains a full complement of DNA modifying enzymes, and an unusually high copy number (4) of a new, ancestral family of polB type nucleotidyltransferases designated as MNT (minimal nucleotidyltransferases). Considering its optimal growth at 72°C, *D. turgidum* has an anomalously low G+C content of 39.9% that may account for the presence of reverse gyrase, usually associated with hyperthermophiles.

**Keywords:** *Dictyoglomus turgidum*, thermophile, biomass degradation, phage, *Dictyoglomi*, DNA polymerase, glucanase, reverse gyrase

## OPEN ACCESS

### Edited by:

Kian Mau Goh,  
Universiti Teknologi Malaysia, Malaysia

### Reviewed by:

Biswarup Mukhopadhyay,  
Virginia Tech, USA  
Ida Helene Steen,  
University of Bergen, Norway

### \*Correspondence:

Phillip J. Brumm  
pbrumm@c56technologies.com

### Specialty section:

This article was submitted to  
Extreme Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 28 July 2016

**Accepted:** 25 November 2016

**Published:** 20 December 2016

### Citation:

Brumm PJ, Gowda K, Robb FT and Mead DA (2016) The Complete Genome Sequence of Hyperthermophile *Dictyoglomus turgidum* DSM 6724™ Reveals a Specialized Carbohydrate Fermentor. *Front. Microbiol.* 7:1979. doi: 10.3389/fmicb.2016.01979

## INTRODUCTION

*Dictyoglomus* species are genetically distinct and divergent from known taxa, and have been assigned to their own phylum, *Dictyoglomi* (Saiki et al., 1985; Euzéby, 2012). They have been cultivated from or detected in anaerobic, hyperthermophilic hot spring environments (Patel et al., 1987; Svetlichny and Svetlichnaya, 1988; Mathrani and Ahring, 1991; Kublanov et al., 2009; Gumerov et al., 2011; Kochetkova et al., 2011; Burgess et al., 2012; Sahm et al., 2013; Coil et al., 2014; Menzel et al., 2015) or isolated from paper-pulp factory effluent (Mathrani and Ahring, 1992), but only two *Dictyoglomus* species have been validly described in the literature (Saiki et al., 1985; Svetlichny and Svetlichnaya, 1988). Both strains grow up to 80°C, are Gram negative, and exhibit unusual morphologies consisting of filaments, bundles, and spherical bodies. The first described *Dictyoglomus* species, *Dictyoglomus thermophilum* was isolated from Tsuetate Hot Spring in Kumamoto Prefecture, Japan (Saiki et al., 1985). The genome of *D. thermophilum* has been

sequenced (Coil et al., 2014), and a number of potentially useful enzymes including amylase (Fukusumi et al., 1988; Horinouchi et al., 1988), xylanases (Gibbs et al., 1995; Morris et al., 1998), a mannanase (Gibbs et al., 1999) and an endoglucanase (Shi et al., 2013) have been cloned and characterized. The second described species, *Dictyoglomus turgidus*, was isolated from a hot spring in the Uzon Caldera, in eastern Kamchatka, Russia (Svetlichny and Svetlichnaya, 1988). The name *Dictyoglomus turgidus* was subsequently corrected to *Dictyoglomus turgidum* (Euzéby, 1998). Unlike *D. thermophilum*, *D. turgidum* was reported to grow on a wide range of substrates including starch, cellulose, pectin, carboxymethylcellulose, lignin, and humic acids, but not on pentose sugars such as xylose and arabinose (Svetlichny and Svetlichnaya, 1988). Because of the wide range of substrates utilized, *D. turgidum* was selected for enzyme library construction and carbohydrase screening (Brumm et al., 2011) as well as whole genome sequencing. Here we describe the complete genome sequence of *D. turgidum*, bioinformatic analysis of the metabolism of this unusual organism, and comparative analysis with the genome of *D. thermophilum*. We also present functional analysis of its DNA Pol I gene and a number of novel carbohydrases.

## MATERIALS AND METHODS

*D. turgidum* strain 6724<sup>T</sup> was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ). 10G electrocompetent *E. coli* cells, pEZSeq (a lac promoter vector), Taq DNA polymerase and OmniAmp DNA polymerase were obtained from Lucigen, Middleton, WI. Azurine cross-linked-labeled polysaccharides were obtained from Megazyme International (Wicklow, Ireland). 4-methylumbelliferyl- $\beta$ -D-cellobioside (MUC), 4-methylumbelliferyl- $\beta$ -D-xylopyranoside (MUX), and 4-methylumbelliferyl- $\beta$ -D-glucopyranoside (MUG) were obtained from Research Products International Corp. (Mt. Prospect, IL). CellLytic IIB reagent, pNP- $\beta$ -glucoside, pNP- $\beta$ -cellobioside, 4-methylumbelliferyl- $\alpha$ -D-arabinofuranoside (MUA), 4-methylumbelliferyl- $\beta$ -D-lactopyranoside (MUL), 5-Bromo-4-chloro-3-indolyl  $\alpha$ -D-galactopyranoside (X- $\alpha$ -Gal, XAG), and 5-Bromo-4-chloro-3-indolyl  $\beta$ -D-galactopyranoside (X-gal, XG) were purchased from Sigma-Aldrich (St. Louis, MO). All other chemicals were of analytical grade.

*D. turgidum* DSM 6724<sup>TM</sup> was obtained from the DSMZ culture collection and maintained on DSM Medium 516 reduced with Na<sub>2</sub>S and N<sub>2</sub> at 75°C in Balch tubes with a headspace of N<sub>2</sub>. Cultures grown in 1 L stoppered flasks were harvested for DNA preparation. YT plate media (16 g/l tryptone, 10 g/l yeast extract, 5 g/l NaCl and 16 g/l agar) was used in all molecular biology screening experiments. Terrific Broth (12 g/l tryptone, 24 g/l yeast extract, 9.4 g/l K<sub>2</sub>HPO<sub>4</sub>, 2.2 g/l KH<sub>2</sub>PO<sub>4</sub>, and 4.0 g/l glycerol added after autoclaving) was used for liquid cultures.

A cell concentrate of *D. turgidum* strain 6724<sup>TM</sup> was lysed using a combination of SDS and proteinase (Sambrook et al., 1989) and genomic DNA was purified using phenol/chloroform extraction. The genomic DNA was precipitated, treated

with RNase to remove residual contaminating RNA, and fragmented by hydrodynamic shearing (HydroShear apparatus, GeneMachines, San Carlos, CA) to generate fragments of 2–4 kb. The fragments were purified on an agarose gel, end-repaired, and ligated into pEZSeq (Lucigen Corp., Middleton WI). The recombinant plasmids were then used to transform electrocompetent cells. A copy of the library containing the *Dictyoglomus turgidum* genomic DNA was submitted to the Joint Genome Institute of the Department of Energy for whole genome sequencing; a second copy of the library was used for carbohydrase screening experiments.

The genome of *D. turgidum* DSM 6724<sup>TM</sup> was sequenced at the Joint Genome Institute (JGI) using a combination of 3 and 8 kb DNA libraries. In addition to 20x Sanger sequencing, 454 pyrosequencing was done to a depth of 20x coverage. Draft assemblies were based on 32,817 total reads. The Phred/Phrap/Consed software package was used for sequence assembly and quality assessment (Ewing and Green, 1998; Gordon et al., 1998). After the shotgun stage, reads were assembled with parallel phrap. Possible mis-assemblies were corrected with Dupfinisher or transposon bombing of bridging clones. Gaps between contigs were closed by editing in Consed, custom primerwalking or PCR amplification. A total of 80 additional reactions were necessary to close gaps and to raise the quality of the finished sequence. The completed genome sequence of *D. turgidum* DSM 6724<sup>TM</sup> contains 34,756 reads, achieving an average of 17.3x coverage. The Accession number for the complete genome is NC\_011661.

Genes were identified using Prodigal (Hyatt et al., 2010) as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. These data sources were combined to assert a product description for each predicted protein. Non-coding genes and miscellaneous features were predicted using tRNAscan-SE (Lowe and Eddy, 1997), RNAMmer (Lagesen et al., 2007), Rfam (Griffiths-Jones et al., 2003), TMHMM (Krogh et al., 2001), CRISPRFinder (Grissa et al., 2007), and signalP (Krogh et al., 2001). RAST annotations (Aziz et al., 2008) of *D. turgidum* and *D. thermophilum* were carried out in parallel to further clarify genomic relationships using SEED genome comparison tools (Overbeek et al., 2005).

The phylogeny of *D. turgidum* was determined using its 16S ribosomal RNA (rRNA) gene sequence as well as those of the most closely related 16S rRNA sequences identified by BLASTn. 16S rRNA gene sequences were aligned using MUSCLE (Edgar, 2004), pairwise distances were estimated using the maximum composite likelihood (MCL) approach, and initial trees for heuristic search were obtained automatically by applying the neighbor-joining method in MEGA7 (Kumar et al., 2016). The alignment and heuristic trees were then used to infer the phylogeny using the maximum likelihood method based on Tamura-Nei (Tamura and Nei, 1993; Tamura et al., 2011). The phylogeny of the reverse gyrase protein sequence was inferred

using the Neighbor-Joining method. The optimal tree with the sum of branch length = 1.99686421 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The analysis involved 7 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 3230 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016).

The *endo*-glucanase specificity of enzymes was determined in 0.50 ml of 50 mM acetate buffer, pH 5.8, containing 0.2% azurine cross-linked-labeled (AZCL) insoluble substrates and 50  $\mu$ l of clarified lysate. Each purified enzyme was evaluated for *endo*-activities using the following set of substrates: AZCL-arabinan (AR), AZCL-arabinoxylan (AX), AZCL- $\beta$ -glucan (BG), AZCL-curdlan (CU), AZCL-galactan (GL), AZCL-galactomannan (GM), AZCL-hydroxyethyl cellulose (HEC), AZCL-pullulan (PUL), AZCL-rhamnogalacturonan (RH), and AZCL-xyloglucan (XG). Assays were performed at 70°C, with shaking at 1000 rpm, for 60 min in a Thermomixer R (Eppendorf, Hamburg, Germany). Tubes were clarified by centrifugation and absorbance values at 600 nm determined using a Bio-Tek EL<sub>800</sub> plate reader. The *exo*-glucanase specificity of enzymes was determined by spotting 2.0  $\mu$ l of clarified lysate directly on agar plates containing 10 mM 4-methylumbelliferyl substrate. Plates were placed in a 70°C incubator for 60 min and then examined using a hand-held UV lamp and compared to negative and positive controls for fluorescence.

Amplification efficacy was compared between Dtur, Taq and OmniAmp DNA polymerases (DNAP) in side by side PCR reactions using four different sized amplicons (0.9, 2.8, 5.0, and 10.0 Kb). PCR reaction conditions contained 1–20 ng of template DNA, 2.5U of Taq DNAP or 5U Dtur or OmniAmp DNAP (Lucigen Corp.), 200  $\mu$ M dNTPs, and 0.5  $\mu$ M primers in a 50  $\mu$ l reaction. DNAP buffer (1X) contained 10 mM Tris-HCl (pH 8.8), 10 mM KCl, 10 mM NH<sub>2</sub>SO<sub>4</sub>, 2 mM MgSO<sub>4</sub>, 0.1% tritonX-100, and 15% sucrose. Cycling conditions were 94°C 2 min and 30 cycles of 94°C for 15 s, 60°C for 30 s, and 72°C for 1 min per kb. The templates and PCR primers are as follows: pUC19 0.9 kb amplicon primers (CCC CTA TTT GTT TAT TTT TCT AAA ATT CAA TAT GTA TCC GCT and TTA CCA ATG CTT AAT CAG TGA GGC ACC TAT CT), *E. coli* 2.8 kb amplicon primers (TAC TGT CTG CCA TGG TTC AGA TCC CCC AAA ATC CAC TTA TCC TTG TAG A and TTA TCT GTG GTC GAC TTA GTG CGC CTG ATC CCA GTT TTC GCC ACT CCC CA), *E. coli* 5 kb amplicon primers (TCT CTC CGA CCA AAG AGT TG and GAA ACA TTG AGC GAA GAG GA), and *E. coli* 10 kb amplicon primers (CTA TGA TTA TCT AGG CTT AGG GTC AC and CAG TGT AGA GAG ATA GTC AGG AGT TA).

Functional screening for active carbohydrase enzymes involved plating transformed *E. coli* cells containing 2–4 kb Dtur genomic DNA inserts in the pEZSeq vector on YT agar

containing IPTG (for *lacZ* promoter induction) and one of the fluorescent substrates MUC, MUG or MUX. A long wavelength UV lamp was used to locate colonies that were fluorescent, which were sequenced by Sanger chemistry to identify the gene. Genes identified in the functional screen as well as additional genes of interest from the completed genome were amplified without their respective signal sequence, ligated into pET28A, and transformed into BL21(DE3) *E. coli* competent cells. Recombinant clones were cultured overnight at 37°C, 100 rpm, in 100 ml Luria Broth containing 50 mg/l kanamycin. Expression was induced using 1 mM IPTG, and cultures were harvested 18 h after induction. Cells were pelleted by centrifugation, and the pellets were lysed using Cellytic B reagent. Proteins were purified using standard methods for His-tagged proteins (Spriestersbach et al., 2015), and their purity and identity verified by SDS PAGE.

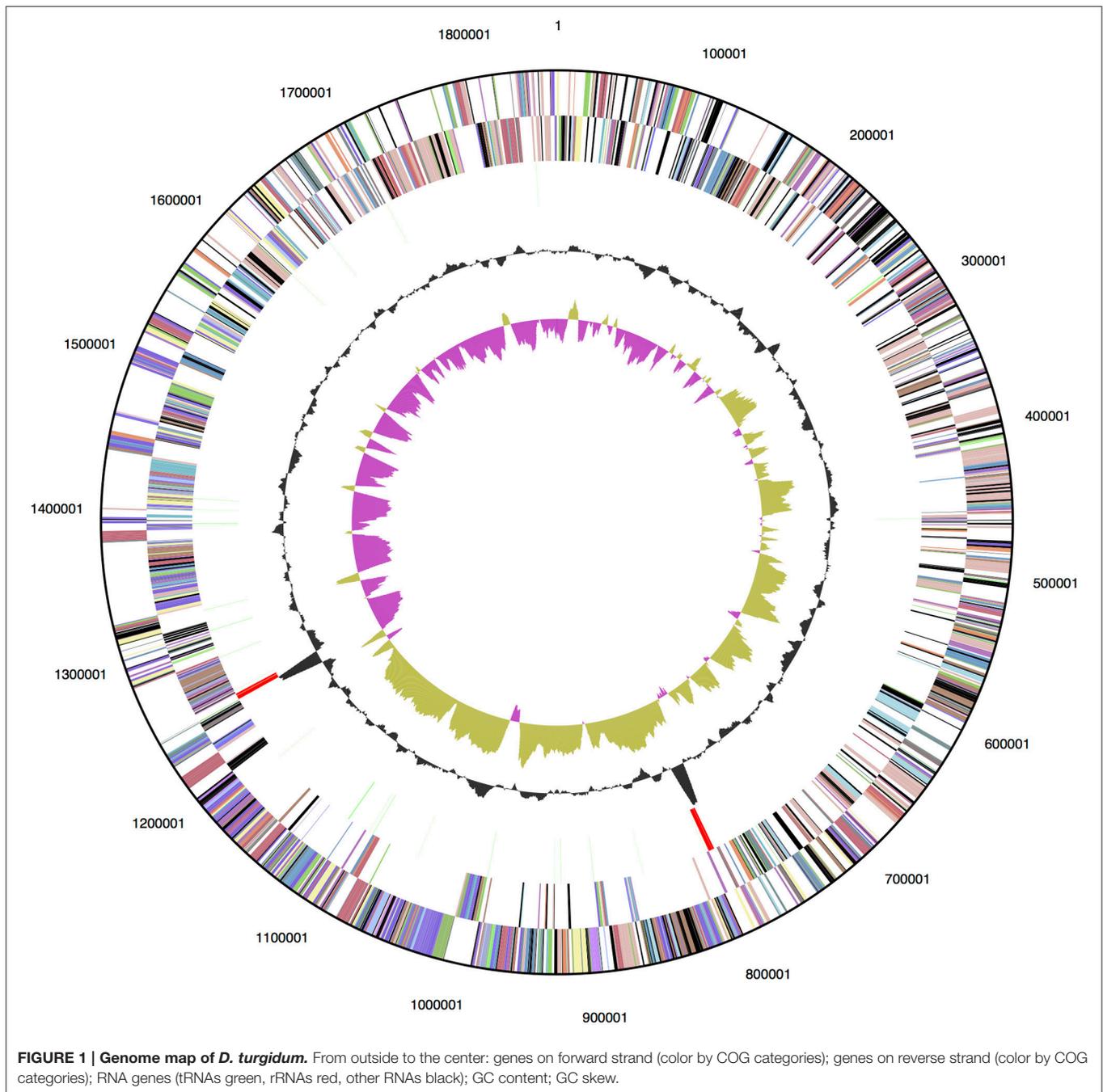
*D. turgidum* DNA polymerase I (Dtur DNAP) was cloned by PCR amplification using the proofreading enzyme Phusion (NEB, Waltham MA) and forward and reverse 24 base oligonucleotides that spanned the start and stop codons. The amplified DNA was inserted into the rhamnose promoter vector pRham containing an N terminal histidine tag and transformed into 10G competent *E. coli* cells (Lucigen Corp.). Recombinant Dtur DNAP production was induced by rhamnose and the enzyme was purified using standard methods for His-tagged proteins (Spriestersbach et al., 2015).

## RESULTS

### Genome of *D. turgidum*

The genome of *D. turgidum* DSM 6724<sup>TM</sup> consists of a single chromosome of 1,855,560 bp and no plasmids or extrachromosomal elements. The GC content of the chromosome is 33.96% based on the genome sequence, slightly higher than the reported value of 32.5% (Svetlichny and Svetlichnaya, 1988) and is predicted to contain 1813 protein-coding genes and 52 RNA genes (**Figure 1**). The completed genome sequence is available from GenBank (GenBank: CP001251.1). Based on 16S rRNA gene sequence analysis, *D. turgidum* DSM 6724 and *D. thermophilum* are separate species. This is confirmed by average nucleotide analysis (ANI), where *D. turgidum* and *D. thermophilum* are calculated to have 82.4% average nucleotide identity, below the threshold for members of the same species.

Of the 1813 protein-coding genes, 1354 genes (72.6%) were assigned to COGs categories (**Table 1**). The fraction of the genes annotated as members of COG class G, carbohydrate transport and metabolism (highlighted in bold), 13.4%, is greater than the fraction observed for 95% of genomes in the MicrobesOnline database (Dehal et al., 2010). This represents the lower limit of proteins involved in carbohydrate metabolism, because it does not include any proteins in categories R, S or not in COGS that were not identified by the algorithm as being involved in carbohydrate metabolism. A number of pectate lyases, for example, are not identified as members of COGs class G. No other COGs category had a significantly higher than average number of members, and no COGs category had a significantly lower than average percentage of members.



## Genomic Insights into the Relationship of *D. turgidum* to *D. thermophilum* and Other Organisms

While being separate species, an in-depth comparison of the two *Dictyoglomi* genomes shows that *D. turgidum* is closely related to *D. thermophilum* on a number of levels. The genomes are similar in size, with *D. turgidum* being slightly smaller than the genome of *D. thermophilum* (1,855,560 bp vs. 1,959,987 bp) and containing approximately 100 fewer protein coding genes (1813

vs. 1912). The two organisms have a highly conserved set of genes present in their genomes. Over 95% of the proteins present in *D. turgidum* have orthologs in *D. thermophilum*. There are only 43 proteins of greater than 100 amino acids present in *D. turgidum* without orthologs in *D. thermophilum*, and there are only 109 proteins of greater than 100 amino acids present in *D. thermophilum* without orthologs in *D. turgidum*. Of the proteins with orthologs in both species, there are 614 proteins with >90% sequence identity.

**TABLE 1 | Number of genes associated with general COG functional categories.**

Code	Value	Percentage	Description
J	168	11.0%	Translation, ribosomal structure and biogenesis
K	76	5.0%	Transcription
L	61	4.0%	Replication, recombination and repair
B	1	0.1%	Chromatin structure and dynamics
D	19	1.2%	Cell cycle control, Cell division, chromosome partitioning
V	40	2.6%	Defense mechanisms
T	48	3.1%	Signal transduction mechanisms
M	87	5.7%	Cell wall/membrane biogenesis
N	20	1.3%	Cell motility
U	18	1.2%	Intracellular trafficking and secretion
O	61	4.0%	Posttranslational modification, protein turnover, chaperones
C	79	5.2%	Energy production and conversion
<b>G</b>	<b>205</b>	<b>13.4%</b>	<b>Carbohydrate transport and metabolism</b>
E	170	11.1%	Amino acid transport and metabolism
F	60	3.9%	Nucleotide transport and metabolism
H	73	4.8%	Coenzyme transport and metabolism
I	44	2.9%	Lipid transport and metabolism
P	77	5.0%	Inorganic ion transport and metabolism
Q	18	1.2%	Secondary metabolites biosynthesis, transport and catabolism
R	130	8.5%	General function prediction only
S	58	3.8%	Function unknown
–	511	27.4%	Not in COGs

Highlighted in bold, COG class G. The fraction of the genes annotated as members of this class is greater than the fraction observed for 95% of genomes in the MicrobesOnline database.

Synteny plots were generated using both RAST and IMG annotation methods. The two annotation methods gave essentially identical plots, as did plots based on DNA or protein sequences. The plots show the genomes of *D. turgidum* and *D. thermophilum* have highly conserved large and small-scale organization (**Figure 2A**). This conserved organization appears to be an unusual phenomenon. Two sets of thermophilic organisms with similar ANI values, *T. thermophilus* and *T. aquaticus* (84.3% ANI, **Figure 2B**) and *C. bescii* and *C. saccharolyticus* (82.0% ANI, **Figure 2C**) show only limited short-range synteny and no extensive long-range synteny. It is unclear if this conserved genomic organization is limited to these two species, or is present in all *Dictyoglomi* genomes.

The relationship of these two *Dictyoglomus* species to other organisms appears significantly more complicated, depending on the type of analysis and interpretation (Love et al., 1993; Rees et al., 1997; Takai et al., 1999; Ding et al., 2000; Wagner and Wiegel, 2008). Phylogenetic analysis using 16S rRNA shows the two *Dictyoglomus* species appear most closely related to *Thermotoga* species before bootstrapping (data not shown). After bootstrapping, the relationship shifts dramatically, with the two *Dictyoglomus* species becoming most closely related to *Caldicellulosiruptor* species (**Figure 3**). Previous work using average nucleotide identity (ANI) calculations (Nishida

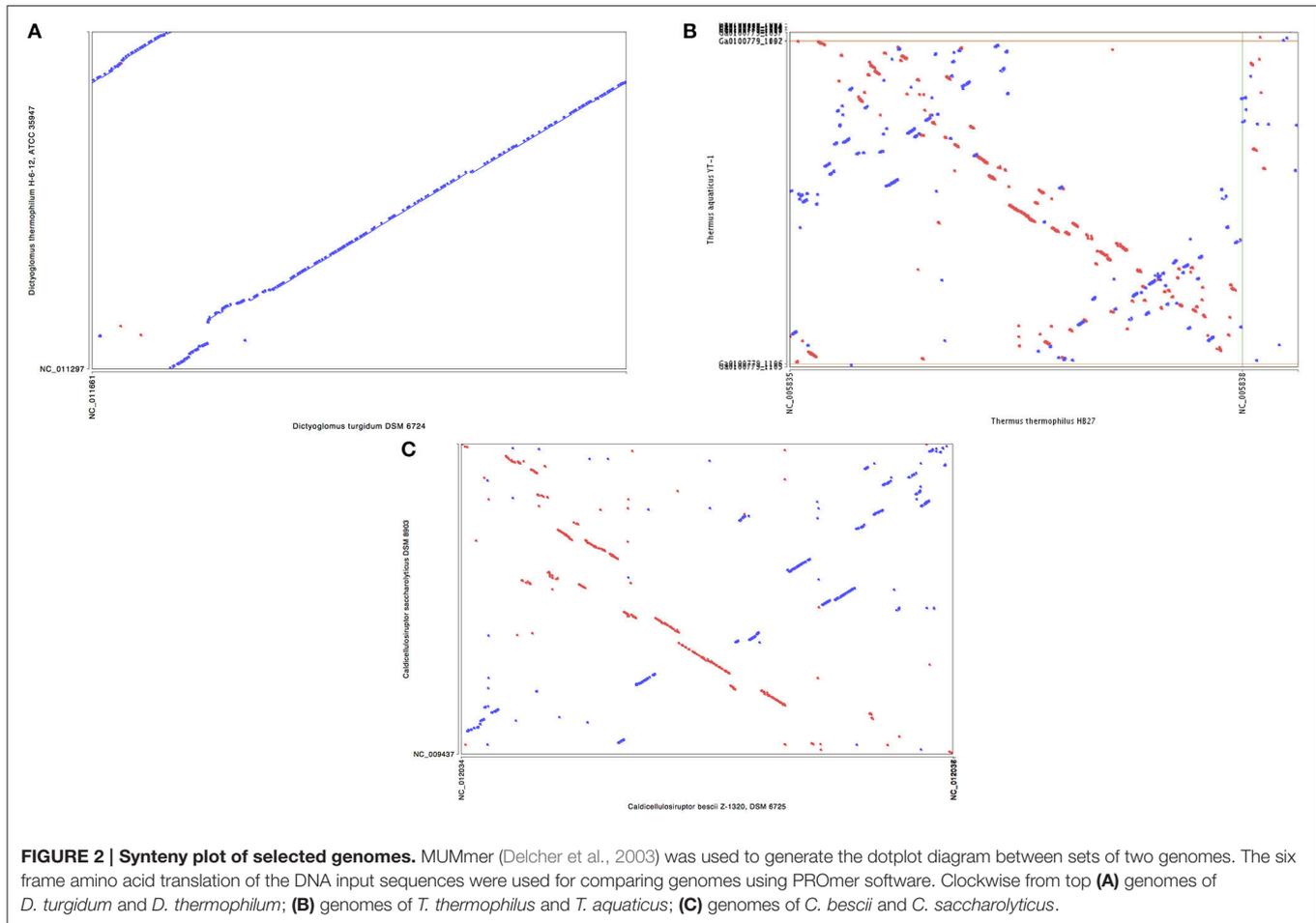
et al., 2011) identified *Thermotoga* species as the closest relatives to *Dictyoglomus*. ANI values were generated using the *D. thermophilum* genome, eight finished, closed *Thermotoga* genomes and three finished, closed *Caldicellulosiruptor* genomes. ANI values (Kim et al., 2014) were computed as pairwise bidirectional best nSimScan hits of genes having 70% or more identity and at least 70% coverage of the shorter gene. ANI calculations performed as described above yielded 82.4% identity between the genomes of *D. turgidum* and *D. thermophilum*, based on 1584 proteins (87% of the genome) that met the criteria. The value of 82.4% is well below the cut-off value of 98% for strains of the same species, and confirms that *D. turgidum* and *D. thermophilum* are separate species. The ANI calculations found 67–68% identity between *D. turgidum* and the three *Caldicellulosiruptor* species, based on 124–129 proteins per genome that met the criteria for the calculation (approximately 7% of the genome). ANI calculations found 66–68% identity between *D. turgidum* and the eight *Thermotoga* species, based on the 36–64 proteins per genome that met the criteria (approximately 2–4% of the genome). Rather than identifying relationships among these organisms, the low number of proteins in *D. turgidum* with at least 70% identity to the proteins in these 11 strains (on which these ANI values are calculated) further demonstrates the uniqueness of this organism.

## Protein and Amino Acid Metabolism

Based on the MEROPS database (Rawlings et al., 2014), the *D. turgidum* genome codes for 55 potential peptidases. This value is within the range of peptidases reported in the database for *Thermotoga* species (52–67) and *Caldicellulosiruptor* species (54–74). Of the 55 potential peptidases, only a single peptidase, Dtur\_0603, possesses an annotated signal sequence and is predicted to be secreted. While possessing only a single secreted peptidase to generate amino acids and peptides, *D. turgidum* possesses nine potential membrane transporter systems to transport amino acids and peptides into the cell. These nine transporters include seven annotated oligopeptide/dipeptide ABC transporter systems (Dtur\_0082 through Dtur\_0086; Dtur\_0158 through Dtur\_0162; Dtur\_0214 through Dtur\_0217; Dtur\_0664 through Dtur\_0668; Dtur\_1061 through Dtur\_1064; Dtur\_1704 and Dtur\_1707; Dtur\_1719 through Dtur\_1722) as well as two amino acid ABC transporter systems (Dtur\_1051 through Dtur\_1053 and Dtur\_0932 through Dtur\_0936).

*D. turgidum* appears to utilize the amino acids and peptides taken up for protein synthesis, but it is unable to metabolize most amino acids as an energy or carbon source. Based on the BioCyc (Karp et al., 2005; Caspi et al., 2014) and SEED (Devoid et al., 2013) metabolic reconstructions from the genome sequence, *D. turgidum* is lacking degradation pathways for the following 13 amino acids: aspartate, asparagine, cysteine, histidine, isoleucine, leucine, lysine, phenylalanine, proline, serine, tryptophan, tyrosine, and valine. Arginine is not metabolized, but may be converted to putrescine.

Only four amino acids appear to be metabolized by *D. turgidum*. Glutamate is converted to methyl aspartate using glutamate mutase (Dtur\_1345 through Dtur\_1347) and then to pyruvate and acetate. Threonine can be degraded to glycine and acetaldehyde via threonine aldolase (Dtur\_0449), and the



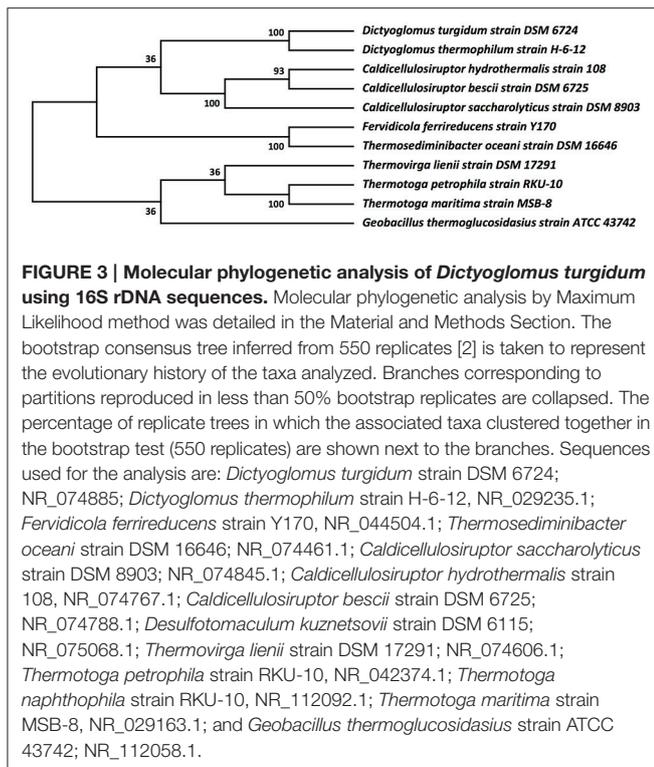
acetaldehyde generated is then converted to acetyl-CoenzymeA (acetyl-CoA) via aldehyde dehydrogenase (Dtur\_0484). Alanine can be converted to pyruvate by alanine dehydrogenase (Dtur\_1049), and glycine can be converted to ammonium 5,10-methylenetetrahydrofolate via glycine dehydrogenase and glycine cleavage system T protein (Dtur\_1515 through Dtur\_1518). The ability to utilize these four amino acids may be responsible for the observation of growth by *D. turgidum* on yeast extract, peptone, and casamino acids (Svetlichny and Svetlichnaya, 1988).

## Monosaccharide Metabolism

Based on the genomic reconstruction of Dtur, the organism is able to metabolize most five and six carbon sugars, and the following pathways are predicted. Arabinose is utilized via isomerization to L-ribulose (Dtur\_0379, or other isomerase), phosphorylation by L-ribulose kinase (Dtur\_1748) to L-ribulose-5-phosphate, and isomerization by L-ribulose-5-phosphate-4-epimerase (Dtur\_1734) to D-xylulose-5-phosphate, which is then metabolized via the pentose phosphate pathway. Rhamnose is utilized via isomerization by L-rhamnose isomerase to L-rhamulose (Dtur\_0427), phosphorylation by L-rhamulose kinase (Dtur\_1748) to L-rhamulose-1-phosphate, and cleavage into

dihydroxyacetone phosphate and L-lactaldehyde. Xylose is utilized via isomerization by xylose isomerase (Dtur\_0036 or other sugar isomerase) to xylulose, and the xylulose is phosphorylated by xylulose kinase to (Dtur\_0920) to D-xylulose-5-phosphate, which is then metabolized via the pentose phosphate pathway.

Fucose is utilized via isomerization by L-fucose isomerase to L-fuculose (Dtur\_0410), phosphorylation by L-fuculokinase (Dtur\_0920) to L-fuculose-1-phosphate, and cleavage into dihydroxyacetone phosphate and L-lactaldehyde. Galactose is phosphorylated by galactose kinase (Dtur\_1195) to galactose-1-phosphate, which is converted to UDP-galactose by galactose-1-phosphate uridylyl transferase (Dtur\_1196), isomerized by UDP-glucose-4-epimerase (Dtur\_1352) to UDP-glucose, and finally to glucose-1-phosphate by UTP-glucose-1-phosphate uridylyltransferase (Dtur\_1627). Mannose is phosphorylated by mannose kinase (Dtur\_0176; Dtur\_0716 or other annotated sugar kinase) to generate mannose-1-phosphate. The mannose-1-phosphate is isomerized to mannose-6-phosphate by phosphomannomutase/phosphoglucosyltransferase (Dtur\_0067) and then to fructose-6-phosphate by phosphoglucose/phosphomannose isomerase (Dtur\_1271). UDP-glucose is either isomerized to fructose, or oxidized



to UDP-glucuronate using either Dtur\_575 or Dtur\_718. The UDP-glucuronate can then be further oxidized to ribulose-5-phosphate by 6-phosphogluconate dehydrogenase (Dtur\_0197).

Galacturonate generated by pectin degradation may be epimerized by one of the six UDP sugar epimerase genes found in the genome. Rarely-encountered sugars may be handled by any of a number of sugar isomerases. Dtur rhamnose isomerase (Dtur\_0427) isomerizes seven monosaccharides: L-rhamnose, L-lyxose, L-mannose, L-xylulose, L-fructose, D-allose, and D-ribose (Kim et al., 2013). The Dtur fucose isomerase (Dtur\_0410) isomerizes L-fucose, D-arabinose, D-altrose, and L-galactose (Hong et al., 2012). Dtur also possesses a cellobiose 2-epimerase that may isomerize non-metabolized disaccharides into easily-degradable ones (Kim et al., 2012).

## Polysaccharide Degradation and Transport

Polysaccharide degradation by *D. turgidum* is of interest for a number of reasons. Analysis of the *D. turgidum* genome shows an enrichment in COGS family members annotated as involved in carbohydrate transport and metabolism (Table 1). *D. turgidum* is reported to utilize polysaccharides such as starch, cellulose, pectin, glycogen, and carboxymethyl cellulose (Svetlichny and Svetlichnaya, 1988) while *D. thermophilum* is reported to utilize starch, but not cellulose. Finally, a number of carbohydrates with potential industrial applications have been identified in the two *Dictyoglomus* species including amylases and xylanases. A combination of genomic and enzymatic analyses was carried out to clarify the polysaccharide degradation potential of *D. turgidum*.

**TABLE 2 | Annotated secreted polysaccharide-degrading enzymes.**

Gene	GH family	Annotated activity	Nearest ortholog	Identity
Dtur_0097	GH 44	$\beta$ -mannanase	Calkro_0851	70.1%
Dtur_0172	GH 28	pectinase	Cphy_3310	47.1%
Dtur_0243	GH 11	xylanase	Calkro_0081	83.7%
Dtur_0276	GH 5	cellulase	Mahau_0466	59.9%
Dtur_0277	GH 26	$\beta$ -mannanase	BG52_11385	52.7%
Dtur_0430	PL 1	pectate lyase	SNOD_03765	42.1%
Dtur_0431	PL 1	pectate lyase	M769_0111315	60.1%
Dtur_0432	PLNC	pectate lyase	CSE_02370	57.3%
Dtur_0433	CE 8	pectin esterase	Calkro_0154	56.0%
Dtur_0628	GH 12	curdlanase	CTN_1107	48.4%
Dtur_0669	GH 5	cellulase	Mahau_0466	54.7%
Dtur_0675	GH 57	$\alpha$ -amylase	ANT_11030	41.3%
Dtur_0676	CBM9	$\alpha$ -amylase	COCOR_00322	39.7%
Dtur_0857	GH 53	$\beta$ -galactanase	TRQ7_08325	56.5%
Dtur_1586	GH 5	cellulase	BSONL12_10711	41.5%
Dtur_1675	GH 13	$\alpha$ -amylase	CAAU_0986	51.6%
Dtur_1715	GH 10	xylanase	Pmob_0231	46.9%
Dtur_1729	GH 43	$\beta$ -xylosidase	Csac_1560	67.9%
Dtur_1739	GH 51	$\beta$ -xylosidase	Calhy_1625	58.9%
Dtur_1740	GH 39	$\beta$ -xylosidase	TRQ7_03440	38.3%

Analysis of the *D. turgidum* genome reveals a wide range of genes coding for annotated extracellular and intracellular polysaccharide degrading enzymes. The CAZy database (Lombard et al., 2014) identifies 57 glycosyl hydrolases (GH), 3 polysaccharide lyases (PL) and 6 carbohydrate esterases (CE) in the Dtur genome. Based on signal sequence predictions (Petersen et al., 2011), 20 of the polysaccharide-degrading enzymes are secreted into the medium (Table 2), where they degrade polysaccharides into oligosaccharides and monosaccharides. After polysaccharide degradation, 18 annotated three-component ABC carbohydrate transporters are predicted to transport monosaccharides and oligosaccharides into the cell. *D. turgidum* is reported to utilize fructose, glucose, rhamnose, inositol, mannitol, and sorbitol (Svetlichny and Svetlichnaya, 1988), indicating ABC carbohydrate transporters exist for these monosaccharides and sugar alcohols. *D. turgidum* cannot utilize arabinose, fucose, galactose, mannose, or xylose, indicating a lack of dedicated transport systems for these monosaccharides. These sugars may be transported into the cell as oligosaccharides by the oligosaccharide transporters and degraded to monosaccharides in the cytoplasm. Once inside the cell, oligosaccharides are degraded into monosaccharides by a combination of 46 *exo*-acting and *endo*-acting enzymes (Table 3). Working together, these 46 enzymes appear capable of degrading oligosaccharides from most plant-based polysaccharides to monosaccharides.

BLAST analysis was used to determine the closest orthologs of the 66 Dtur CAZymes. Of these 66 enzymes, 56 have their closest orthologs in *D. thermophilum*, with 80–90% amino acid identity. The remaining 10 enzymes have no orthologs in *D. thermophilum*. Seven of the ten unique enzymes in

**TABLE 3 | Annotated intracellular polysaccharide-degrading enzymes.**

Gene	GH family	Annotated activity	Nearest ortholog	Identity
Dtur_0081	GH 2	$\beta$ -galactosidase	Calhy_1828	60.9%
Dtur_0157	GH 4	$\alpha$ -glucosidase	Mc24_02443	47.5%
Dtur_0171	GH 31	$\alpha$ -glucosidase	A500_11654	44.9%
Dtur_0219	GH 3	$\beta$ -glucosidase	<i>D. tunisiensis</i> bglB3	67.3%
Dtur_0222	GH 20	$\beta$ -hexosaminidase	CDSM653_01797	67.2%
Dtur_0242	CE NC	feruloyl esterase	TM_0033	55.1%
Dtur_0265	CE 7	acetyl xylan esterase	Tmari_0074	66.4%
Dtur_0289	GH 3	$\beta$ -glucosidase	Cst_c03130	66.8%
Dtur_0315	GH 29	$\alpha$ -fucosidase	Tthe_0662	60.7%
Dtur_0320	GH 31	$\alpha$ -glucosidase	Csac_1354	65.9%
Dtur_0321	GH 3	$\beta$ -glucosidase	Cst_c12090	50.1%
Dtur_0384	GH 4	$\alpha$ -glucosidase	CTER_5006	48.4%
Dtur_0435	PL 1	pectate lyase	MB27_42800	36.0%
Dtur_0440	GH 4	$\alpha$ -galacturonidase	BTS2_1711	61.6%
Dtur_0450	CE 4	deacetylase	Tnap_0743	67.4%
Dtur_0451	GH 16	curdlanase	TRQ7_04835	50.9%
Dtur_0462	GH 1	$\beta$ -glucosidase	CLDAP_02840	48.5%
Dtur_0490	GH 31	$\alpha$ -glucosidase	Tbis_2416	45.4%
Dtur_0502	GH 127	$\beta$ -L-arabinofuranosidase	CTN_0404	56.3%
Dtur_0505	GH 42	$\beta$ -galactosidase	Mahau_1293	59.2%
Dtur_0523	GH 18	chitinase	Bccel_2454	50.1%
Dtur_0551	GH 32	invertase	Calhy_2186	47.6%
Dtur_0629	GH 26	$\beta$ -mannanase	Calkro_1144	54.5%
Dtur_0650	GH 31	$\alpha$ -glucosidase	TheetDRAFT_1156	45.2%
Dtur_0658	GH 130	$\alpha$ -D-mannosyltransferase	X274_02975	41.2%
Dtur_0670	GH 5	cellulase	Mahau_0466	61.7%
Dtur_0671	GH 5	cellulase	TM_1752	58.7%
Dtur_0770	GH 57	$\alpha$ -amylase	BROSI_A0626	37.3%
Dtur_0794	GH 13	$\alpha$ -amylase	AC812_10325	35.3%
Dtur_0852	GH 3	$\beta$ -glucosidase	M164_2324	58.2%
Dtur_0895	GH 57	$\alpha$ -amylase	TSIB_1115	46.5%
Dtur_0896	GH 57	$\alpha$ -amylase	Calab_2422	40.9%
Dtur_1539	GH 2	$\beta$ -glucuronidase	Calkro_0120	60.7%
Dtur_1647	GH 10	xylanase	PaelaDRAFT_3013	51.2%
Dtur_1670	GH 36	$\alpha$ -galactosidase	Calla_1244	77.7%
Dtur_1677	GH 4	$\beta$ -glucosidase	L21TH_1859	47.0%
Dtur_1714	GH 67	$\alpha$ -glucuronidase	Mc24_01903	69.4%
Dtur_1723	GH 3	$\beta$ -glucosidase	<i>C. polysaccharolyticus</i> Xyl3A	46.7%
Dtur_1735	GH 51	$\beta$ -xylosidase	COB47_1422	70.2%
Dtur_1749	GH 4	$\alpha$ -glucosidase	TRQ7_00895	68.7%
Dtur_1758	GH 38	$\alpha$ -mannosidase	CTN_0786	41.3%
Dtur_1799	GH 1	$\beta$ -glucosidase	Hore_15280	57.7%
Dtur_1800	GH 43	$\beta$ -xylosidase	Athe_2555	82.9%
Dtur_1802	GH 2	$\beta$ -galactosidase	Thewi_0408	42.2%

*D. turgidum* are secreted enzymes, including three of the four predicted pectin-degrading enzymes, three of the four predicted xylan-degrading enzymes and the predicted *endo*-arabinase. Only two Dtur enzymes, Dtur\_0243 and Dtur\_1800, have non-*Dictyoglomus* orthologs with over 80% identity (Tables 2, 3). The nearest non-*Dictyoglomus* orthologs of most of the 66 have

<60% identity, showing the uniqueness of the Dtur enzymes. The wide range of organisms these orthologs are found in further demonstrates the uniqueness of this organism. Of the 66 enzymes, 13 have nearest orthologs in *Caldicellulosiruptor* species and 13 have nearest orthologs in *Thermotoga* species. Five orthologs are found in mesophilic *Clostridia* species,

four in mesophilic *Mahella* species, and three in thermophilic *Thermoanaerobacter* species. The remaining 28 orthologs are spread over a wide range of mesophilic and thermophilic organisms.

### Degradation of Polymeric Substrates Substrates Reported to Be Degraded for Which Genomic and Enzymatic Support Exists

Both *D. turgidum* and *D. thermophilum* are reported to utilize starch, and a number of  $\alpha$ -amylases have been cloned and characterized from *D. thermophilum*. The genome of *D. turgidum* codes for three annotated extracellular  $\alpha$ -amylases (Dtur\_0675; Dtur\_0676, and Dtur\_1675) as well as four annotated intracellular  $\alpha$ -amylases (Dtur\_0770; Dtur\_0794; Dtur\_0895, and Dtur\_0896) and six annotated  $\beta$ -glucosidases (Dtur\_0157; Dtur\_0171; Dtur\_0320; Dtur\_0384; Dtur\_0490, and Dtur\_1749). These intracellular enzymes may function in both degradation of starch oligosaccharides transported into the cell as well as degradation of glycogen stored in the cell.

*D. turgidum* is reported to utilize pectin, while no data on pectin utilization was reported for *D. thermophilum*. The genome of *D. turgidum* possesses three annotated secreted pectin lyases (Dtur\_0430, Dtur\_0431, and Dtur\_0432), one secreted pectin esterase (Dtur\_0433), an annotated intracellular pectin lyase (Dtur\_0435), and an annotated intracellular  $\alpha$ -galacturonidase (Dtur\_0440).

*D. turgidum* is reported to utilize carboxymethyl cellulose. Because carboxymethyl cellulose is a man-made chemically-modified derivative of cellulose, there are no specific annotated carboxymethyl cellulases present in nature. Unlike cellulose which is crystalline and insoluble in water, carboxymethyl cellulose is an amorphous polymer that is soluble in aqueous solutions. As a result of this solubility, carboxymethyl cellulose is used as a substrate in the assay of a number of enzyme families including xylanases, cellulases, and  $\beta$ -glucanases. The genome of *D. turgidum* has two annotated, secreted xylanases (Dtur\_0243 and Dtur\_1715) and three secreted annotated cellulases (Dtur\_0276; Dtur\_0669, and Dtur\_1586). The organism also possesses one annotated intracellular xylanase (Dtur\_1647), two intracellular cellulases (Dtur\_0670 and Dtur\_0671) as well as six annotated  $\beta$ -glucosidases (Dtur\_0219; Dtur\_0289; Dtur\_0321; Dtur\_0462; Dtur\_1723, and Dtur\_1799). Assay of the enzymes expressed and purified in this work showed three cellulases (Dtur\_0276; Dtur\_0670 and Dtur\_0671) and two xylanases (Dtur\_1647 and Dtur\_1715) utilized carboxymethyl cellulose as substrate, producing high levels of reducing sugars from a carboxymethyl cellulose solution (data not shown). These enzymatic assay results confirm the genomic analyses indicating that *D. turgidum* can utilize carboxymethyl cellulose.

### Substrates Reported to Be Degraded for Which Genomic and Enzymatic Support Does Not Exist

*D. turgidum* is reported to utilize crystalline cellulose (Svetlichny and Svetlichnaya, 1988), though the authors report “the organism grew markedly less readily on microcrystalline cellulose, lignin and humic acids.” This in contrast to

*D. thermophilum*, which is reported unable to utilize cellulose. Microbial degradation of crystalline cellulose requires the expression and secretion of multiple cellulases and accessory proteins to decrystallize the cellulose chains and generate soluble, low molecular weight cellodextrins (Brumm, 2013). These cellodextrins are then taken up via membrane transporters and further degraded into glucose monomers in the cytoplasm. Genomic and enzymatic analysis of *D. turgidum* indicate the organism is most likely unable to degrade crystalline cellulose. Comparison of the two genomes indicates *D. turgidum* contains no additional annotated cellulases not found in *D. thermophilum*. All five of the *D. turgidum* cellulases (Dtur\_0276; Dtur\_0669; Dtur\_0670; Dtur\_0671 and Dtur\_1586) have orthologs in *D. thermophilum* (Dicth\_0008; Dicth\_0505; Dicth\_0506; Dicth\_0508 and Dicth\_1476, respectively). Analysis of the genome shows a lack of GH9, GH6, GH8, GH12, or GH48 cellulases found in truly cellulytic organisms (Brumm, 2013). Close examination of the genome reveals no cellulases containing CBM2 or CBM3 modules present in cellulose-degrading *Caldicellulosiruptor* species or celulosomal structures present in cellulose-degrading *C. thermocellum* species within the genome. Assay of the enzymes expressed and purified in this work showed three intracellular enzymes (Dtur\_1647; Dtur\_0670 and Dtur\_0671) produced low levels of reducing sugars from crystalline cellulose (data not shown). The two secreted xylanases (Dtur\_0276 and Dtur\_1715) produced no reducing sugar from the crystalline cellulose. The lack of activity by the secreted enzymes confirms the genomic analyses indicating that *D. turgidum* most likely cannot utilize crystalline cellulose as a growth substrate. The microcrystalline cellulose preparation used in the original study may have contained glucan, xylan or mannan, resulting in the observed weak growth of the organism on cellulose used in the experiments (Svetlichny and Svetlichnaya, 1988).

The ability of *D. turgidum* to utilize lignin and humic acids is questionable for many of the reasons described above. The authors do not describe the source, purification, and analysis of the lignin and humic acids used in the growth experiments. Depending on the method of purification, lignin is often contaminated with mannan, cellulose and hemicellulose. Humic acids and lignin also contain sugars chemically bonded via ester linkages. Utilization of these sugars may be responsible for the low-level growth seen with these substrates. Thermophilic organisms capable of degrading aliphatic and aromatic organic compounds such as *Geobacillus* species, contain clearly identifiable extended gene clusters with these functions. For example, *Geobacillus* species Y41MC52 possesses three clusters annotated for degradation of aromatic acid molecules, GYMC52\_1956 through GYMC52\_1962; GYM C52\_1990 through GYMC52\_2001, and GYMC52\_3134 through GYMC52\_3141. Three similar clusters are found in the related strain *Geobacillus* species Y41MC61. Manual annotation of the *D. turgidum* genome failed to identify orthologs of any of the genes present in the three clusters, confirming that *D. turgidum* cannot utilize the aromatic ring structures found in lignin and humic acids.

## Substrates Not Reported to Be Degraded for Which Genomic and Enzymatic Support Exists

No data was reported on xylan utilization by *D. turgidum*, however a xylanase has been cloned and expressed from *D. thermophilum*. The genome of *D. turgidum* has two annotated, secreted xylanases (Dtur\_0243 and Dtur\_1715) and three secreted  $\beta$ -xylosidases (Dtur\_1729; Dtur\_1739, and Dtur\_1740). Genes for intracellular enzymes annotated as feruloyl esterase (Dtur\_0242), acetyl xylan esterase (Dtur\_0265), xylanase (Dtur\_1647),  $\alpha$ -glucuronidase (Dtur\_1714), and two  $\beta$ -xylosidases (Dtur\_1735 and Dtur\_1800) may be involved in degradation of oligosaccharides derived from xylan.

Mannans and glucans comprise a diverse group of plant-based polysaccharides that share a  $\beta$ -linked hexose backbone. Among the members of these two groups are mannan, glucomannan, galactomannan, galactoglucomannan,  $\beta$ -glucan, curdlan, and xyloglucan. No data was reported on mannan or glucan utilization by either *D. turgidum* or *D. thermophilum*. The genome of *D. turgidum* codes for two annotated, secreted  $\beta$ -mannanases (Dtur\_0097, and Dtur\_0277) and one intracellular  $\beta$ -mannanase (Dtur\_0629), three secreted annotated cellulases (Dtur\_0276; Dtur\_0669, and Dtur\_1586) and two intracellular cellulases (Dtur\_0670 and Dtur\_0671), as well as six annotated  $\beta$ -glucosidases (Dtur\_0219; Dtur\_0289; Dtur\_0321; Dtur\_0462; Dtur\_1723, and Dtur\_1799).

Arabinogalactan is a polysaccharide found in many plants, with the highest concentration in larch wood. No data was reported on arabinogalactan utilization by either *D. turgidum* or *D. thermophilum*. The genome of *D. turgidum* codes for a secreted annotated  $\beta$ -galactanase (Dtur\_0857) as well as four cytoplasmic  $\beta$ -galactosidases (Dtur\_0081; Dtur\_0081; Dtur\_0505, and Dtur\_1802) and one cytoplasmic  $\alpha$ -galactosidase (Dtur\_1670). Together these enzymes may be adequate for degradation of arabinogalactan as well as galactose-containing oligosaccharides. Annotated genes also code for intracellular fucosidase (Dtur\_0315), invertase (Dtur\_0551), galacturonidase (Dtur\_0440) and  $\beta$ -glucuronidase (Dtur\_1539), pectate lyase (Dtur\_0435), and chitinase (Dtur\_0523).

To verify the activities of some of these enzymes, cloning, expression, and purification was attempted for 30 of the annotated carbohydrase genes. Of these thirty, 16 Dtur enzymes were successfully expressed, purified, and characterized (Table 4). The 16 included two each of GH1 and GH3, four GH5, two GH10, and one each of GH36, GH42, GH43, GH53, GH57, and GH67. The remaining genes either failed to give amplicons of the correct size, or failed to express a soluble protein of the correct molecular weight (Brumm et al., 2011).

The two GH1 family members, Dtur\_0462 and Dtur\_1799, are annotated as  $\beta$ -glucosidases. These two cloned enzymes possess not only the predicted  $\beta$ -glucosidase activity, but also possess  $\beta$ -cellobiosidase,  $\beta$ -galactosidase,  $\beta$ -xylosidase and  $\beta$ -arabinofuranosidase activities (Table 4). GH 3 family members Dtur\_0852 and Dtur\_1723, also annotated as  $\beta$ -glucosidases, show  $\beta$ -glucosidase,  $\beta$ -xylosidase and  $\beta$ -arabinofuranosidase activity. Dtur\_0852 also possesses  $\beta$ -cellobiosidase activity, which is absent in Dtur\_1723 (Table 4).

TABLE 4 | Enzymatic activity of cloned gene products.

Gene	GH family	Annotated activity	Substrates hydrolyzed
Dtur_0462	GH 1	$\beta$ -glucosidase	MUA, MUC, MUG, MUX, XG
Dtur_1799	GH 1	$\beta$ -glucosidase	MUA, MUC, MUG, MUX, XG
Dtur_0852	GH 3	$\beta$ -glucosidase	MUA, MUC, MUG, MUX
Dtur_1723	GH 3	$\beta$ -glucosidase	MUA, MUG, MUX
Dtur_0276	GH 5	Cellulase	BG
Dtur_0669	GH 5	Cellulase	BG, GM
Dtur_0670	GH 5	Cellulase	AX, BG, GM, HEC, XG
Dtur_0671	GH 5	Cellulase	GM
Dtur_1647	GH 10	Xylanase	AX, ARA, BG, HEC, MUG, MUX
Dtur_1715	GH 10	Xylanase	AX
Dtur_1670	GH 36	$\alpha$ -galactosidase	XAG
Dtur_0505	GH 42	$\beta$ -galactosidase	MUA, MUX
Dtur_1729	GH 43	$\alpha$ -arabinase	ARA
Dtur_0857	GH 53	$\beta$ -galactanase	MUA, XAG
Dtur_0675	GH 57	$\alpha$ -amylase	PUL
Dtur_1714	GH 67	$\alpha$ -glucuronidase	Xylan, xylooligosaccharides

Legend: MUC, 4-methylumbelliferyl- $\beta$ -D-cellobioside; MUX, 4-methylumbelliferyl- $\beta$ -D-xylopyranoside; MUG, 4-methylumbelliferyl- $\beta$ -D-glucopyranoside; MUA, 4-methylumbelliferyl- $\alpha$ -D-arabinofuranoside; XAG, 5-Bromo-4-chloro-3-indolyl  $\alpha$ -D-galactopyranoside; XG, 5-Bromo-4-chloro-3-indolyl  $\beta$ -D-galactopyranoside; AR, AZCL-arabinan; AX, AZCL-arabinoxylan; BG, AZCL- $\beta$ -glucan; GL, AZCL-galactan; GM, AZCL-galactomannan; HEC, AZCL-hydroxyethyl cellulose; PUL, AZCL-pullulan; XG, AZCL-xyloglucan.

The four GH5 family annotated cellulases show a wide range of activities. Three of the GH5 family members hydrolyze a limited number of substrates. Dtur\_0276 possesses only  $\beta$ -glucanase activity, while Dtur\_0671 possesses only  $\beta$ -mannanase activity. Dtur\_0669 possesses both  $\beta$ -mannanase and  $\beta$ -glucanase activities. None of the three possess  $\beta$ -glucosidase,  $\beta$ -cellobiosidase,  $\beta$ -galactosidase, or  $\beta$ -xylosidase activity. In contrast to these three enzymes, Dtur\_0670 (Dtur CelA) possesses both endo-activity and exo-activity on a wide range of substrates (Table 4). Dtur\_0670 possesses endoglucanase activity on a number of insoluble chromogenic substrates including AZCL-HE cellulose, AZCL- $\beta$ -glucan, and AZCL-xyloglucan, endomannanase activity on AZCL-glucomannan, endoxylanase activity on AZCL-arabinoxylan, as well as  $\beta$ -glucosidase and  $\beta$ -cellobiosidase activity (Brumm et al., 2011). None of the GH5 family members released physiologically relevant amounts of sugar from crystalline cellulose even under prolonged incubation.

The two GH10 family xylanases show significantly different activities. Dtur\_1715 possesses only endoxylanase activity, with no other detectable endo- or exo-activities. Dtur\_1647 (XynA) displays endo-activity on  $\beta$ -(1,4)-linked pentose substrates such as xylan, arabinoxylan, and linear arabinan and  $\beta$ -(1,4)-linked hexose substrates such as  $\beta$ -glucan and hydroxyethyl cellulose. XynA also possesses  $\beta$ -glucosidase,  $\beta$ -xylosidase and  $\beta$ -cellobiosidase activity (Table 4).

The GH42 family member, Dtur\_0857 predicted to be a  $\beta$ -galactanase, possesses no endo-activity, but instead possesses

$\beta$ -galactosidase and  $\beta$ -arabinofuranosidase activities. The GH43 family member, Dtur\_1729, does not possess  $\beta$ -xylosidase and  $\beta$ -arabinofuranosidase activity as expected from the annotation, but instead possesses only *endo*- $\beta$ -arabinase activity. The GH57 family member, Dtur\_0675, possesses  $\alpha$ -amylase activity as predicted. The GH67 family member, Dtur\_1714, shows strong  $\alpha$ -glucuronidase activity on xylan and xylan oligosaccharides as predicted (Gao et al., 2011). Comparison to structural orthologs indicate that all 16 enzymes possess a single active site, with the differences in substrate range being a function of active site accessibility for each enzyme. There is no evidence of multiple active sites in any of the enzymes examined. The properties of these 16 enzymes show *D. turgidum* possesses the enzymes capable of hydrolyzing arabinoxylan, arabinan and arabinogalactan,  $\beta$ -glucan, mannan, and glucomannan to usable sugars. Failure to obtain active pectinase clones prevented us from confirming the ability of the organism to utilize pectin.

## Energy Generation

Dtur is predicted to utilize the Embden–Meyerhof–Parnas pathway to produce ATP, reducing equivalents, and fermentation products from monosaccharides. Predicted products from pyruvate include lactate (Dtur\_0700), acetate via acetyl-CoenzymeA (acetyl-CoA) (Dtur\_0260; Dtur\_0261, and Dtur\_0262), and ethanol via acetyl CoA (Dtur\_0260; Dtur\_0261, and Dtur\_0262) and acetaldehyde (Dtur\_0484 and Dtur\_1632). Hydrogen production is predicted by the presence of three hydrogenase gene clusters. The annotation reveals a partial *hypA* operon (Dtur\_0074 through Dtur\_0080) upstream of a hydrogenase assembly cluster (Dtur\_0086 through Dtur\_0090). Additional hydrogenase genes are located in a downstream cluster (Dtur\_0556 through Dtur\_0561). These metabolic predictions are in agreement with the published microbiological studies (Svetlichny and Svetlichnaya, 1988) showing the organism produces acetate, lactate, ethanol, CO<sub>2</sub>, and H<sub>2</sub> during fermentation on sugars.

Based on the genome annotation, Dtur possesses an incomplete, reductive TCA cycle. This cycle allows the organism to convert acetate to pyruvate, oxaloacetate, and eventually to  $\alpha$ -ketoglutarate. The  $\alpha$ -ketoglutarate generated in this pathway can then be utilized for production of glutamate and other amino acids. Based on the annotation, the organism is unable to synthesize citrate from oxaloacetate and acetyl CoA.

Dtur has an extremely simple respiratory system. The genome codes for no respiratory cytochromes. The pathways for production of aminolevulinic acid from either glycine and succinyl-CoA or glutamate are both absent in Dtur. No tetrapyrroles (hemes, sirohemes, or corrinoids) are synthesized by Dtur, and the organism has no ABC-type heme transporters to utilize exogenous heme. The organism is also lacking the pathway for ubiquinone biosynthesis, indicating either ubiquinone is scavenged from the environment, or an alternate electron acceptor is utilized, like ferridoxin (Dtur\_0730) or ferredoxin-like proteins (Dtur\_0076; Dtur\_0457; Dtur\_0556; Dtur\_0730; Dtur\_0774, and Dtur\_1717). The proton gradient needed for ATP generation is produced by NADH oxidoreductase (Dtur\_0558; Dtur\_0559; Dtur\_0916; Dtur\_0919,

and Dtur\_1091), and succinate dehydrogenase (Dtur\_0445). ATP is generated by proton translocation via an F<sub>0</sub>F<sub>1</sub>-type ATP synthase (Dtur\_129 through Dtur\_135). *D. turgidum* also possesses a V-type ATP synthase (Dtur\_1499 through Dtur\_1506), the function of which is unclear. The V-type ATP synthase may also be used for ATP generation, or may it may hydrolyze ATP to generate proton or ion gradients for transport.

## Other Metabolic Pathways

As expected from its small genome size, *D. turgidum* does not possess a full set of biosynthetic and metabolic capabilities. *D. turgidum* appears able to synthesize all 20 amino acids from carbohydrate precursors, but as mentioned previously, is unable to metabolize the majority back to carbohydrates. Like the amino acid situation, *D. turgidum* is able to synthesize fatty acids, but not degrade them. The organism appears to synthesize folate, pyridoxal 5'-phosphate, thiamine, NAD from aspartate, and ascorbate from glucose or galactose, but is lacking pathways for biosynthesis of biotin, pantothenate or flavins.

Carbon monoxide is utilized by strict anaerobes via the Wood-Ljungdahl pathway (Techtmann et al., 2009), using the anaerobic CO dehydrogenase/acetyl CoA synthase complex. This complex catalyzes the complex multistep anaerobic reactions that include oxidizing CO to CO<sub>2</sub>, formation of H<sub>2</sub>, and biosynthesis of acetyl CoA. This pathway is found in thermophilic anaerobes such as *T. tengcongensis* and *M. thermoacetica* as well as in two *G. thermoglucosidarius* species. Manual curation of the genome indicates that *D. turgidum* does not possess the Wood-Ljungdahl pathway and is unable to utilize carbon monoxide as a carbon and energy source.

## DNA Replication, Recombination, and Repair

IMG (DOE JGI) annotation methods identified 61 COG functional category L members (Table 1) for *D. turdigum*. Manual reannotation of each L category gene uncovered six mis-annotated genes. Five genes annotated as excinuclease ATPase subunits (Dtur\_0247, 1011, 1053, 1153, 1667) are more likely ABC type transporters and an endonuclease (Dtur\_0036) has supporting evidence to be annotated as a xylose isomerase. These genes were removed from the compilation shown in Table 5. A complete review of the annotated genes for *D. turdigum* identified a number of missed and overlooked genes that properly belong in the L COG family which totals 85 members in our revised tabulation of the genome (Table 5). Even though the 16S rRNA genes of the only two described species of *Dictyoglomus*, *D. turgidum* (CP001251) and *D. thermophilum* (CP001146), share 99% sequence identity, the divergence of their orthologous replication proteins is significant, as described below.

The genome of *D. turdigum* possesses 85 annotated genes for DNA replication, recombination and repair (Table 5), 78 of which have their closest ortholog to Diclth genes, whereas 6 have no orthologs in Diclth (Dtur\_0102, 0317, 0545, 1284, 1294, 1297). The last four genes are part of a prophage that is not present in *D. thermophilum*. There are two genes in this category that are present in *D. thermophilum* but not *turdigum*, including a deoxyribodipyrimidine photolyase (Diclth\_0072)

**TABLE 5 | *D. turgidum* annotated DNA replication, recombination, and repair enzymes.**

Dtur gene	Annotation	Nearest neighbor/ % identity
0745	ATP-dependent DNA helicase	<i>Thermoanaerobaculum aquaticum</i> /43
1397	ATP-dependent DNA helicase PcrA	<i>Caloranaerobacter azorensis</i> /49
1140	ATP-dependent DNA helicase RecG	<i>Halothermothrix orenii</i> /52
0780	ATP-dependent DNA ligase	uncultured <i>Acidobacteria</i> /56
1514	Bacterial nucleoid DNA-binding protein	<i>Symbiobacterium thermophilum</i> /66
1151	Cell division protein FtsK	<i>Thermoanaerobacterium aotearoense</i> /41
0001	Chromosomal replication initiator protein DnaA	<i>Pelotomaculum thermopropioni</i> /59
1468	Chromosome segregation and condensation protein ScpA	<i>Caldsalinibacter Kiritimati</i> /28
1467	Chromosome segregation and condensation protein ScpB	<i>Dethiobacter alkaliphilus</i> /45
0947	Chromosome segregation protein SMC	<i>Caloranaerobacter azorensis</i> /27
1104	Competence protein ComE	<i>Clostridium purinilyticum</i> /29
1103	Competence protein ComEA	<i>Clostridium</i> sp. CAG:273/55
1614	Crossover junction endodeoxyribonuclease RuvC	<i>Alkaliphilus oremlandii</i> /50
0771	Deoxyinosine 3' endonuclease (endonuclease V)	bacterium JGI-24/54
1268	DNA and RNA Helicase	candidate division <i>Zixibacteria</i> /57
1547	DNA gyrase B subunit	<i>Thermoanaerobacter</i> sp. YS13/66
1263	DNA gyrase, A subunit	<i>Mahella australiensis</i> /54
1478	DNA integrity scanning protein DisA	<i>Bacillus</i> sp. CMAA 1185/52
1078	DNA mismatch repair protein MutL	<i>Caldanaerobacter subterraneus</i> /42
1077	DNA mismatch repair protein MutS	<i>Clostridium thermocellum</i> /46
0884	DNA or RNA helicase of superfamily II (UvrB)	<i>Caldanaerobacter subterraneus</i> /68
0104	DNA polymerase beta domain-containing protein	<i>Syntrophaceticus schinkii</i> /57
0317	DNA polymerase beta domain-containing protein	marine sediment metagenome/38
0545	DNA polymerase beta domain-containing protein	<i>Microgenomates</i> bacterium/39
1295	DNA polymerase beta domain-containing protein*	<i>Chloracidobacterium thermophilum</i> /60
0882	DNA polymerase I 3'-5' exonuclease & polymerase domains	<i>Tepidanaerobacter acetatoxydans</i> /44
1391	DNA polymerase III alpha subunit	<i>Caloranaerobacter</i> sp. TR13/52
1105	DNA polymerase III delta subunit	<i>Alkaliphilus oremlandii</i> /29
0257	DNA polymerase III gamma/tau subunits	<i>Thermincola potens</i> /41
0789	DNA polymerase III gamma/tau subunits	marine sediment metagenome/39

(Continued)

**TABLE 5 | Continued**

Dtur gene	Annotation	Nearest neighbor/ % identity
1551	DNA polymerase III sliding clamp subunit beta	<i>Clostridium acidurici</i> /37
1600	DNA Polymerase X	<i>Caldisericum exile</i> /55
1316	DNA primase N	<i>Lachnospiraceae</i> bacterium/38
1527	DNA protecting protein DprA	<i>Caldanaerobacter subterraneus</i> /43
1549	DNA recombination protein RecF	<i>Clostridium aceticum</i> /36
1625	DNA repair exonuclease	<i>Thermobaculum terrenum</i> /33
0327	DNA repair photolyase	<i>Parcubacteria</i> bacterium/41
0463	DNA repair photolyase	<i>Caldicellulosiruptor owensensis</i> /66
1479	DNA repair protein RadA	<i>Thermoanaerobacterium xylanolyticum</i> /48
0881	DNA repair protein RADC	<i>Paenibacillus</i> sp./50
1047	DNA repair protein RecN	<i>Caloramator australicus</i> /42
1321	DNA repair protein RecO	<i>Microgenomates</i> bacterium/29
0015	DNA replication and repair protein RecF	<i>Fervidobacterium nodosum</i> /80
1526	DNA topoisomerase type I	<i>Carboxydotherrmus hydrogenoformans</i> /54
1522	Double-stranded DNA repair protein Rad50	<i>Thermoanaerobacter siderophilus</i> /26
1626	Double-stranded DNA repair protein Rad50	<i>Thermofilum</i> sp./25
0264	Endonuclease IV	<i>Thermincola potens</i> /36
1485	Excinuclease ABC subunit C	<i>Acetohalobium arabaticum</i> /45
0885	Excinuclease ATPase subunit (ABC-ATPase UvrA)	<i>Thermosediminibacter oceani</i> /64
1613	Holliday junction DNA helicase RuvA	<i>Thermodesulfobivrio yellowstonii</i> /34
1612	Holliday junction DNA helicase RuvB	<i>Caldicellulosiruptor obsidiansis</i> /54
0792	Holliday junction resolvosome helicase subunit	<i>Carboxydotherrmus hydrogenoformans</i> /54
1284	Integrase*	<i>Acetothermus autotrophicum</i> /35
0886	Methylated DNA-protein cysteine methyltransferase	<i>Anaerococcus lactolyticus</i> /40
1227	Mg-dependent Dnase—deoxyribonuclease TatD	<i>Bacillus</i> sp. SA1-12/47
1308	Mismatch repair ATPase (MutS family)	<i>Peptococcaceae</i> bacterium/40
1294	Modification methylase, type III R/M system*	<i>Sulfurihydrogenibium yellowstonense</i> /65
0341	N6-adenine-specific methylase	<i>Caloranaerobacter azorensis</i> /65
1141	N6-adenine-specific methylase	<i>Thermoanaerobacter ethanolicus</i> /44
1497	Poly(A) polymerase	<i>Peptococcaceae</i> bacterium/36
0683	Predicted EndoIII-related endonuclease	<i>Thermodesulfatator indicus</i> /58
0846	Predicted EndoIII-related endonuclease	<i>Candidatus Methanoperedens</i> sp./32
1024	Predicted endonuclease involved in recombination	<i>Veillonella atypica</i> /44

(Continued)

TABLE 5 | Continued

Dtur gene	Annotation	Nearest neighbor/ % identity
1530	Predicted endonuclease related to Holliday junction resolvase	<i>Caldicellulosiruptor saccharolyticus</i> /42
1709	Predicted exonuclease	marine sediment metagenome/61
0102	Predicted nucleic acid-binding protein contains PIN domain	<i>Leptospira wolbachii</i> /47
1441	Primosomal protein N' (replication factor Y)	human gut metagenome/29
1507	Putative chromosome partitioning protein	<i>Plasmodium knowlesi</i> /38
1162	RecA/RadA recombinase	<i>Thermosediminibacter oceani</i> /73
1524	Recombinase XerD	<i>Symbiobacterium thermophilum</i> /47
0259	Recombination protein RecR	<i>Oxobacter pfennigii</i> /58
1358	Replicative DNA helicase	<i>Acetohalobium arabaticum</i> /56
0014	Reverse gyrase	<i>Fervidobacterium nodosum</i> /79
0708	Ribonuclease HII	<i>Kosmotoga pacifica</i> /35
1531	Ribonuclease HII	<i>Clostridium</i> sp./59
0118	Single stranded nucleic acid binding protein (R3H domain)	<i>Butyrivibrio</i> sp./45
1602	Single-stranded DNA exonuclease RecJ	<i>Clostridium thermocellum</i> /36
1362	Single-stranded DNA-binding protein	<i>Syntrophothermus lipocalidus</i> /52
0786	Single-stranded nucleic acid binding protein (R3H domain)	<i>Syntrophobacter fumaroxidans</i> /59
0880	Smc1 chromosome segregation protein, putative	marine sediment metagenome/28
0202	Thermostable 8-oxoguanine DNA glycosylase	<i>Thermotoga maritima</i> /59
1511	Transcription-repair coupling factor	marine sediment metagenome/40
1297	Type III restriction endonuclease subunit R*	bacterium JGI-6/67
1473	Tyrosine recombinase XerD	<i>Tepidanaerobacter acetatoxydans</i> /51
1393	Uracil-DNA glycosylase	<i>Nitrospira moscoviensis</i> /54

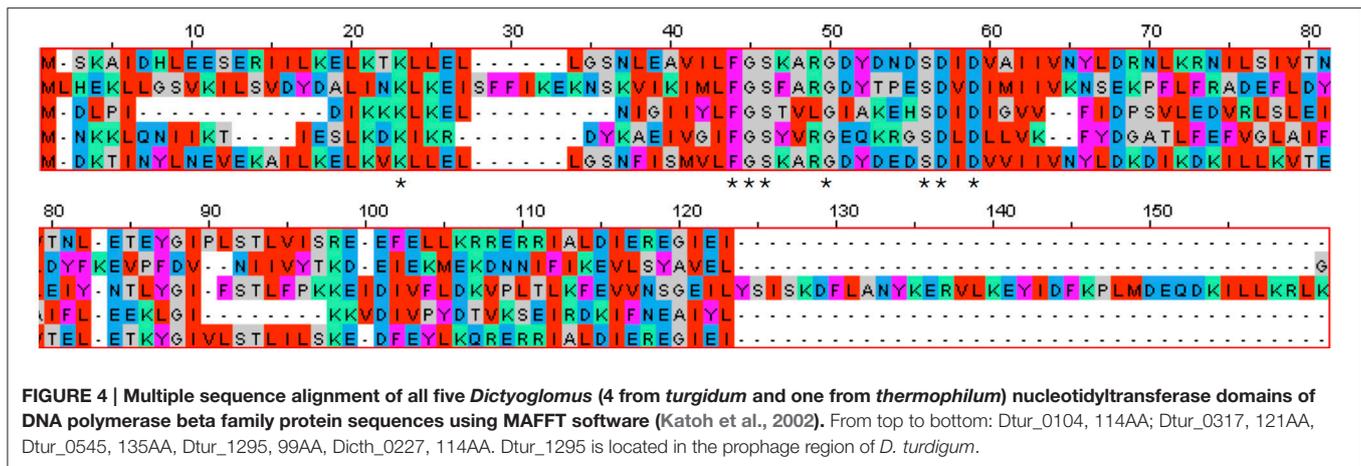
Prophage related genes are marked\*.

and a DNA modification methylase (Dichth\_0253). Only one gene (Dtur\_1514), annotated as a bacterial nucleoid DNA-binding protein has 100% identity to its Dichth ortholog. On average the genes in this COG category share 85% identity with their Dichth counterparts, with the lowest at 66% identity (Dtur\_1626, double-stranded DNA repair protein Rad50). Based on blastP analysis the most striking feature of this class of genes is how dissimilar they are to other sequenced genes and genomes in the database, other than *D. thermophilum*. This is apparent when the next nearest neighbors to *D. turgidum* DNA replication proteins are tabulated (Table 5). On average the genes in this COG category share 48% amino acid identity to nearest neighbor genes (25–80% range) and the cross section of homology is widespread among taxa that are primarily

anaerobic, thermophilic, or halophilic. Only two of the genes share homology to Thermotogae (Dtur\_0202 and 0708) and 3 to *Caldicellulosiruptor* (Dtur\_0463, 1530, and 1612). The greatest frequency of nearest neighbor orthologs after *D. thermophilum* are to clostridial (9%) and *Thermoanaerobacterium* (8%) genus members, followed by unknown metagenomic genes (6%). The low homology of *Dictyoglomus* replication proteins to orthologs in other organisms is another testament to how phylogenetical unique the genus is.

In spite of its preferred growth temperature of 72°C (Svetlichny and Svetlichnaya, 1988), *D. turgidum* has an extremely low (33.96 mol%) G+C content, which seems counterintuitive to genome stability and repair (Ishino and Narumi, 2015). Dtur does possess a reverse gyrase (Dtur\_0014), a hallmark enzyme that is systematically present in all hyperthermophiles (Brochier-Armanet and Forterre, 2007), which introduces positive supercoils in DNA and thereby protects it from unwinding. Dtur and Dichth do not appear to contain genes for exonuclease III, a DNA-repair enzyme that hydrolyzes the phosphodiester bond 5' to an abasic site in DNA, which is commonly induced by heat. However, they both possess endonuclease IV, which has been shown to perform a similar abasic site processing function in *Thermotoga maritima* (Haas et al., 1999).

*D. turgidum* possesses 7 DNA replication and repair genes annotated to contain a nucleotidyltransferase (NT) domain, a superfamily that includes DNA polymerase beta domain-containing proteins (NT\_Pol-beta), family X and poly-A DNA polymerases, as well as other proteins (Aravind and Koonin, 1999). The majority of the NTs are characterized by a distinct amino acid residue pattern, namely hG[GS]x(9,13)Dh[DE]h (x indicates any amino acid and h indicates a hydrophobic amino acid) that are essential for catalysis, which is true for all 7 NT domain containing genes in *D. turgidum*. Three of the *D. turgidum* NT domain-containing DNA repair proteins are larger than 450 amino acids (Dtur\_0257, 1497, and 1600), whereas four members further annotated as belonging to the Pol-beta subfamily only encode 99–135 amino acids (Figure 4). DNA polymerase B is a proofreading-proficient enzyme thought to be involved with DNA repair activities in eubacteria (Wijffels et al., 2005) and replication in archaea (Kelman and Kelman, 2014), however there are no known thermophilic bacterial DNA polymerase B genes. A new, ancestral family of polB type nucleotidyltransferases designated as MNT (minimal nucleotidyltransferases) has been described (Aravind and Koonin, 1999) that are one-half to one-third the size of the larger orthologs. They are not uncommon as 258 cases can be found in the protein NCBI database “dna polymerase beta domain-containing protein” in 129 different microbes as of January 2016. However, there are no known biochemical studies showing whether these diminutive NTs are catalytically functional monomeric enzymes or whether they are part of a larger multimeric complex. The four NT\_Pol-beta genes found in *D. turgidum*, one of which is associated with the prophage element discussed below, is an unsolved mystery as to the function these diminutive proteins might play,



particularly with regard to the lack of a PolB enzyme in this hyperthermophile.

## Functional Analysis of *D. turgidum* DNA Polymerase I

*D. turgidum* possesses four sets of DNA polymerases: Pol X (Dtur\_1600), Poly(A) polymerase (Dtur\_1497), DNA polymerase I (Dtur\_0882), and a minimal DNA polymerase III set of subunits (alpha, Dtur\_1391; beta, Dtur\_1551; delta, Dtur\_1105, and gamma/tau, Dtur\_0257 and \_0789). In a survey looking for thermostable reverse transcriptase (RT) activity, the DNA polymerase I (PolI) from *Dictyoglomus thermophilum* strain Rt46B.1 has been cloned and expressed (Shandilya et al., 2004). While the enzyme did not exhibit RT activity, it did show significant thermal stability at 85°C compared to eight other enzymes being studied. Presumably its ortholog behaves the same, as Dicho\_0729 shares 90% identity (770/856) and 96% positives (826/856) at the amino acid level with Dtur\_0882.

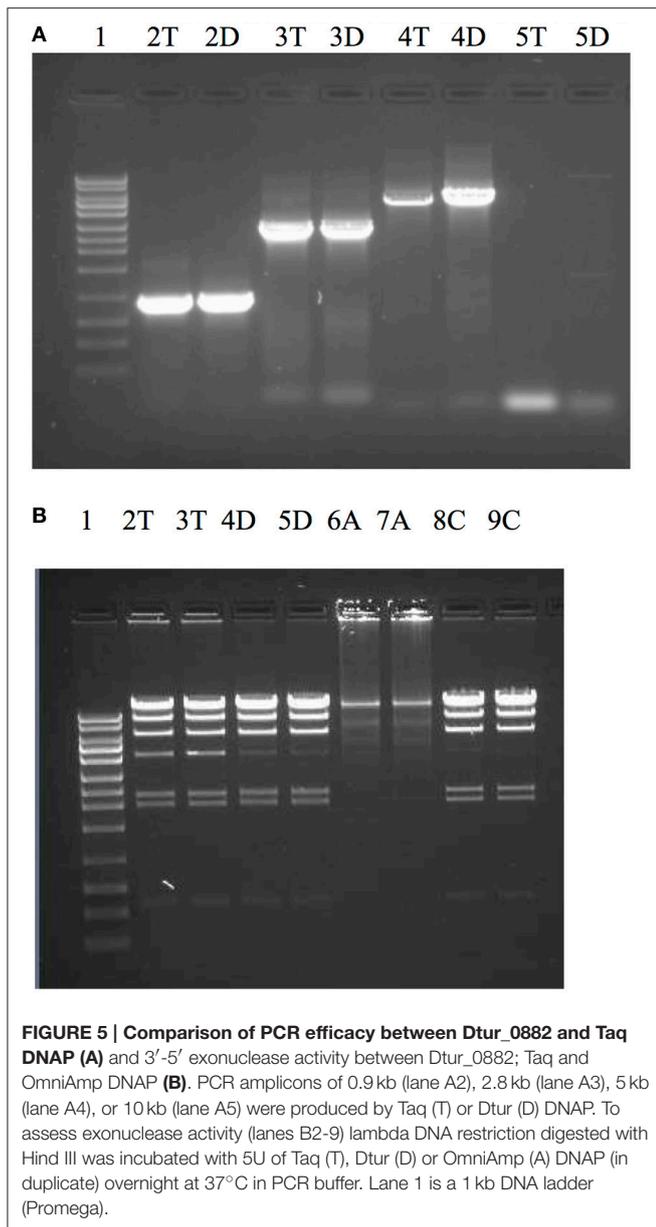
Dtur\_0882 is a PolA type polymerase annotated to contain a 5'-3' and 3'-5' exonuclease domain in addition to the DNA-directed DNA polymerase domain. With the exception of *Rhodothermus marinus* PolA (PMID:11483153) and the enzyme OmniAmp polymerase (Chander et al., 2014) every thermostable PolA enzyme characterized thus far lacks a functional 3'-5' exonuclease domain associated with proofreading and enzyme fidelity. Because the high temperature growth conditions for *D. turgidum* are very similar to *Thermus aquaticus* (Taq), but the amino acid identities are so different between the two PolI enzymes (41%, 351/847), the utility of Dtur\_0882 as a PCR enzyme was evaluated and compared with Taq DNAP (Figure 5). The 3'-5' exonuclease activity of both enzymes were also compared with the thermostable proof reading PolA enzyme called OmniAmp polymerase (Chander et al., 2014). Dtur\_0882 produced the same yield of amplicon for the 0.9 and 2.8 kb reactions as Taq DNAP (Figure 5A lanes 2, 3), but was more efficient at amplifying the 5 and 10 kb primer templates (Figure 5A lanes 4, 5). As with Taq DNAP, Dtur\_0882 does not appear to have any measurable 3'-5' exonuclease activity (Figure 5B lanes 2/3 compared to 4/5) as opposed to a strong

exonuclease activity demonstrated by the proofreader OmniAmp DNAP (Figure 5 lanes 6/7).

## Prophage and CRISPR Elements

*D. turgidum* possesses two regions containing CRISPR repeats, suggesting the previous exposure to phage(s). The first CRISPR region, located between 470870 and 474424 nucleotides in the genome codes for 54 repeats. The repeat sequence is 30 nucleotides long, and the spacer average length is 36 nucleotides long. The second CRISPR region, located between nucleotides 6151530 and 617311 in the genome codes for 33 repeats. The repeat sequence is the same as the first CRISPR region, and the spacer average length is 36 nucleotides long. No CRISPR-associated proteins are in the vicinity of the first CRISPR region. Upstream of the second CRISPR region are eight CRISPR-associated proteins (Table 5), while downstream of the second CRISPR region is a four-gene insert coding for biosynthetic enzymes (Dtur\_0614—Dtur\_0617) and eight additional CRISPR-associated proteins. *D. thermophilum* shows a similar organization of its CRISPR-associated proteins (Table 6), with a larger, 15-gene insert coding for biosynthetic enzymes (not shown).

An incomplete prophage sequence was identified by PHAST (Hubisz et al., 2011; Zhou et al., 2011) as a 17,407 base insert from 1,301,419 to 1,318,825 (Dtur\_1284-1300) and confirmed as foreign DNA by IslandViewer Software (Hsiao et al., 2003) (1,298,422 to 1,317,048 containing genes Dtur\_1282 through Dtur\_1297). The prophage contains an integrase (Dtur\_1284) followed by four annotated putative lipoproteins that potentially form part of a beta-barrel assembly machinery (Dtur\_1285; Dtur\_1286; Dtur\_1288, and Dtur\_1289) and a Type III restriction system (Dtur\_1294 and Dtur\_1297). The four annotated lipoprotein genes are closely related, as they share 79–88% amino acid identity. Inspection of the amino acid sequence reveals a unique periodicity of aspartic (D) and glutamic (E) residues to hydrophobic residues. The reason for this novel periodicity is due to the seven back to back, nearly perfect 49–52 amino acid tandem repeats found in these proteins (data not shown). Tandem repeat proteins are ubiquitous (Jernigan and



Bordenstein, 2015), but the unique signature sequence found here is only partially common to a handful of hypothetical proteins found in bacteria. Additional work is needed to clarify the function of these four repeated proteins with seven tandem internal repeats. This genomic island is unique to *D. turgidum*. *D. thermophilum* possesses an ortholog to only one of the four lipoproteins and no Type III restriction system proteins.

### Thermophily, Stress Responses, and Heat Shock Proteins

*D. turgidum* in common with *D. thermophilum* has a complement of heat shock proteins typical of thermophilic bacteria, including a single GroEL/ES locus encoding Hsp60 and a DnaK/DnaJ locus encoding Hsp70 and cochaperones GrpE. Interestingly, in both

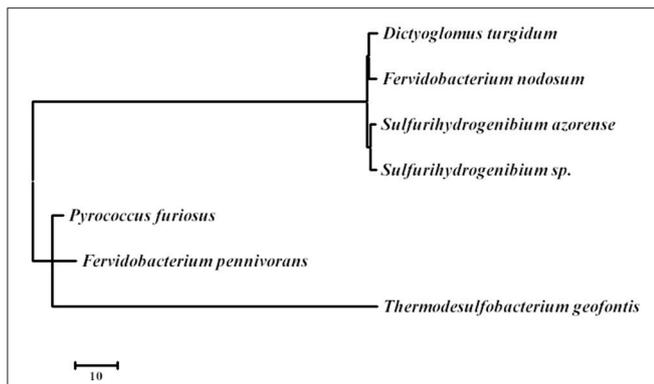
**TABLE 6 | *D. turgidum* CRISPR-associated proteins.**

Gene	Annotation	Orthologs
Dtur_0606	CRISPR-associated protein, TM1812 family	Dicth_0458
Dtur_0607	CRISPR-associated protein DxTHG motif protein	Dicth_0187
Dtur_0608	CRISPR-associated RAMP protein, Csm5 family	Dicth_0186
Dtur_0609	CRISPR-associated RAMP protein, Csm4 family	Dicth_0185
Dtur_0610	CRISPR-associated RAMP protein, Csm3 family	Dicth_0184
Dtur_0611	CRISPR-associated RAMP protein, Csm2 family	Dicth_0183
Dtur_0612	CRISPR-associated RAMP protein, Csm1 family	Dicth_0182
Dtur_0613	CRISPR-associated protein, Csx3 family	Dicth_0181
Dtur_0618	CRISPR-associated protein, Cas2 family	Dicth_0165
Dtur_0619	CRISPR-associated protein, Cas1 family	Dicth_0164
Dtur_0620	CRISPR-associated endonuclease, Cas4 family	Dicth_0163
Dtur_0621	CRISPR-associated helicase, Cas3 family	Dicth_0162
Dtur_0622	CRISPR-associated protein, Cas5h family	Dicth_0161
Dtur_0623	CRISPR-associated protein, Csh2 family	Dicth_0160
Dtur_0624	CRISPR-associated protein, Csh1 family	Dicth_0159
Dtur_0625	CRISPR-associated protein, Cas6 family	Dicth_0158

*D. turgidum* and *D. thermophilum*, the reverse gyrase is encoded in a gene cluster shared with recJ and revG, closely linked to the DnaK/DnaJ operon. Heat shock regulation is enigmatic since the genome lacks both CIRCE elements and sigma32 SOS regulation. It is tempting to speculate that conditional expression of the reverse gyrase under high temperature growth conditions might be a mechanism for regulating DNA positive supercoiling in concert with the heat shock response. The phylogeny of the reverse gyrase is extraordinary. The reverse gyrase is most closely related to orthologs in *Fervidobacterium* species as shown in Figure 6. This phylogenetic position of the reverse gyrase does not conform to the 16S rRNA phylogeny (Figure 3) where *D. turgidum* is most closely related to *Caldicellulosiruptor* species. This suggests that lateral gene transfer of the reverse gyrase may have taken place.

### Morphological Characteristics

Microbiological testing of *D. turgidum* indicate the organism stains Gram-negative (Svetlichny and Svetlichnaya, 1988). The *D. turgidum* genome contains a cluster of 10 genes potentially coding for outer membrane proteins characteristic of Gram-negative organisms and outer membrane lipid biosynthesis (Dtur\_0814 through Dtur\_0824). This cluster contains genes coding for a TamB (Dtur\_0814), two BamA orthologs (Dtur\_0815 and Dtur\_0816), and two outer membrane chaperone Skp (OmpH) orthologs (Dtur\_0817 and Dtur\_0818), all potentially involved in outer membrane transport and assembly. This cluster of genes provides genomic support for the observed Gram-negative membrane structure observed in electron micrographs (Svetlichny and Svetlichnaya, 1988). Following these five genes, the *D. turgidum* genome contains a lipid biosynthesis cluster coding for the first four enzymes of Lipid A biosynthesis, LpxD (Dtur\_0819), LpxC (Dtur\_0820), LpxA (Dtur\_0821), and (Dtur\_0823) and an ortholog of FabZ (beta-hydroxyacyl-(acyl-carrier-protein) dehydratase,



**FIGURE 6 | Evolutionary relationships of reverse gyrase.** The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Tamura et al., 2011). The tree with the highest log likelihood (-22290.6302) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 7 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 1122 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016). UniProt sequences used for the analysis were: B8DYH3, *Dictyoglomus turgidum* strain DSM 6724; A7HMS7, *Fervidobacterium nodosum* strain DSM 5306; H9UDK4, *Fervidobacterium pennivorans* strain DSM 9078; C1DT23, *Sulfurihydrogenibium azorense* strain DSM 15241; B2V6S9, *Sulfurihydrogenibium* sp. strain YO3AOP; F8C2X1, *Thermodesulfobacterium geofontis* strain OPF15, and P95479; *Pyrococcus furiosus* strain DSM 3638.

Dtur\_0821). The final gene in the cluster (Dtur\_0824) is a hypothetical protein related to LpxB. This 10-gene cluster has the identical organization in *D. thermophilum*, with individual genes averaging 80% identity to its Dtur ortholog. The cluster appears to be unique to *Dictyoglomus* species, as no similar cluster is found in any other sequenced organisms. The individual genes also show little homology to orthologs in other species, with observed amino acid identities being  $\leq 30\%$  for all genes in the cluster.

Downstream of this ten-gene cluster is an annotated cluster of six proteins potentially involved in Gram-negative outer membrane efflux including orthologs of an ABC transporter ATP-binding protein (Dtur\_0835), two of TolC (Dtur\_0836 and Dtur\_0837), HlyD (Dtur\_0838) an ABC transporter permease protein (Dtur\_0839) and a predicted transmembrane protein (Dtur\_0840). An identical cluster is found in *D. thermophilum* (DICTH\_0678 through DICTH\_0683). A similar cluster is found in *Meiothermus taiwanensis* DSM 14542 as well as other *Meiothermus* and *Thermus* species. No orthologs of this cluster are found in any sequenced Thermotogales or Firmicutes species.

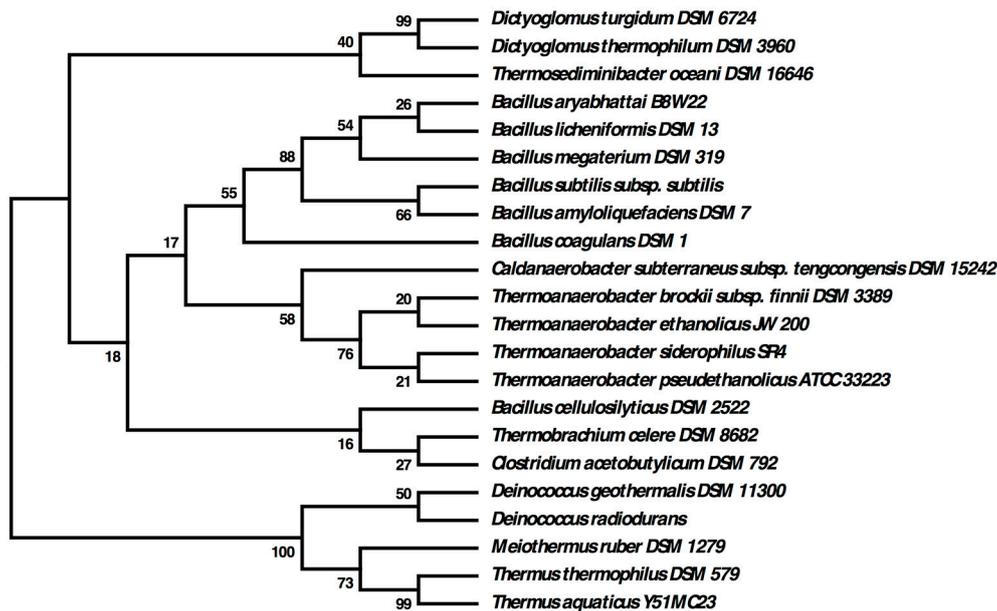
Many thermophilic bacteria possess complex morphologies, with varying shapes seen under different growth conditions. Examples of this include “rotund bodies” in *Thermus aquaticus* (Brumm P. J. et al., 2015), the outer membrane “toga” of *Thermotoga maritima* (Huber et al., 1986) and the multicellular spheres of *D. turgidum* (Svetlichny and Svetlichnaya, 1988). Regulation of these morphologies may be controlled by the action

of SpoVS (Brumm P. J. et al., 2015). *D. turgidum* possesses a gene coding for SpoVS (Rigden and Galperin, 2008) (Dtur\_0800) similar to SpoVS proteins found in sporulating Firmicutes species as well as in the non-sporulating *Thermus-Deinococcus* and *Thermotoga* groups, but not in non-sporulating Firmicutes species. Phylogenetic reconstruction indicates that *D. turgidum* SpoVS is most closely related to the *D. thermophilum* SpoVS, (Figure 7) followed by the SpoVS of *Thermosediminibacter oceani* DSM 16646. These three SpoVS molecules form a separate clade from the SpoVS of the Firmicutes and *Deinococcus/Thermus* species. The SpoVS orthologs show much higher homology than orthologs of other *Dictyoglomus* proteins, suggesting an important conserved function. *Thermotoga* SpoVS orthologs have 53–78% identity, a *Clostridium thermocellum* ortholog has 74% identity, *Caldicellulosiruptor* orthologs have 65–71% identity, and *Thermus* orthologs have 56% identity to Dtur\_0800. SpoVS may be an important regulator of cell morphology and differentiation in both sporulating thermophiles where it regulates the transition from vegetative growth to spore formation as well as the non-sporulating thermophiles where it regulates the transition from vegetative growth to formation of multiple morphologies (Brumm P. J. et al., 2015).

## DISCUSSION

We report here the genome sequence, sequence analysis, and cloning of key enzymes of *D. turgidum*, an anaerobic, thermophile reported to degrade a wide range of biomass components including starch, cellulose, pectin and lignin [14]. This hyperthermophile has a small 1.8 M bp genome with a G+C content of the 34%. COGS analysis shows the organism is enriched in genes coding for carbohydrate transport and metabolism. While *Dictyoglomus* make up 25% of the species identified by 16S rRNA sequencing in some environments currently only two species, *D. thermophilum* (Saiki et al., 1985) and *D. turgidus* (Svetlichny and Svetlichnaya, 1988), corrected to *D. turgidum.*, *D. turgidum*, and *D. thermophilum*, have been sequenced and annotated. This first comparison of the two genomes shows that, while the two organisms are unique species, they show extremely high levels of orthologous genes, average nucleotide identity, and synteny. No unique metabolic pathways are present in either organism. Approximately 1/3 of the proteins present in Dtur have orthologs in *D. thermophilum* with over 90% amino acid identity, and less than 10% of the proteins present in either genome have no ortholog in the other genome. The two organisms show extensive short-range and long-range synteny. Genome sequences of additional *Dictyoglomus* species are needed to determine if this is coincidence or a conserved feature of the *Dictyoglomi*. Additional work is also needed to confirm that the differences in synteny observed between the two genomes are real and are not artifacts of the assembly of the genomes.

The genome of *D. turgidum* provides insights into an organism that is strangely foreign and vaguely familiar at the same time. At first glance, the genome is remarkably unremarkable, containing no novel pathways or secondary products. In fact, the organism is lacking many of the pathways



**FIGURE 7 | Evolutionary relationships of SpoVS proteins.** The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model. The bootstrap consensus tree inferred from 550 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (550 replicates) are shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The analysis involved 22 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 86 positions in the final dataset.

normally associated with microbes, including amino acid and fatty acid degradation pathways and energy harvesting via proteins containing hemes, sirohemes, or quinones and appears to be the genome of a strict carbohydrate fermentor. Yet, at the same time, *D. turgidum* cannot be identified as similar to any one organism or Phylum. Results presented here and elsewhere (Nishida et al., 2011; Vesth et al., 2013) show the chameleon-like nature of the organism. Changing the method of comparison radically changes the resulting relationships between *D. turgidum* and other organisms.

The metabolic reconstruction based on the *D. turgidum* genome reveals two unusual features. Most evident is the importance of carbohydrate metabolism for the organism, because *D. turgidum* lacks the ability to metabolize fatty acids and most amino acids. The genome, while lacking in enzymes to degrade crystalline cellulose, possesses genes coding for utilization of most other biomass-derived polymers including xylans, glucans, pectins, arabinans and galactans. Utilization of these polysaccharides appears to involve secretion of enzymes that degrade the polysaccharides to oligosaccharides, transport of the oligosaccharides into the cytoplasm, and degradation of the oligosaccharides to monosaccharides in the cytoplasm. A similar strategy is utilized by thermophilic *Geobacillus* species (Brumm P. et al., 2015). Genes for utilization of carbohydrates are distributed randomly throughout the *D. turgidum* genome, unlike the *Geobacillus* genomes, where genes for individual polysaccharide degradation pathways are organized into distinct operons. The second feature is the obligate fermentative

nature of *D. turgidum*. Unlike many other thermophilic anaerobes including *Thermus*, *Geobacillus*, *Caldicellulosiruptor*, and *Thermotoga* species, *D. turgidum* possesses no genes for production or utilization of either cytochromes or quinones. Dtur is predicted to utilize the EMP pathway, to produce ATP, reducing equivalents, and fermentation products from monosaccharides. The predicted fermentation products, lactate, acetate, ethanol and hydrogen are in agreement with the published microbiological studies (Svetlichny and Svetlichnaya, 1988) showing the organism produces these four products during fermentation on sugars. The proton gradient needed for ATP generation is produced by NADH oxidoreductase and succinate dehydrogenase, and the ATP is generated by an F<sub>0</sub>F<sub>1</sub>-type and a V-type ATP synthases.

Sixteen *D. turgidum* carbohydrases were cloned, expressed and characterized to better understand their function in the metabolism of the organism. The 16 included two each of GH1 and GH3, four GH5, two GH10, and one each of GH36, GH42, GH43, GH53, GH57, and GH67. Based on the proposed mechanism for polysaccharide utilization, *D. turgidum* produces oligosaccharides using secreted enzymes, and degrades the oligosaccharides using intracellular enzymes. The secreted enzymes would be expected to have high substrate specificity to generate oligosaccharides recognized by the transporter systems. The cloned enzymes predicted to be secreted showed activity only on one or two substrates, showing activity on xylan, arabinan, *beta*-glucan, starch, or mannan. Conversely, the intracellular enzymes would be expected to have low specificity, allowing

them to degrade multiple substrates and linkages efficiently. The cloned intracellular enzymes typically showed a broader range of activities. GH1 and GH3 enzymes possess *exo*-activity on four or five different carbohydrate substrates. The cloned intracellular xylanase and cellulase both possess both *exo*-activity and *endo*-activity, as well as activity on multiple substrates.

Replication, recombination, and repair enzymes are critical to the genome maintenance and integrity of all cells. Many proteins from this COG functional category are expected to share conserved domains and motifs that could in theory be used to understand the phylogenetic relationship of *D. turgidum* to other organisms. The 16S rRNA genes of *D. turgidum* and *D. thermophilum* share 99% sequence identity. The fraction of replication proteins having 100–90% similarity between the two species is 35%, with 42% sharing 89–80% similarity, 17% sharing 79–70% similarity and 5% with similarity below 69%. The similarity of *D. turgidum* replication proteins to other taxa drops off considerably from that with *D. thermophilum*. The fraction of replication proteins having 100–90% similarity between *D. turgidum* and the next nearest neighbor species is 0%, with 1% sharing 89–80% similarity, 3% sharing 79–70% similarity, 13% sharing 69–60 similarity, 32% sharing 59–50% similarity, 28% sharing 49–40 similarity, and 32% similarity below 40%. This informal comparison again demonstrates how unique *Dictyoglomi* are compared to other species. The number and type of replication proteins found in *D. turgidum* is similar to those found in other hyperthermophilic bacteria using IMG tools

at the Joint Genome Institute (data not shown). The phylogenetic position of the reverse gyrase does not conform to the 16S rRNA phylogeny suggesting that lateral gene transfer may have taken place.

## AUTHOR CONTRIBUTIONS

FR analyzed data and contributed to manuscript preparation. PB wrote the manuscript, produced and purified Dtur proteins, and performed carbohydrate analyses. KG produced and purified Dtur proteins. DM managed genome sequencing and analysis, performed DNAP analyses, and contributed to manuscript preparation.

## FUNDING

This work was completely funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494 and DOE OBP Office of Energy Efficiency and Renewable Energy DE-AC05-76RL01830). FR acknowledges support from the NASA Exobiology Program.

## ACKNOWLEDGMENTS

The authors would like to thank Robb lab and Elizabeth O'Connor for assistance in the growth and preparation of the cells used for genomic DNA isolation.

## REFERENCES

- Aravind, L., and Koonin, E. V. (1999). DNA polymerase  $\beta$ -like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history. *Nucleic Acids Res.* 27, 1609–1618. doi: 10.1093/nar/27.7.1609
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75
- Brochier-Armanet, C., and Forterre, P. (2007). Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers. *Archaea* 2, 83–93. doi: 10.1155/2006/582916
- Brumm, P. J., De Maayer, P., Cowan, D. A., and Mead, D. A. (2015). Genomic analysis of six new Geobacillus strains reveals highly conserved carbohydrate degradation architectures and strategies. *Front. Microbiol.* 6:430. doi: 10.3389/fmicb.2015.00430
- Brumm, P. J. (2013). Bacterial genomes: what they teach us about cellulose degradation. *Biofuels* 4, 669–681. doi: 10.4155/bfs.13.44
- Brumm, P. J., Hermanson, S., Luedtke, J., and Mead, D. A. (2011). Identification, cloning and characterization of *Dictyoglomus turgidum* CelA, an endoglucanase with cellulase and mannanase activity. *J. Life Sci.* 5, 488–496.
- Brumm, P. J., Monsma, S., Keough, B., Jasinovica, S., Ferguson, E., Schoenfeld, T., et al. (2015). Complete genome sequence of thermus aquaticus Y51MC23. *PLoS ONE* 10:e0138674. doi: 10.1371/journal.pone.0138674
- Burgess, E. A., Unrine, J. M., Mills, G. L., Romanek, C. S., and Wiegel, J. (2012). Comparative geochemical and microbiological characterization of two thermal pools in the Uzon Caldera, Kamchatka, Russia. *Microb. Ecol.* 63, 471–489. doi: 10.1007/s00248-011-9979-4
- Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., et al. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 42, D459–D471. doi: 10.1093/nar/gkt1103
- Chander, Y., Koelbl, J., Puckett, J., Moser, M. J., Klingele, A. J., Liles, M. R., et al. (2014). A novel thermostable polymerase for RNA and DNA loop-mediated isothermal amplification (LAMP). *Front. Microbiol.* 5:395. doi: 10.3389/fmicb.2014.00395
- Coil, D. A., Badger, J. H., Forberger, H. C., Riggs, F., Madupu, R., Fedorova, N., et al. (2014). Complete genome sequence of the extreme thermophile *Dictyoglomus thermophilum* H-6-12. *Genome Announc.* 2:e00109-14. doi: 10.1128/genomeA.00109-14
- Dehal, P. S., Joachimiak, M. P., Price, M. N., Bates, J. T., Baumohl, J. K., Chivian, D., et al. (2010). MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res.* 38, D396–D400. doi: 10.1093/nar/gkp919
- Delcher, A. L., Salzberg, S. L., and Phillippy, A. M. (2003). Using MUMmer to identify similar regions in large sequence sets. *Curr. Protoc. Bioinformatics* Chapter 10, Unit 10.13. doi: 10.1002/0471250953.bi1003s00
- Devoid, S., Overbeek, R., DeJongh, M., Vonstein, V., Best, A. A., and Henry, C. (2013). Automated genome annotation and metabolic model reconstruction in the SEED and Model SEED. *Methods Mol. Biol.* 985, 17–45. doi: 10.1007/978-1-62703-299-5\_2
- Ding, Y. H., Ronimus, R. S., and Morgan, H. W. (2000). Sequencing, cloning, and high-level expression of the pfp gene, encoding a PP(i)-dependent phosphofructokinase from the extremely thermophilic eubacterium *Dictyoglomus thermophilum*. *J. Bacteriol.* 182, 4661–4666. doi: 10.1128/JB.182.16.4661-4666.2000
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Euzéby, J. (2012). List of new names and new combinations previously effectively, but not validly, published. *Int. J. Syst. Evol. Microbiol.* 62, 1–4. doi: 10.1099/ijs.0.039487-0

- Euzéby, J. P. (1998). Taxonomic note: necessary correction of specific and subspecific epithets according to rules 12c and 13b of the international code of nomenclature of bacteria (1990 Revision). *Int. J. Syst. Evol. Microbiol.* 48, 1073–1075. doi: 10.1099/00207713-48-3-1073
- Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8, 186–194. doi: 10.1101/gr.8.3.186
- Fukusumi, S., Kamizono, A., Horinouchi, S., and Beppu, T. (1988). Cloning and nucleotide sequence of a heat-stable amylase gene from an anaerobic thermophile, *Dictyoglomus thermophilum*. *Eur. J. Biochem.* 174, 15–21. doi: 10.1111/j.1432-1033.1988.tb14056.x
- Gao, D., Uppugundla, N., Chundawat, S. P., Yu, X., Hermanson, S., Gowda, K., et al. (2011). Hemicellulases and auxiliary enzymes for improved conversion of lignocellulosic biomass to monosaccharides. *Biotechnol. Biofuels* 4:5. doi: 10.1186/1754-6834-4-5
- Gibbs, M. D., Reeves, R. A., and Bergquist, P. L. (1995). Cloning, sequencing, and expression of a xylanase gene from the extreme thermophile *Dictyoglomus thermophilum* Rt46B.1 and activity of the enzyme on fiber-bound substrate. *Appl. Environ. Microbiol.* 61, 4403–4408.
- Gibbs, M. D., Reeves, R. A., Sunna, A., and Bergquist, P. L. (1999). Sequencing and expression of a  $\beta$ -mannanase gene from the extreme thermophile *Dictyoglomus thermophilum* Rt46B.1, and characteristics of the recombinant enzyme. *Curr. Microbiol.* 39, 351–0357.
- Gordon, D., Abajian, C., and Green, P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res.* 8, 195–202. doi: 10.1101/gr.8.3.195
- Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S. R. (2003). Rfam: an RNA family database. *Nucleic Acids Res.* 31, 439–441. doi: 10.1093/nar/gkg006
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52–W57. doi: 10.1093/nar/gkm360
- Gumerov, V. M., Mardanov, A. V., Beletskii, A. V., Bonch-Osmolovskaia, E. A., and Ravin, N. V. (2011). [Molecular analysis of microbial diversity in the Zavarzin Spring, the Uzon caldera]. *Mikrobiologiya* 80, 258–265. doi: 10.1134/s002626171102007x
- Haas, B. J., Sandigursky, M., Tainer, J. A., Franklin, W. A., and Cunningham, R. P. (1999). Purification and characterization of *Thermotoga maritima* endonuclease IV, a thermostable apurinic/aprimidinic endonuclease and 3'-repair diesterase. *J. Bacteriol.* 181, 2834–2839.
- Hong, S. H., Lim, Y. R., Kim, Y. S., and Oh, D. K. (2012). Molecular characterization of a thermostable L-fucose isomerase from *Dictyoglomus turgidum* that isomerizes L-fucose and D-arabinose. *Biochimie* 94, 1926–1934. doi: 10.1016/j.biochi.2012.05.009
- Horinouchi, S., Fukusumi, S., Ohshima, T., and Beppu, T. (1988). Cloning and expression in *Escherichia coli* of two additional amylase genes of a strictly anaerobic thermophile, *Dictyoglomus thermophilum*, and their nucleotide sequences with extremely low guanine-plus-cytosine contents. *Eur. J. Biochem.* 176, 243–253. doi: 10.1111/j.1432-1033.1988.tb14275.x
- Hsiao, W., Wan, L., Jones, S. J., and Brinkman, F. S. (2003). IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics* 19, 418–420. doi: 10.1093/bioinformatics/btg004
- Huber, R., Langworthy, T. A., König, H., Thomm, M., Woese, C. R., Sleytr, U. B., et al. (1986). *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Arch. Microbiol.* 144, 324–333.
- Hubisz, M. J., Pollard, K. S., and Siepel, A. (2011). PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinformatics* 12, 41–51. doi: 10.1093/bib/bbq072
- Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal prokaryotic dynamic programming genefinding algorithm. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Ishino, Y., and Narumi, I. (2015). DNA repair in hyperthermophilic and hyperthermoresistant microorganisms. *Curr. Opin. Microbiol.* 25, 103–112. doi: 10.1016/j.mib.2015.05.010
- Jernigan, K. K., and Bordenstein, S. R. (2015). Tandem-repeat protein domains across the tree of life. *Peer J.* 3:e732. doi: 10.7717/peerj.732
- Karp, P. D., Ouzounis, C. A., Moore-Kochlacs, C., Goldovsky, L., Kaipa, P., Ahrén, D., et al. (2005). Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res.* 33, 6083–6089. doi: 10.1093/nar/gki892
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Kelman, L. M., and Kelman, Z. (2014). Archaeal DNA replication. *Annu. Rev. Genet.* 48, 71–97. doi: 10.1146/annurev-genet-120213-092148
- Kim, J. E., Kim, Y. S., Kang, L. W., and Oh, D. K. (2012). Characterization of a recombinant cellobiose 2-epimerase from *Dictyoglomus turgidum* that epimerizes and isomerizes  $\beta$ -1,4- and  $\alpha$ -1,4-gluco-oligosaccharides. *Biotechnol. Lett.* 34, 2061–2068. doi: 10.1007/s10529-012-0999-z
- Kim, M., Oh, H. S., Park, S. C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 64(Pt 2), 346–351. doi: 10.1099/ijs.0.059774-0
- Kim, Y. S., Shin, K. C., Lim, Y. R., and Oh, D. K. (2013). Characterization of a recombinant L-rhamnose isomerase from *Dictyoglomus turgidum* and its application for L-rhamnose production. *Biotechnol. Lett.* 35, 259–264. doi: 10.1007/s10529-012-1069-2
- Kochetkova, T. V., Rusanov, I. I., Pimenov, N. V., Kolganova, T. V., Lebedinsky, A. V., Bonch-Osmolovskaya, E. A., et al. (2011). Anaerobic transformation of carbon monoxide by microbial communities of Kamchatka hot springs. *Extremophiles* 15, 319–325. doi: 10.1007/s00792-011-0362-7
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315
- Kublanov, I. V., Perevalova, A. A., Slobodkina, G. B., Lebedinsky, A. V., Bidzhieva, S. K., Kolganova, T. V., et al. (2009). Biodiversity of thermophilic prokaryotes with hydrolytic activities in hot springs of Uzon Caldera, Kamchatka (Russia). *Appl. Environ. Microbiol.* 75, 286–291. doi: 10.1128/AEM.00607-08
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178
- Love, C. A., Patel, B. K. C., Ludwig, W., and Stackebrandt, E. (1993). The phylogenetic position of *Dictyoglomus thermophilum* based on 16S rRNA sequence analysis. *FEMS Microbiol. Lett.* 107, 317–320. doi: 10.1111/j.1574-6968.1993.tb06050.x
- Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.0955
- Mathrani, I., and Ahring, B. (1991). Isolation and characterization of a strictly xylan-degrading *Dictyoglomus* from a man-made, thermophilic anaerobic environment. *Arch. Microbiol.* 157, 13–17. doi: 10.1007/BF00245328
- Mathrani, I., and Ahring, B. (1992). Thermophilic and alkalophilic xylanases from several *Dictyoglomus* isolates. *Appl. Microbiol. Biotechnol.* 38, 23–27. doi: 10.1007/BF00169413
- Menzel, P., Gudbergsdóttir, S. R., Rike, A. G., Lin, L., Zhang, Q., Contursi, P., et al. (2015). Comparative metagenomics of eight geographically remote terrestrial hot springs. *Microb. Ecol.* 70, 411–424. doi: 10.1007/s00248-015-0576-9
- Morris, D. D., Gibbs, M. D., Chin, C. W., Koh, M. H., Wong, K. K., Allison, R. W., et al. (1998). Cloning of the xynB gene from *Dictyoglomus thermophilum* Rt46B.1 and action of the gene product on kraft pulp. *Appl. Environ. Microbiol.* 64, 1759–1765.
- Nishida, H., Beppu, T., and Ueda, K. (2011). Whole-genome comparison clarifies close phylogenetic relationships between the phyla Dictyoglomi and Thermotogae. *Genomics* 98, 370–375. doi: 10.1016/j.ygeno.2011.08.001
- Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., et al. (2005). The subsystems approach to genome annotation and its use

- in the project to annotate 1000 genomes. *Nucleic Acids Res.* 33, 5691–5702. doi: 10.1093/nar/gki866
- Patel, B. K., Morgan, H. W., Wiegel, J., and Daniel, R. M. (1987). Isolation of an extremely thermophilic chemoorganotrophic anaerobe similar to *Dictyoglomus thermophilum* from New Zealand hot springs. *Arch. Microbiol.* 147, 21–24. doi: 10.1007/BF00492899
- Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701
- Rawlings, N. D., Barrett, A. J., and Bateman, A. (2014). Using the MEROPS database for proteolytic enzymes and their inhibitors and substrates. *Curr. Protoc. Bioinformatics* 48, 1.25.1–33. doi: 10.1002/0471250953.bi0125s48
- Rees, G. N., Patel, B. K., Grassia, G. S., and Sheehy, A. J. (1997). Anaerobaculum thermoterrenum gen. nov., sp. nov., a novel, thermophilic bacterium which ferments citrate. *Int. J. Syst. Bacteriol.* 47, 150–154. doi: 10.1099/00207713-47-1-150
- Rigden, D. J., and Galperin, M. Y. (2008). Sequence analysis of GerM and SpoVS, uncharacterized bacterial ‘sporulation’ proteins with widespread phylogenetic distribution. *Bioinformatics* 24, 1793–1797. doi: 10.1093/bioinformatics/btn314
- Sahm, K., John, P., Nacke, H., Wemheuer, B., Grote, R., Daniel, R., et al. (2013). High abundance of heterotrophic prokaryotes in hydrothermal springs of the azores as revealed by a network of 16S rRNA gene-based methods. *Extremophiles* 17, 649–662. doi: 10.1007/s00792-013-0548-2
- Saiki, T., Kobayashi, Y., Kawagoe, K., and Beppu, T. (1985). *Dictyoglomus thermophilum* gen. nov., sp. nov., a Chemoorganotrophic, Anaerobic, Thermophilic Bacterium. *Int. J. Syst. Evol. Microbiol.* 35, 253–259.
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Shandilya, H., Griffiths, K., Flynn, E. K., Astatke, M., Shih, P. J., Lee, J. E., et al. (2004). Thermophilic bacterial DNA polymerases with reverse-transcriptase activity. *Extremophiles* 8, 243–251. doi: 10.1007/s00792-004-0384-5
- Shi, R., Li, Z., Ye, Q., Xu, J., and Liu, Y. (2013). Heterologous expression and characterization of a novel thermo-halotolerant endoglucanase Cel5H from *Dictyoglomus thermophilum*. *Bioresour. Technol.* 142, 338–344. doi: 10.1016/j.biortech.2013.05.037
- Spriestersbach, A., Kubicek, J., Schäfer, F., Block, H., and Maertens, B. (2015). Purification of his-tagged proteins. *Meth. Enzymol.* 559, 1–15. doi: 10.1016/bs.mie.2014.11.003
- Svetlichny, V. A., and Svetlichnaya, T. P. (1988). *Dictyoglomus turgidus* sp. nov., a new extremely thermophilic eubacterium isolated from hot springs of the Uzon volcano caldera. *Mikrobiologiya* 57, 435–441.
- Takai, K., Inoue, A., and Horikoshi, K. (1999). *Thermaerobacter marianensis* gen. nov., sp. nov., an aerobic extremely thermophilic marine bacterium from the 11,000 m deep Mariana Trench. *Int. J. Syst. Bacteriol.* 49(Pt 2), 619–628. doi: 10.1099/00207713-49-2-619
- Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10, 512–526.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Techtmann, S. M., Colman, A. S., and Robb, F. T. (2009). That which does not kill us only makes us stronger: the role of carbon monoxide in thermophilic microbial consortia. *Environ. Microbiol.* 11, 1027–1037. doi: 10.1111/j.1462-2920.2009.01865.x
- Vesth, T., Ozen, A., Andersen, S. C., Kaas, R. S., Lukjancenko, O., Bohlin, J., et al. (2013). Veillonella, firmicutes: microbes disguised as Gram negatives. *Stand. Genomic Sci.* 9, 431–448. doi: 10.4056/sigs.2981345
- Wagner, I. D., and Wiegel, J. (2008). Diversity of thermophilic anaerobes. *Ann. N.Y. Acad. Sci.* 1125, 1–43. doi: 10.1196/annals.1419.029
- Wijffels, G., Dalrymple, B., Kongsuwan, K., and Dixon, N. E. (2005). Conservation of eubacterial replicases. *IUBMB Life* 57, 413–419. doi: 10.1080/15216540500138246
- Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucl. Acids Res.* 39, W347–W352. doi: 10.1093/nar/gkr485

**Conflict of Interest Statement:** At the time this work was performed, the authors PB and DM were employees and shareholders of C5-6 Technologies Inc. (WI, USA), a company that created bio-based solutions to efficiently convert biomass into five and six carbon sugars. The company ceased operation in December of 2014. PB has since purchased the assets of the company and started C5-6 Technologies LLC (WI, USA), a company focused on supplying reagent enzymes for carbohydrase research. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed. No writing assistance was utilized in the production of this manuscript.

The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Brumm, Gowda, Robb and Mead. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.