Genomic Comparison of Facultatively Anaerobic and Obligatory Aerobic *Caldibacillus debilis* Strains GB1 and Tf Helps Explain Physiological Differences

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Canadian Journal of Microbiology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>cjm-2018-0464.R2</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>18-Dec-2018</td>
</tr>
</tbody>
</table>
| Complete List of Authors: | Wushke, Scott; University of Manitoba  
Froese, Alan; University of Manitoba, Microbiology  
Fristensky, Brian; University of Manitoba, Plant Science  
Zhang, Xiang; University of Manitoba, Plant Science  
Spicer, Victor; University of Manitoba, Physics & Astronomy  
Krokhin, Oleg; University of Manitoba, Physics & Astronomy  
Levin, David B.; University of Manitoba, Biosystems Engineering  
Sparling, Richard; University of Manitoba, Microbiology |
| Keyword:          | Thermophile, Fermentation, Genomics                   |
| Is the invited manuscript for consideration in a Special Issue? : | Not applicable (regular submission)   |
Genomic Comparison of Facultatively Anaerobic and Obligatory Aerobic *Caldibacillus debilis* Strains GB1 and Tf Helps Explain Physiological Differences

Scott Wushke¹, Alan Froese¹, Brian Fristensky², Xiang Li Zhang², Vic Spicer³, Oleg V. Krokhin⁴, David B. Levin⁵, Richard Sparling†*

¹Department of Microbiology, University of Manitoba, Winnipeg, Manitoba, Canada

²Department Plant Science, University of Manitoba, Winnipeg, Manitoba, Canada

³Department Physics & Astronomy, University of Manitoba, Winnipeg, Manitoba, Canada

⁴Department of Internal Medicine & Manitoba Centre for Proteomics and Systems Biology, University of Manitoba, Winnipeg, Manitoba, Canada

⁵Department of Biosystems Engineering, University of Manitoba, Winnipeg, Manitoba, Canada

* Corresponding Author: Richard Sparling, Department of Microbiology, University of Manitoba, Winnipeg, Manitoba, Canada, R3T 2N2; Tel: (204) 474-8320; Fax: (204) 474-8320; E-mail: richard.sparling@ad.umanitoba.ca
Abstract

*Caldibacillus debilis* strains GB1 and Tf display distinct phenotypes. *C. debilis* GB1 is capable of anaerobic growth and can synthesize ethanol while *C. debilis* Tf cannot. Comparison of the GB1 and Tf genome sequences revealed that the genomes were highly similar in gene content and showed a high level of synteny. At the genome scale, there were several large sections of DNA that appeared to be from lateral gene transfer into the GB1 genome. Tf did have unique genetic content but at a much smaller scale; 300 genes in Tf verses 857 genes in GB1 that matched at ≤90% sequence similarity. Gene complement and copy number of genes for the glycolysis, tricarboxylic acid (TCA) cycle, and electron transport chain (ETC) pathways were identical in both strains. While Tf is an obligate aerobe, it possesses the gene complement for an anaerobic lifestyle (ldh, ak, pta, adhE, pfl). As a species, other strains of *C. debilis* should be expected to have the potential for anaerobic growth. Assaying the whole cell lysate for ADH activity revealed an approximately 2-fold increase in the enzymatic activity in GB1 when compared to Tf

Keywords: Thermophilic, Fermentation, Genomics
Importance

This work characterizes the genomes of the thermophilic *C. debilis* strains GB1 and Tf with regard to core metabolism. Comparison of two highly related but physiologically distinct strains highlights the physiology and genome potential of the single species genus *Caldibacillus* especially with regards to fermentation. *Caldibacillus* spp. may be of interest in industrial process, namely biofuel production, a deep understanding of core metabolism is needed in order to optimize these processes.
1.0 Introduction

*Caldisbacillus debilis* GB1 was isolated from a co-culture found to convey aerotolerance to *Clostridium thermocellum* when co-cultured aerobically on cellulose (Wushke et al. 2013). The recently defined single species genus *Caldisbacillus* has been the subject of only a few physiological studies in pure culture (Banat et al. 2004; Wushke et al. 2015). *Caldisbacillus debilis* Tf was originally isolated from a cool soil environment in Northern Ireland (Banat et al. 2004). Contrary to what was observed with *C. debilis* GB1, *Caldisbacillus debilis* Tf was unable to create a micro-aerophilic environment suitable for co-culturing with the strict anaerobe *Clostridium thermocellum* (Wushke et al. 2015). We believe that the inability of *C. debilis* Tf to grow anaerobically may affect its ability to create the reduced anaerobic microenvironment needed for *C. thermocellum* to thrive. Previously, Wushke et al. (2015) compared the physiological profile of two strains of *C. debilis*, GB1 and Tf. They showed near identical characteristics in growth rate, temperature optimum, substrate utilization profile, and types of end-products produced under aerobic and oxygen-limiting conditions, with the notable exception of ethanol production. A *C. debilis* Tf draft genome has been generated and annotated by the Joint Genome Institute (JGI) (Grigoriev et al. 2012), but there has not been a detailed genome analysis published. The *C. debilis* GB1 genome was previously sequenced and its proteome was analyzed with respect to aerobic and anaerobic growth (Wushke et al. 2017). Because of the major physiological differences observed, including anaerobic growth and ethanol synthesis, we hypothesized that a comparison of the genomes sequences of the GB1 and Tf, with a focus on core metabolism, could explain the observed physiological differences. Part of the work presented herein was adapted from the Ph.D. thesis of Scott Wushke (2017).
2.0 Materials and Methods

2.1 Cell Culturing

The methods for growing *Caldibacillus debilis* strain GB1 DSM 29516 and *C. debilis* Tf DSM 16016 cultures were described previously (Wushke *et al.* 2013, 2015). *C. debilis* strains were grown on modified 1191 (M-1191) medium (Islam *et al.* 2006) with a lower concentration of yeast extract (0.76 g/L) and the pH was adjusted to 7.2 for all experiments. Sealed Balch tubes (26 mL) from Bellco Glass Inc. (Vineland, NJ) and 1L Corning bottles supplied from Fisher Scientific (Toronto, ON), were used to carry out experiments. Anaerobic and aerobic environments in Balch tubes and Corning bottles were prepared as previously described by Wushke *et al.* (2015).

2.2 Genomic Comparison

The GB1 genome was sequenced and annotated as previously described (Wushke *et al.* 2017). The GB1 and Tf genome sequences can both be accessed through the National Center for Biotechnology Information (NCBI) under the catalogue numbers AZRV00000000 and ARVR00000000, respectively. Joint Genome Institute (JGI) Integrated Microbial Genomes (IMG) Expert Review (IMG/ER) was used to compare the bacterial genomes (Markowitz *et al.* 2008). NUCmer plots were generated through IMG/ER. Gview was used to create genome-scale comparative images (Stothard *et al.* 2017, Petkau *et al.* 2010). Protein and DNA alignments were done with Multiple Sequence Alignment Tool (MUSCLE) (Edgar *et al.* 2004). NCBI Blast was used to create a 2D model of domains and key amino acids involved in function (Johnson *et al.* 2008). 3D models of the ADHE proteins were created using RAPTORX (Källberg *et al.* 2012).
2.3 Protein Extraction

Samples for protein extraction were collected during early exponential in aerobic growth, OD$_{600}$ 0.2, shaken at 150rpm for both strains GB1 and Tf of *C. debilis*. Approximately ~200 mg cell pellets were washed twice in phosphate buffered saline (PBS) buffer then re-suspended in 1 mL of lysis buffer [1% sodium dodecyl sulfate (SDS), 100 mM dithiothreitol (DTT)] in 100 mM ammonium bicarbonate (ABC). The mixture was vortexed for 30 seconds (sec), then placed in boiling water for 5 minutes (min), and then cooled for 5 min. Samples were then sonicated three times with 15 sec pulses, and 1 min breaks in-between then were centrifuged at 16000 x g for 20 min. Supernatants containing the proteins were kept for further processing, and stored at -80 ºC. Aliquots (20 μL) of the protein containing supernatants were diluted with 180 μL of 100 mM ABC buffer. Twenty (20) μL of 500 mM iodoacetic acid (IAA) were added, vortexed for 5 sec and stored in the dark for 45 min to complete alkylation. To quench excess IAA, 33 μL of 100mM DTT was added. Then, 150 μL of ABC were added to bring total volume to ~380 μL to reduce the SDS concentration in order to allow digestion. Trypsin (4 μg) was added, and CaCl$_2$ was added to a final concentration of 10 mM. Samples were incubated for 12 hours at 37 ºC to allow trypsinization to occur. Samples were then dried using roto-evaporation (Savant Speedvac-sc110), and 200 μL of 3 M KCl was added and vortexed for 10 min, then were spun at 16000 x g for 25 min and supernatants were gently removed from the precipitate (containing the SDS, which is incompatible mass spectrometry methods). The samples were diluted 2 fold with 0.5% trifluoroacetic acid (TFA) then desalted by HPLC using a C18 column as described previously (Wushke *et al.* 2017).
2.4 Mass Spectrometry Methods

One-dimensional liquid chromatography followed by mass spectrometry (1D-LC-MS) proteomics were used to identify proteins by their peptide sequences, as previously described by Gungormusler-Yilmaz et al. (2014). Proteomic profiles of the *C. debilis* strains GB1 and Tf grown to early exponential phase on cellobiose were compared using an in-house data analysis system called UNITY (Table S5), as previously described (Fu et al. 2015). The XML raw output file from an X!tandem search and its corresponding TIC intensities can be found at http://hs2.proteome.ca/thegpm-cgi/plist.pl?path=/gpm/archive/C.debilis_GB1/GB1A-BASERESULTS.xml&proex=-1&npep=0 and http://hs2.proteome.ca/thegpm-cgi/plist.pl?path=/gpm/archive/C.debilis_GB1/GB1B-BASERESULTS.xml&proex=-1&npep=0 and http://hs2.proteome.ca/thegpm-cgi/plist.pl?path=/gpm/archive/C.debilis_GB1/DSM1606AX-BASERESULTS.xml&proex=-1&npep=0.

2.5 Proteogenomics

To enhance the genome sequence and annotation, a proteogenomic approach was used, on the basis of the proteomes described above. We have previously described the use of proteomic data to identify and correct improperly annotated genes using a naïve assembler (Wushke et al. 2017) using the same methodology as Verbeke et al. (2014). We used these proteomic data to verify unique regions and genes missing in the *C. debilis* GB1 and Tf genome by analyzing the peptides generated with respect to both the GB1 and Tf genomes.
2.6 Whole cell extract enzymatic assays

*C. debilis* cells grown to early exponential (OD$_{600}$ 0.2) on cellobiose were pelleted through centrifugation at 13,000 rpm for 2 minutes. The pelleted cells were resuspended in lysis buffer containing 10 mM MOPS, pH 7.4, 20 mM NaCl, and 5 mM DTT, and lysed using sonication for 6 minutes with 30 seconds on / 30 seconds off cycles. The alcohol dehydrogenase (ADH) assay reaction mixture contained 0.2 mM NADH or NADPH, 10 mM acetaldehyde, 10 mM DTT, 100 mM MOPS, and 20 µl cell extract. The final volume was 200 µl, the assay temperature was 50°C, the pH was 7.4, and the assay was started by the addition of acetaldehyde. The ADH reaction was also done in the reverse direction using 1.36 M ethanol, 5 mM NAD$^+$ or NAD(P)$^+$, 10 mM DTT, 100 mM MOPS, and 20 µl cell extract. In the ALDH forward reaction the acetaldehyde in the reaction mixture was replaced with 0.5 mM acetyl-CoA. In the reverse direction for the ALDH reaction 1 mM CoA was added and the NAD(P)H was replaced with 2.5 mM NAD(P)$^+$. Background activity was measured by adding water instead of the usual starter chemical, and was subtracted from the reaction activity recorded. The changes in NAD(P)(H) concentrations were recorded in a BioTek synergy 4 plate reader by monitoring the absorbance at 340 nm. All enzymatic assays were done in triplicate.
3.0 Results

3.1 GB1 and Tf Genome Comparison

A comparison of the GB1 (Wushke et al. 2017) and Tf (Mukherjee et al. 2017) permanent draft genomes containing 49 and 41 contigs respectively can be found in Table 1 (complete genome comparison summary in Table S1). CheckM analysis (Parks et al. 2014) indicates 99.42 completeness for Tf and 99.27 for GB1 indicating that these genomes are mostly complete with respect to functional genes. A pairwise ANI of the two genomes is consistent with two strains of the same species (Table S2). The genome sequences of C. debilis strains GB1 and Tf are highly similar in both gene identity and genome synteny (2354 genes in matching cassettes, where a cassette consists of blocks of associated genes in sets of 2 or more). GB1 has 2570 genes matching Tf with a sequence identity ≥90%. Figure S2 shows a NUCmer plot demonstrating their synteny produced using IMG/ER. When the genome sequences of the two strains were compared, there were 853 genes in GB1 that did not match with a sequence identity greater than or equal to 90% in Tf. Since these strains primarily have highly related genes, matches with ≤90% sequence identity were considered to be unique. Many of the unmatched genes found in GB1 were considered hypothetical (408). Many of the unmatched genes could be attributed to two structural differences: i) the addition of 207 genes on contig 4 which appear to form a conjugative element; ii) the addition of the full length genome of a cryptic prophage on contig 16 described elsewhere Wushke et al. (2018), as well as several partial copies of this prophage associated with the end of various other contigs. There were also 16 transposase-associated genes dispersed through the genome several of which were near identical copies. Figure 1 provides a graphical representation produced by G-View of the GB1 and Tf aligned genome contigs with genes matching at ≤90% sequence similarity highlighted. In
contrast to GB1, Tf has only 300 unique genes that do not match sequences in the GB1 genome at a ≤90% sequence similarity. Unlike GB1 most of the unique genes in Tf were scattered throughout the various contigs and did not appear in any large contiguous sections. The unique genes seen in Tf came from a variety of genes including different transposase-associated genes, different small hypothetical genes, and differences in ABC cassettes. Figure S1 is a Venn diagram of unique genes of GB1 and Tf at a 90% sequence similarity cut off.

Sequence analyses of GB1 revealed that there are 24 tRNA genes (Cdeb_01241-Cdeb_01300) on contig 4 that are associated with the conjugative element described above. There are 85 tRNA genes in GB1 compared to only 58 in Tf, as shown in Table S3.

Investigation of differences in the gene complement of the Phosphotransferase (PT) systems show that GB1 has several extra subunits associated with putative cellulose transport specificity (Table S4), consistent with its isolation from cellulose enrichment cultures and plates containing cellulose as carbon and energy substrate (Wushke et al. 2013).

We found two annotation differences between GB1 and Tf during our investigation: i) LSU ribosomal protein L36P (Cdeb_03413), while Tf possessed the exact same DNA as GB1 sequence but the ORF was missing in Tf. ii) GB1 has its pyruvate kinase split into two ORF’s (Cdeb_02915, Cdeb_02914) which was annotated as separate but adjacent genes representing the two domains of the pyruvate kinase in GB1 and one gene, a pyruvate kinase, in Tf even though DNA in the coding region is 100% identical.

3.2 Proteogenomics

The general features of the C. debilis GB1 genome have been described previously, and proteogenomics used to correct some annotation start sites (Wushke et al. 2017). In this study we
used a further single shallow proteome for each strain (Table S5), GB1 and Tf (704 and 473 proteins detected respectively), during early aerobic growth (Figure S3) which allowed us to: i) further check for genes that were missed due to incomplete assembly of sequences and ii) check whether or not genes in regions which appeared unique to GB1 were expressed. No new genes were found in GB1 or the Tf using proteogenomics that would indicate incomplete sequencing of major transcribed and translated genes. The regions that appeared to be unique to GB1, a cryptic prophage region within contig 16 and conjugation equipment comprising contig 4, resulted in 5 and 3 proteins detected respectively. When the Tf proteome was used to check the GB1 genome unique regions on contig 16 and contig 4 no proteins were observed, indicating that these regions were most likely absent in Tf, but unique to GB1.

3.4 Pyruvate Fermentation

A list of pyruvate metabolism genes and their abbreviations GB1 and Tf can be found in Table 2, the full table with corresponding locus tags can be found in the supplementary (Table S6). When focusing on pyruvate fermentation pathways, small differences were observed in regions within 200bp up stream of a several genes of interest shown in Figure S4. When examining the genes each appeared to have a Pribnow box and transcription start site. In some cases, the differences in the upstream regions between GB1 and Tf were extensive (ldh, pfl, adh), compared to others (ak, pta, adhE, aldh) where differences were observed but were less extensive, as seen in Figure S4. Some of the coding sequences for pyruvate fermentation contained differences (adhE, pfl activating enzyme, one of two aldh, and four of the adh genes), shown in Figure S5, and others (one of two aldh, pfl C-acetyltransferase, ldh and ak) did not, as would be expected in closely related strains. The activity of key enzymes PFL, LDH, AK, and ADHE could be surmised by end-product analysis under oxygen-limited conditions, as their
corresponding products were detected, with the exception of ethanol in Tf, even though the
genome of strain Tf encodes ADHE. Both strains of *C. debilis* have a single copy of *adhe* as well
as discrete copies of *adh* and *aldh*.

While lactic acid production was low in both organisms, differences in lactic acid
production were observed between the strains, with GB1 typically producing significantly less
than Tf under comparable conditions (Wushke *et al.* 2015). There are significant differences in
upstream regions of the *ldh* gene between the two strains; these changes may have affected the
transcriptional regulation of LDH.

When cultured under oxygen-limiting conditions both strains, GB1 and Tf, synthesize
formate and acetate as end-products indicating that both are able to use PFL (Wushke *et al.*
2015). There is a difference in the amount of formate produced by each strain under oxygen-
limited conditions; significant differences in the upstream regions of PFL are observed.

3.5 Alcohol Aldehyde Dehydrogenase (*adhE*) fine analysis and Alcohol Aldehyde
Dehydrogenase whole lysate assays

While the genomes of both strains encoded a putative ADHE enzyme, detailed analyses
revealed sequence differences in regions upstream as well as within the coding sequence of the
*adhE* gene. Nucleotide substitutions in the ADHE coding sequence, shown in Figure S6, result in
different amino acids: GB1 → Tf = 269A→E, 482D→E, 495V→I, 825D→A, 829R→K,
831P→H. These amino acid changes, shown graphically in Figure S6 are caused by a single
nucleotide polymorphism in all but one case, 495V→I where two bp were changed. Changes
482D→E and 495V→I, are most likely benign sense mutations. Two codon changes, 825 D→A
and 829R→K, are seen in other functional ADHE alignments and seem to be normal variations
in coding sequence (Extance *et al.* 2013). However, 269A→E is a missense substitution rather
close to a putative catalytic cysteine at position 255. The 831P→H codon change is a missense mutation that appears to place a charged amino acid in an α-helix region on the outer shell of the Tf ADHE protein. The changes in the ADH domain of ADHE are not directly within any putative active site, cofactor binding-site, or known allosteric regulation site (Zheng et al. 2015). The Blast domain analysis is shown in Figure S7. These amino acid substitutions affect a small loop connecting an α-helix region (Extance et al. 2013), and therefore could affect overall protein or multimeric protein structure.

While there are also differences in regions upstream of adhE, they did not appear to affect the promoter binding site and are not as extensive as the differences observed in ldh and pfl (Figure S4).

In order to determine if ADH activity or expression could account for the observed differences in end-product production and anaerobic growth whole cell lysates were assayed. The cell extract ADH activity in the forward direction for GB1 was 2.14 (st. dev = 0.031) U/mg of cell-free extract protein (U = 1 μmol of substrate/min) and for Tf was 1.06 (st. dev. = 0.060) U/mg (Table 3). The ADH activity in the reverse direction for GB1 was 5.75 (st. dev = 0.833) U/mg of cell-free extract protein and for Tf was 0.861 (st. dev. = 0.031) U/mg (Table 3). There was no detectable ADH activity for either GB1 or Tf when using NADPH as the cofactor in the cell extract. The activity of AlDH could not be effectively measured in the reverse direction, possibly due to the presence of an aldh (Cdeb_00412) that only required H₂O, acetaldehyde, and NAD⁺ to form acetate. This enzyme activity masked the CoA dependent activity of aldh, Table S7 shows that activity was the same regardless of the addition of HS-CoA. The forward reaction was not tried as acetyl-CoA is a substrate not only for aldh but also for pyruvate dehydrogenase (PDH) which again would mask results in cell free-extracts.
4.0 Discussion

4.1 Genomics

The genome comparison confirms that these two organisms are distinct strains of the same species despite their divergent capabilities to grow in the absence of oxygen. While there are some genome structural changes observed, the metabolic potential appears to be nearly identical. Interpreting changes in physiology on the basis of minor changes at the nucleotide level in and surrounding some of the core metabolism genes must be approached with caution as such differences are also consistent with what would be expected in distinct strains (Hayashi et al. 2001). Indeed many of the differences noted between GB1 and Tf would not be expected to affect gene expression or activity (e.g. silent mutations, mutations in non-coding areas).

Differences in tRNA complement have the potential to affect gene regulation at a global level at the level of translation. However, all of the tRNA’s unique to GB1 had anticodons that were shared by Tf and GB1, therefore the codon usage ability was not expanded.
4.2 Pyruvate Fermentation

Oxygen depleting conditions, i.e. conditions that become microaerobic, were the most informative to observe differences in end-product profiles between the 2 strains, Table S8 (Wushke et al. 2015). Indeed the end product profile observed previously by Wushke et al. (2015) was key in driving the genomic comparison done in this study.

Many wild type thermophilic and mesophilic organisms synthesize lactate as a major fermentative end-product (Narayanan et al. 2004). While there are differences in the proportion of lactate produced for Tf and GB1 and differences in the DNA sequence of the gene, clear reasons for difference in lactic acid production were not apparent despite the difference in LDH sequence.

Production of formate and acetate alone from pyruvate cannot act as an electron sink for NADH generated by glycolysis. Under microaerobic conditions where oxygen is present and formate and acetate are being produced, Tf is able to use the electrons generated through glycolysis via the electron transport chain (ETC). If typical glycolysis and PFL are the primary means by which acetyl-CoA is generated, then anaerobic growth is not possible unless an additional reduced compound is produced. Strain GB1 is capable of anaerobic growth because it is also able to synthesize ethanol. Typically, expression of an air-sensitive enzyme, such as PFL, is linked to anaerobic metabolism, including in Bacillus subtilis which has a similar end-product profile as C. debilis GB1 (Nakano et al. 1998). The use of PFL can be important for biosynthesis in other Firmicutes (Rydzak et al. 2015). It is more likely that Tf represents a strain that has lost the ability to produce ethanol, preventing anaerobic growth. Tf still maintains regulation and expression of PFL since formate production is seen under oxygen limiting conditions (Wushke et al. 2015; 2017). Since both strains demonstrate the ability to express PFL, as indicated by the
The presence of formate, the focus should be on ethanol production genes, since this is where there is a clear physiological difference.

The synthesis of ethanol appears to be the main mechanism used by *C. debilis* GB1 to dispose of excess electrons from glycolysis under anaerobic conditions, as lactic acid production seems to be limited in both strains under the growth conditions used (Wushke et al. 2015; 2017). Ethanol synthesis may therefore be an anaerobic requirement in *C. debilis* GB1. ADHE is associated with the bulk of ethanol production in many anaerobic organisms (Carere et al. 2012, Peng et al. 2008).

### 4.3 Lack of ethanol synthesis in Tf

The lack of ethanol synthesis in *C. debilis* Tf may be understood by looking at the sequence of the bifunctional alcohol/aldehyde dehydrogenase (*adhE*) gene. Differences in *adhE* gene expression regulation or absence of ADHE enzyme activity could result in the ethanol-minus Tf phenotype and explain the lack of anaerobic growth. There is precedence in *Geobacillus thermoglucosidasius*, an organism from closely related genus, which produces similar end-products to *C. debilis* GB1 under both aerobic and anaerobic growth conditions (Extance et al. 2013). *G. thermoglucosidasius* has a similar core metabolism gene complement to *C. debilis* GB1 including several ADH genes, a standalone AIDH and a ADHE Extance et al. (2013) noted that gene knock-outs of *G. thermoglucosidasius* ADHE results in a lack of anaerobic growth and an ethanol-minus phenotype despite the presence of alternate annotated alcohol dehydrogenases and AIDH, very similar to the Tf phenotype. Furthermore, we do observe a ~2 fold increase in acetaldehyde reduction to ethanol activity in *C. debilis* GB1 cell extracts, and a 5 fold difference in the opposite direction. Taken together, the results suggest that
the phenotype differences observed when comparing GB1 and Tf could be due to the differences in the amino acid sequence of the ADHE.

4.4 Respirofermentative Metabolism

Previous experiments done by Wushke et al. (2015) characterised end-product production on cellobiose aerobically and anaerobically for strains GB1 and Tf. Under conditions where oxygen concentrations are low Tf and GB1 both utilise incomplete respiration and non-reductive components of fermentative pathways, also referred to as respirofermentative growth (Siso et al. 1996, Zigha et al. 2007). Incomplete respiration can happen when TCA cycle flux is less than that of glycolysis (Vemuri et al. 2006). During respirofermentative growth a significant proportion of the electrons produced from glycolysis are consumed by respiration, but excess electrons from glycolysis are shunted to fermentation products such as ethanol. GB1 can either use its electrons to create ethanol and/or lactate and/or use them in ETC; based on end-product production under oxygen limiting conditions it appears to do all three. Tf, when undergoing respiro-fermentative growth, utilizes the ETC and produces lactate as alternate electron sink. While organism like Lactobacillus balance its electrons anaerobically by producing lactate alone (De Vries et al. 1970), Tf does not do this under the growth conditions tested. While Tf and GB1 both have additional adh and aldh genes, these may be used primarily for aerobic ethanol oxidation; Wushke et al. (2015) show GB1 is capable of consuming ethanol aerobically.

4.5 Issues with the current genus description for Caldibacillus

Members of the Bacillus and Geobacillus genera appear to fill many niches within both aerobic and facultatily anaerobic environmental communities. The loss of anaerobic metabolism would seem to be a rather large change regarding taxonomical classification, which
would typically lead to different strains being described as different species based on physiological characteristics. However, strains GB1 and Tf do appear to be 2 strains of the same species based on their genome comparison. The inability of Tf to undergo anaerobic metabolism appears to be the result of one or several very small changes in the genomes. The ability of similar strains to fulfill distinct roles within a co-culture has been previously described by Verbeke et al. (2011). However, because of the geographic distance between the sources of these 2 strains it is impossible at this point to state whether loss of anaerobic growth capabilities is due to natural microdiversity within the species or post enrichment genetic drift because of differences in growth conditions in laboratory prior to sequencing.

Overall *C. debilis* GB1 and Tf are very similar organisms at the genetic level, in both synteny and gene complement, including core metabolism pathway genes. Through our genome analysis a gene complement consistent with fermentation (*pfl, ldh, ak, pta,* and *adhE*) was observed in Tf and GB1. Tf likely evolved from a strain that could ferment cellobiose and produce ethanol. The ability of *C. debilis* GB1 to grow under anaerobic conditions and synthesize ethanol clearly distinguishes it from the *C. debilis* the type strain, Tf. The genus description of *Caldibacillus* should be amended to include anaerobic growth as a possible characteristic for strains of this genus, since it is observed in at least one strain of the only species currently described.

**Acknowledgements**

This research was supported by Genome Canada, through the Applied Genomics Research in Bioproducts or Crops (ABC) program for the grant titled, “Microbial Genomics for Biofuels and Co-Products from Biorefining Processes” and by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant (RGPIN-06173-2014) held by Dr. Sparling.
References


Figure 1. Unique Genome of *C. debilis* GB1 (panel A) and Tf (Panel B). Gene bank file (.gbk) comparison of Unique Genome genes that match ≤90% sequence similarity to the corresponding strain are shown in red, ORF’s are shown in brown. This figure is a graphical representation of the differences between the two genomes created by concatenating the contigs of each organism and then comparing them. In panel A, Contig 4 is shown with a green arrow, the full putative phage genome shown with orange arrow, plasmid shown with blue arrow. In panel B the areas of uniqueness are much shorter, typically only corresponding to a few genes. The purple arrow points to a divergent ATP binding cassette transporter.
Table 1. General genome features of *C. debilis* GB1 and Tf (source: IMG/er)

<table>
<thead>
<tr>
<th></th>
<th>GB1</th>
<th>Tf</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA, total number of bases</td>
<td>3346235</td>
<td>3061346</td>
</tr>
<tr>
<td>DNA coding number of bases</td>
<td>2670185</td>
<td>2430336</td>
</tr>
<tr>
<td>% G+C content</td>
<td>51.2</td>
<td>51.6</td>
</tr>
<tr>
<td>DNA scaffolds</td>
<td>49</td>
<td>41</td>
</tr>
<tr>
<td>CRISPR Count</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Genes total number</td>
<td>3374</td>
<td>2896</td>
</tr>
<tr>
<td>Protein coding genes</td>
<td>3264</td>
<td>2800</td>
</tr>
<tr>
<td>RNA genes</td>
<td>110</td>
<td>96</td>
</tr>
<tr>
<td>rRNA genes</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>5S rRNA</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>16S rRNA</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>23S rRNA</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>tRNA genes</td>
<td>85</td>
<td>58</td>
</tr>
<tr>
<td>Other RNA genes</td>
<td>17</td>
<td>25</td>
</tr>
<tr>
<td>Protein coding genes with function prediction</td>
<td>2450</td>
<td>2212</td>
</tr>
<tr>
<td>without function prediction</td>
<td>814</td>
<td>588</td>
</tr>
<tr>
<td>Shared genes (&lt;90% sequence similarity)</td>
<td>2571</td>
<td></td>
</tr>
<tr>
<td>Unique genes (&lt;90% sequence similarity)</td>
<td>853</td>
<td>300</td>
</tr>
</tbody>
</table>
Table 2. Pyruvate metabolism gene complement of GB1 and Tf as specified through KEGG maps built in IMG/ER

<table>
<thead>
<tr>
<th>Pyruvate Metabolism</th>
<th># of gene copies in GB1 (same as Tf)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pyruvate formate lyase (pfl)</td>
<td>2</td>
</tr>
<tr>
<td>Bifunctional aldehyde/alcohol dehydrogenase (adhe)</td>
<td>1</td>
</tr>
<tr>
<td>Acetate kinase (ak)</td>
<td>1</td>
</tr>
<tr>
<td>Phosphotransacetylase (pta)</td>
<td>1</td>
</tr>
<tr>
<td>Lactate dehydrogenase (ldh)</td>
<td>1</td>
</tr>
<tr>
<td>Aldehyde dehydrogenase (aldh)</td>
<td>2</td>
</tr>
<tr>
<td>Alcohol dehydrogenase (adh)</td>
<td>4</td>
</tr>
<tr>
<td><strong>Pyruvate Dehydrogenase (pdh)</strong></td>
<td></td>
</tr>
<tr>
<td>E1 (EC:1.2.4.1)</td>
<td>6</td>
</tr>
<tr>
<td>E2 (EC:2.3.1.12)</td>
<td>3</td>
</tr>
<tr>
<td>E3 (EC:1.8.1.4)</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 3. NADH/NAD\(^+\) dependent ADH activity in whole cell lysate

<table>
<thead>
<tr>
<th>Strain</th>
<th>added to reaction mixture</th>
<th>U/mg protein</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB1</td>
<td>acetaldehyde</td>
<td>2.033</td>
<td>0.031</td>
</tr>
<tr>
<td><strong>Total activity of GB1 ADH in forward direction</strong></td>
<td>2.137</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tf</td>
<td>acetaldehyde</td>
<td>1.281</td>
<td>0.059</td>
</tr>
<tr>
<td><strong>Total activity of Tf ADH in forward direction</strong></td>
<td>1.059</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GB1</td>
<td>ethanol</td>
<td>6.59</td>
<td>0.833</td>
</tr>
<tr>
<td><strong>Total activity of GB1 in reverse direction</strong></td>
<td>5.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tf</td>
<td>ethanol</td>
<td>0.710</td>
<td>0.031</td>
</tr>
<tr>
<td><strong>Total activity of Tf in reverse direction</strong></td>
<td>0.861</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Corrected by subtracting the no carbon substrate control