Supplementary information for

Isolation of two species of *Caldatribacterium* (*Atribacterota*) and the importance of folate preparation on their culturability

Brian P. Hedlund[1]*, Toshio Alvarado[2], Trevor R. Murphy[3], Aries Torres[2], Amanda M. Blocker[3], Cale O. Seymour[1,4], Dengxun Lai[1], Xavier Mayali[5], Alise Muok[6,7], Rui Xiao[8], Erin Levon[3], John T. Phoenix[3], Katherine Mazdra[3], Tyler Sons[3], Jennifer Ledesma[2], Janette Jiries[2], Alexander Manriquez[2], Andreas Vamvakas[2], Adrian Rivera Martinez[2], Alberto Camacho[2], Zofia Owerkowicz[2], Jacob Lacy[2], Peter K. Weber[5], Ariane Briegel[6,9], Duane P. Moser[10], Jennifer Pett-Ridge[5,11,12], Jeremy A. Dodsworth[2]*, and Scott D. Hamilton-Brehm[3]*

*Corresponding Authors Email: brian.hedlund@unlv.edu, Scott.Hamilton-Brehm@siu.edu, and JDodsworth@csusb.edu

List of Supplementary Notes.

List of Supplementary Figures.

Supplementary Note 1. Additional information on the isolation, characterization, and taxonomy of *Thermodesulfobacterium auxiliatoris*.

Isolation of *Thermodesulfobacterium auxiliatoris*. A *Thermodesulfobacterium* strain was isolated on GBS salts medium solidified with 1% gelrite as described for *Caldatribacterium* with the following modifications: sugar and yeast extract were not added, and the medium instead contained 1 mM each of sodium bicarbonate, sodium acetate, and sodium thiosulfate. Plates were incubated at 70 °C for one-week anaerobic incubation vessels with 0.9 atm nitrogen, 0.05 atm carbon dioxide, 0.7 atm hydrogen, and resulting colonies were streaked for isolation three times.

Characterization of *Thermodesulfobacterium auxiliatoris*. Standard growth in liquid cultures was performed in 20 mL serum vials with 10 mL of GBS salts medium as described for *Caldatribacterium* but without added sugar and yeast extract, with 1 mM each of sodium bicarbonate, sodium acetate, and sodium thiosulfate, and with 1 atm nitrogen headspace. Various substrates tested were added to the indicated final concentrations (**Supplementary Data 1**) from concentrated stocks except for elemental sulfur, which was added to medium and subjected to three cycles of heating to 100 °C for 30 minutes and cooling. For testing electron acceptors, a sulfate-free medium was made by excluding sodium sulfate from the GBS salts and replacing the magnesium sulfate heptahydrate with magnesium chloride hexahydrate and the zinc sulfate in the trace metal solution with zinc chloride. Cultures were incubated for up to one week at 70 °C. Growth temperatures were tested from 45-85 °C in 5 °C intervals, and pH was tested at 5.0-9.0 in 0.5 intervals, with samples taken over a 24-hour period after growth was initially observed. Growth was monitored by turbidity and direct microscopic counts.

Taxonomy of *Thermodesulfobacterium auxiliatoris*. Several lines of evidence justification of *T. auxiliatoris* as a distinct species in the genus *Thermodesulfobacterium*. *T. auxiliatoris* has between 80.1-75.8% genomic ANI and 98.42-94.51% 16S rRNA gene identity to other *Thermodesulfobacterium* species, with the highest identity to *Thermodesulfobacterium hveragerdense* DSM 12571^T in both cases. Additionally, the Genome Taxonomy Database release 220 classifies the *T. auxiliatoris* (strain TA1) genome GCF_008630935.1 (Genbank accession number CP043908) as a distinct species within the genus, s__Thermodesulfobacterium sp008630935.

Supplementary Note 2. Additional information on the isolation, characterization, and identification of a *Thermodesulfovibrio yellowstonii* strain from Inyo-BLM 1. A strain of *Thermodesulfovibrio* was isolated from SIUC xylitol cultures using the same medium described above for isolation of *Thermodesulfobacterium* except that 5 mM lactate was included in the medium. For characterization of growth on lactate, H_2, and acetate in liquid culture, it was grown as described above for *T. auxiliatoris*. The 1.946 Mb MAG from the xylitol metagenome is fragmented, containing 3298 contigs with an N50 of 575 and an average coverage of 2.3-fold; however, the ANI of this MAG to *Thermodesulfovibrio yellowstonii* DSM 11347^T was 98.55% (http://enve-omics.ce.gatech.edu/ani/). The 16S rRNA gene sequence of the isolate was 100% identical to that of *Thermodesulfovibrio yellowstonii* DSM 11347^T, thus it likely represents a member of this species.

Supplementary Note 3. Additional discussion on evidence for the genus *Caldatribacterium* and family *Caldatribacteraceae*. The genomes from the two isolates, *C. saccharofermentans*

GBS^T and *C. inferamans* SIUC1^T, were identified by GTDB-Tk v. 2.3.2[1] as *C. saccharofermentans* and a member of the genus *Caldatribacterium*, respectively. Six additional genomes in GTDB RS220[2] are assigned to the genus *Caldatribacterium* and each is considered a different species within GTDB. Those six genomes include the following: the metagenome-assembled genome from *C. saccharofermentans* designated OP9-77CS (GCA_000347575.1) from an in situ corn stover enrichment in GBS; a composite single-amplified genome from *C. californiese* designated OP9-cSCG (GCA_000353875.1) from Little Hot Creek in California, USA; a MAG from an alkaline hot spring in Yellowstone National Park, USA (GCA_028275365.1); a MAG from the Shengli Oilfield in Shangdong Province, China (GCA_014359405.1); a MAG from a brown biofilm from alkaline hot spring in Sunkai, Malaysia (GCA_025060675.1); and a MAG reportedly from seawater collected by researchers at Shandong University, China (GCA_937922705.1). Our ANI results confirmed that strain *C. saccharofermentans* GBS^T and the MAG *C. saccharofermentans* OP9-77CS belong to the same species based on 99% ANI[3] (**Fig. 4b**) and that *C. inferamans* SIUC1^T and all other genomes assigned by GTDB to the genus *Caldatribacterium* are different species based on ANI values <95%[3] (**Fig. 4b**). AAI values supported the assignment of these eight genomes into the genus *Caldatribacterium*, and the distinctness of this genus, based on AAI values ≥66% between these genomes and ≤53% between these genomes and other genomes assigned to the GTDB family *Caldatribacteriaceae*[3].

To assess the genus *Caldatribacterium* and family *Caldatribacteriaceae* independently and thoroughly, we conducted a phylogenomic analysis using entire protein sequences using the bac120 marker protein set. Proteins were identified, aligned, and refined using GTDB-tk v2.3.2[1]. The alignment was subjected to model testing using the best-fit model identified by ModelFinder[4] and maximum-likelihood analysis of the alignment was conducted using IQTree v. 2.2.6[5], and

branch support calculated with ultrafast bootstrapping (UFboot2) and Shimodaira-Hasegawa approximate Likelihood Ratio Tests (SH-aLRT) from 1,000 replicates. The resulting phylogenetic tree was visualized using iTOL[6]. Our phylogenetic analysis confirmed the monophyly of the genus *Caldatribacterium* and family *Caldatribacteriaceae* as designated by GTDB with strong UFboot2 and SH-aLRT support, which was also recently reported in the literature[7]. This analysis excluded three other GTDB genera assigned to the family *Caldatribacteriaceae* within GTDB and supported the monophyly of the genus designated UBA3950.

Supplementary Note 4. Additional characterization of *Caldatribacterium* isolates.

Taxonomic notes. Comparison of the full-length *C. saccharofermentans* GBS^T and *C. inferamans* SIUC1^T 16S rRNA gene sequences revealed 98.9% identity. Pairwise average nucleotide identity (ANI) values of their complete genomes were below threshold for species identity (94.6-91.9%).

Optimal temperature, pH, and substrate utilization. Strains GBS^T and SIUC1^T optimal growth temperature and pH were determined in GBS medium containing 0.01% yeast extract and 0.05% fucose or 10 mM xylitol, respectively. Growth was assessed by direct microscopic enumeration using a Petroff-Hausser Counting Chamber. Specific growth rates (hr^{-1}) were assessed during the exponential growth phase for each strain (**Supplementary Fig. 8**). Strain GBS^T grew from 60 ℃ and 77.5 ℃, with an optimum of 70 ℃ and at pH values from 5. 5 and 9, with an optimum of 7. These data agree with temperatures at which *Caldatribacterium* has been enriched in GBS[8,9] and the circumneutral pH of GBS[10]. Strain SIUC1^T grew from 55 to 75 ℃, with an optimum of 60-70 ℃, and at pH values between 5.9 and 9.5, with an optimum of 7.4-8.7, which

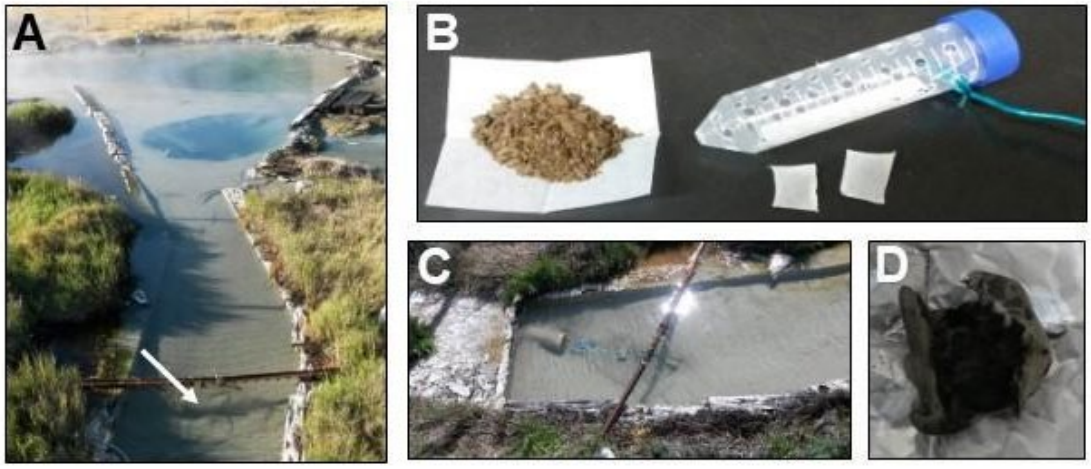
is generally consistent with measurements of Inyo-BLM 1 borehole. Under optimum conditions the doubling times of GBS^T and SIUC1^T are 4.5 hours and 7.6 hours.

Strains GBS^T and SIUC1^T used a variety of fermentation substrates (**Table 1 and Supplemental Table S1**). Growth on simple carbohydrates was predicted by previous genomic analysis[8], and both strains grew on a wide variety of sugars and sugar alcohols, although SIUC1^T utilized complex organic substrates (yeast extract, peptone) and pentoses (xylose, arabinose) more readily than GBS^T. Lithotrophic or autotrophic metabolisms were not observed to support growth for either strain. Strain SIUC1^T grew on trehalose or allantoin while strain GBS^T did not. Use of allantoin as a carbon substrate may be indicative of deep subsurface lifestyles of some *Atribacterota* as previously observed by Bird *et al*, 2019[11] in Baltic Sea marine sediments and Vuillemin *et al*, 2020[12] in Atlantic Abysmal seafloor.

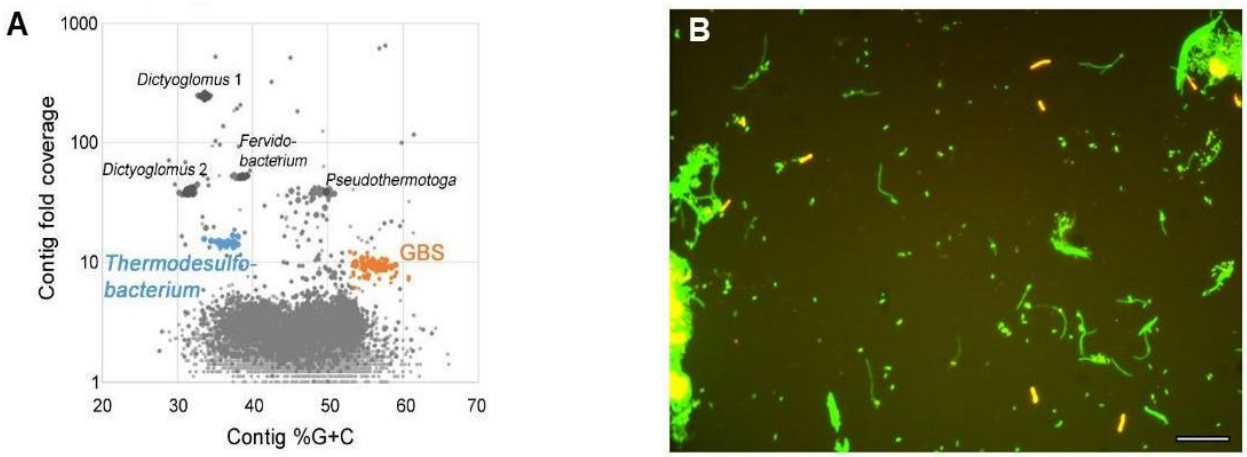
Additional notes on motility and cell ultrastructure. Detached flagella were commonly observed in cryo-EM images of strain *C. inferamans* SIUC^T (11 of 34 imaged fields; **Supplementary Data 8**; **Supplementary Fig. 17**). Flagella measured 17.7 +/- 1.6 nm in diameter. Flagellar motility was observed occasionally in both strains, but more commonly in SIUC^T. Flagella were not visible in cryo-EM images of *C. saccharofermentans* GBS^T. All *Atribacterota* genomes encode complete or near-complete flagellar gene clusters, except for several genomes assigned to the genus *Atribacter* (**Supplementary Data 4**). The presence of genes encoding FlgH and FlgI, forming the L- and P-rings, favor a typical Gram-negative cell envelope structure at the locations where the flagella are inserted. However, since flagella were detached, we don't know the location. We note that *Thermotoga maritima* possesses a lateral flagellum[13], and the width of the periplasm at lateral locations is ~70 nm[14]. In *Caldatribacterium* cells, the distance between the internal LML and the cell surface is similar but highly variable (67.8 +/- 31.0 nm in *C.*

174 *saccharofermentans* GBS^T and 62.1 +/- 27.8 nm in *C. inferamans* SIUC^T), typically with shorter
175 distances in pre-divisional cells and longer distances in more mature, elongated cells
176 (**Supplementary Data 7; Supplementary Data 8**). Distances from the inner LML to cell tips
177 were 372 +/- 46.9 nm in *C. saccharofermentans* GBS^T and 368 +/- 85.1 nm in *C. inferamans*
178 SIUC^T. The width of the space between the two membranes at the cell surface was 9.7 +/- 1.1 nm
179 in *C. saccharofermentans* GBS^T and 10.9 +/- 1.9 nm in *C. inferamans* SIUC^T, at the narrow end
180 of measurements in *E. coli* that range from 10-70 nm[15].

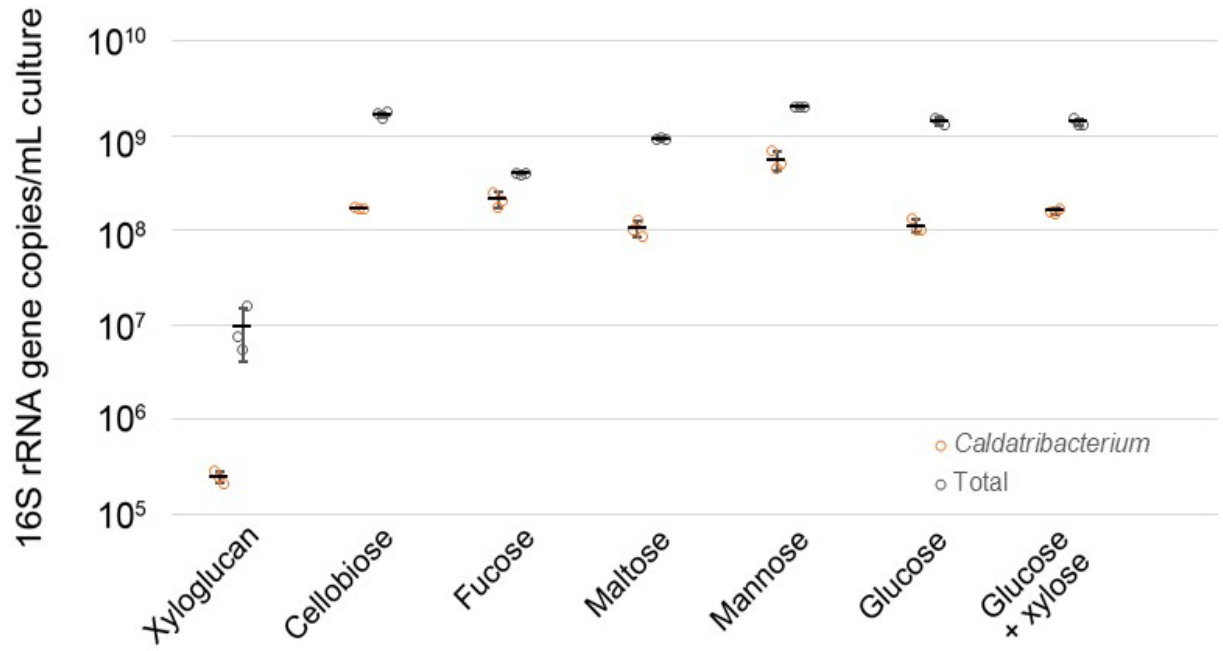
181 **Membrane core lipids.** Lipid fatty acids of the cultured *Atribacteria* species have are
182 shown in **Supplementary Data 10**. Strain SIUC1^T, when grown at 65 ºC, has a fatty acid
183 composition of 33.3% C_{16} and 38.9% C_{18}, with minor populations of C_{15iso} and C_{13iso} at ~11%.
184 Strain GBS^T when grown at 73 ºC has a fatty acid composition of 37.2% of C_{16} and 30% C_{13iso}
185 with 10.1% C_{18}. *A. laminatus*, which grows optimally at 45 ºC, has a fatty acid composition of
186 51% of C_{15} and 21% C_{18} with 10% C_{16} and C_{15iso}. In general, *A. laminatus* has shorter fatty acid
187 chains, which is consistent with patterns seen in less thermophilic bacteria[16,17].
188



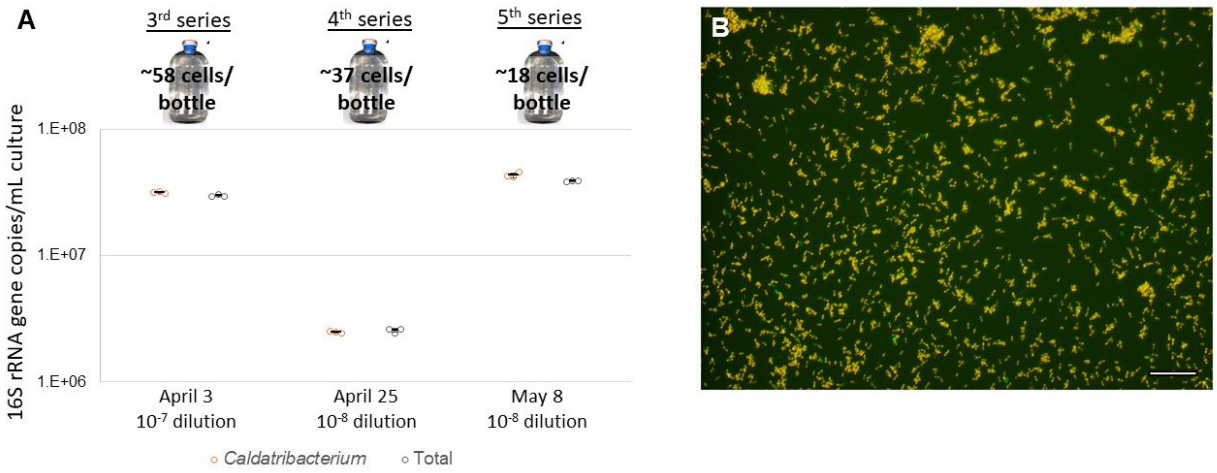
Supplementary Fig. 1. In situ corn stover enrichments in GBS. A) Great Boiling Spring (GBS) near Gerlach, NV, USA, with an arrow indicating where the in situ enrichments were incubated in the outflow. B) Corn stover and enrichment set-up before deployment. C) Corn stover in situ enrichment incubating buried in ~1 cm of sediment in GBS outflow. D) Corn stover after 5 months of incubation (Oct 2013 - March 2014).



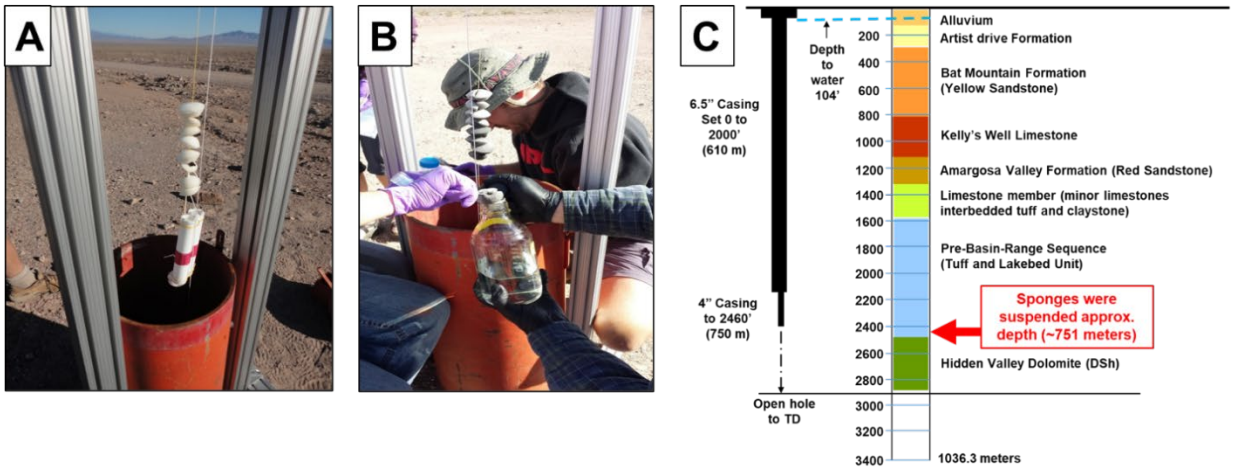
Supplementary Fig. 2. Xyloglucan enrichment metagenome and FISH. (A) Contigs in 2016 xyloglucan culture. Clusters of contigs corresponding to MAGs are indicated, with *C. saccharofermentans* and *Thermodesulfobacterium* MAGs shown in orange and blue, respectively. (B) Fluorescence in situ hybridization of fixed cells from a sample of the xyloglucan culture, overlay of images showing fluorescence with probes for *Caldatribacterium* (Cy3, red/orange) and total bacteria (6FAM, green). Scale bar is 20 µm. The image is representative of ~20 fields examined. Source data are provided as a Source Data file.



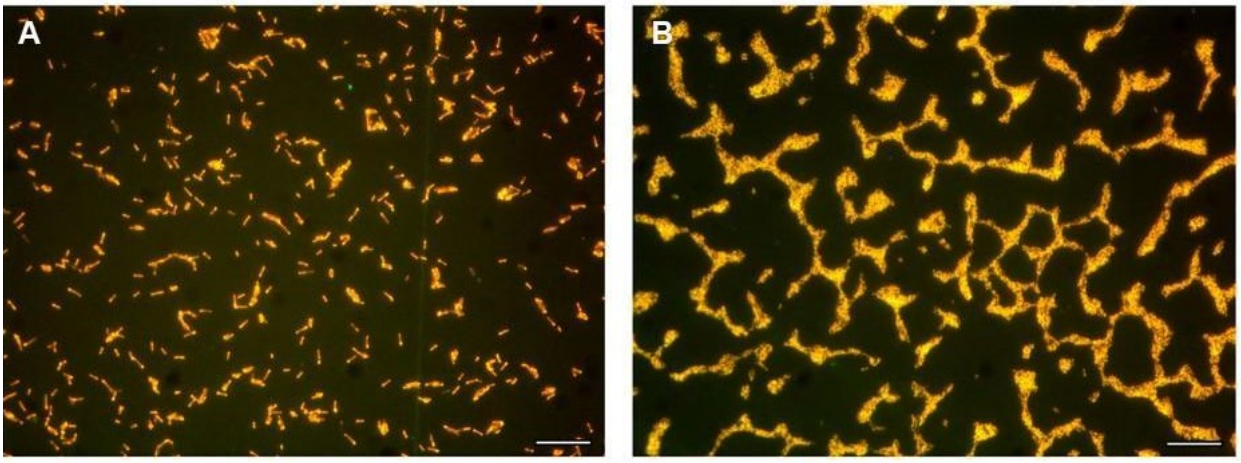
Supplementary Fig. 3. Abundance of *Caldatribacterium* in GBS enrichments grown with different substrates. Abundance of *Caldatribacterium* and total Bacteria/Archaea in cultures with various sole carbon substrates using qPCR with *Caldatribacterium*-specific primers and "universal" primers ("Total" bacteria and archaea). Black bars show means and standard deviations of replicate qPCRs (n=3) from a single culture. Source data are provided as a Source Data file.



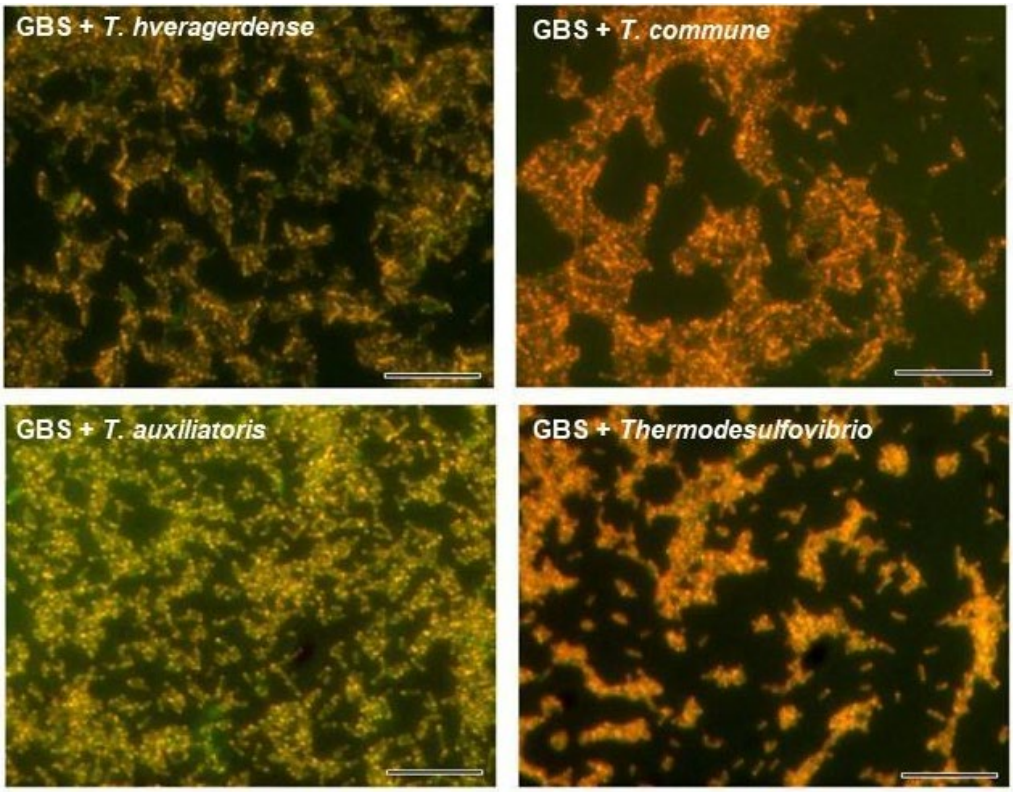
Supplementary Fig. 4. Abundance of *Caldatribacterium* following dilution-to-extinction of GBS fucose enrichments. (A) Abundance of *Caldatribacterium* and total Bacteria/Archaea in dilution-to-extinction cultures using qPCR with *Caldatribacterium*-specific primers and "universal" primers ("Total" bacteria and archaea). Black bars show means and standard deviations of replicate qPCRs (n=3) from a single culture. (B) Fluorescence in situ hybridization of fixed cells from a sample of the fucose culture after dilution-to-extinction, overlay of images showing fluorescence with probes for *Caldatribacterium* (Cy3, red/orange) and total bacteria (FAM, green). Scale bar is 20 µm. The image is representative of ~20 fields examined. Source data are provided as a Source Data file.



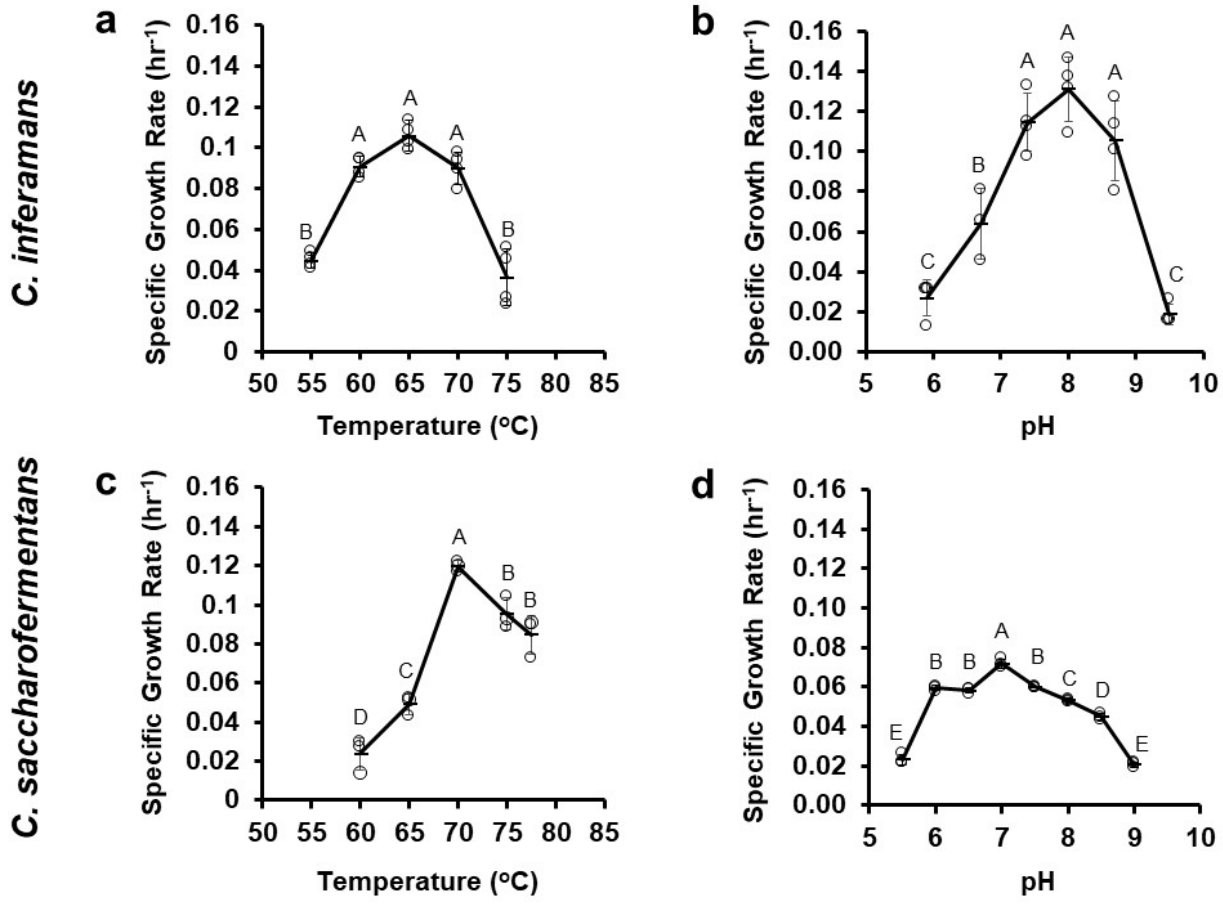
Supplementary Fig. 5. Borehole Inyo-BLM 1B and polyurethane foam plugs used for in situ cultivation. A) Sterile foam plugs ("sponges") being lowered into Inyo-BLM1B in November 2014. B) Sponges were retrieved in February 2015 and quickly stored in premade reduced artificial groundwater media (AGM). (C) Borehole Inyo-BLM1 well design and geological cross section. Total depth at 1036 m before open wall collapse, and relative location below the water table. Location is 36º 24'04.19 N / 116º 28'06.58 W. Image modified from Bredehoeft J, *et al*, 2005.



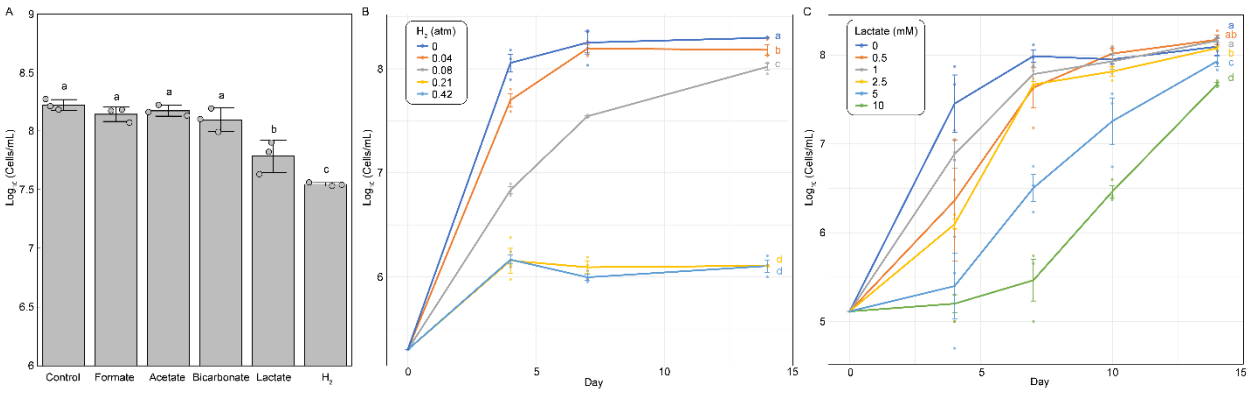
Supplementary Fig. 6. FISH images showing purity of axenic cultures. (A) Strain GBS^T and (B) SIUC1^T. Overlays of images showing fluorescence with probes for *Caldatribacterium* (Cy3, red/orange) and total bacteria (FAM, green). Scale bars are 20 μm. The images are representative of ~20 fields examined.



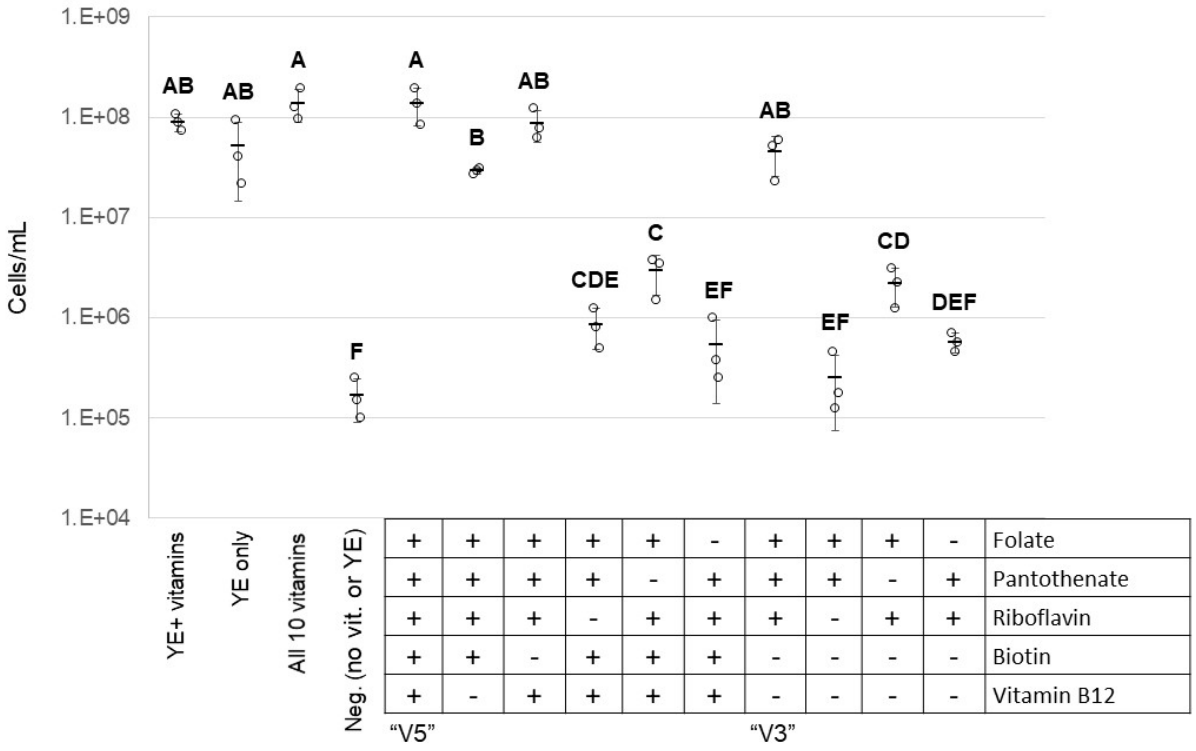
Supplementary Fig. 7. FISH images of strain GBS and various partner SRBs. Representative FISH results on co-cultures showing *C. saccharifermentans* GBS^T cells (GBS, Cy3-labeled *Caldatribacterium* probe, orange cells) and sulfate-reducer partner cells (bacterial 6FAM-labeled probe, green cells), showing that the vast majority of cells in the co-culture are *Caldatribacterium*. SRBs: *Thermodesulfobacterium hveragerdense* DSM 12571^T; *Thermodesulfobacterium commune* DSM 2178^T; *Thermodesulfobacterium auxiliatoris*; and *Thermodesulfovibrio yellowstonii*. Scale bars represent 20 µm. The images are representative of ~20 fields examined.



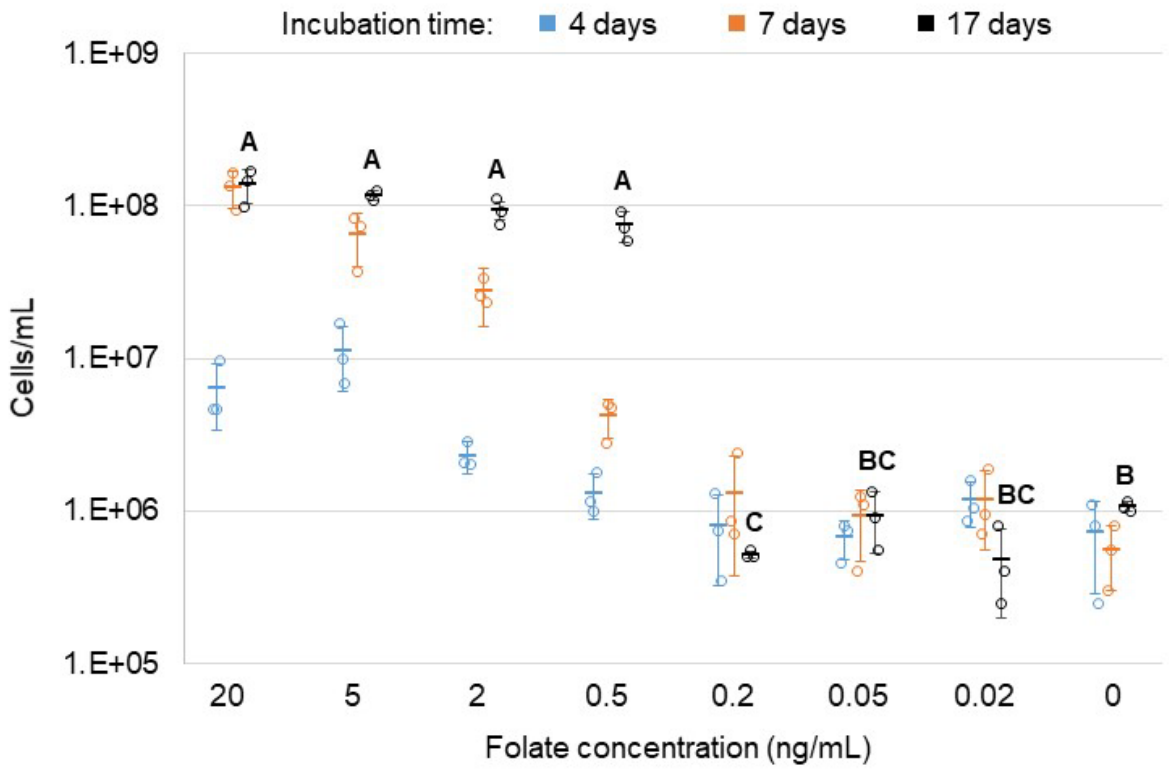
Supplementary Fig. 8. **Growth optimization of *Caldatribacterium* isolates.** Growth ranges and optima for *C. inferamans* strain SIUC1^T (panels a and b) and *C. saccharofermentans* strain GBS^T (panels c and d). Bars show the mean and standard deviation (n=4 for *C. inferamans*, n=3 for *C. saccharofermentans*). Different capital letters show experimental groups that are significantly different ($p < 0.05$; ANOVA with post-hoc Tukey's). Source data are provided as a Source Data file.



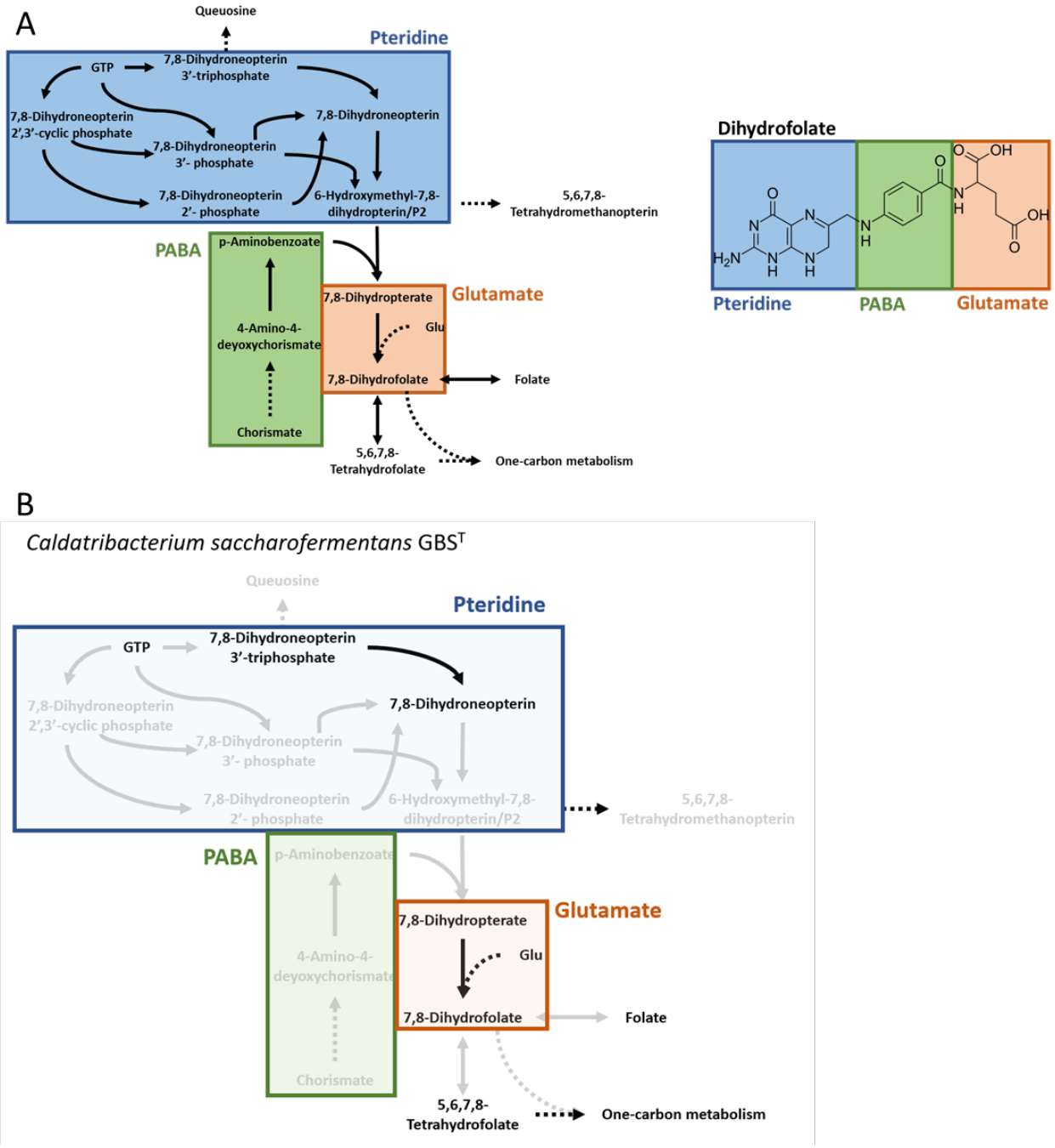
Supplementary Fig. 9. (A) Growth of *Caldatribacteirum saccharofermentans* strain GBS^T in the presence of potential fermentation products at 1 mM or 0.08 atm (for H_2) after seven days of incubation. Data with different letters are significantly different ($p < 0.05$; One-way ANOVA followed by Tukey's post-hoc test). Growth of strain GBS^T at different levels of (B) hydrogen or (C) lactate over a time course of two weeks. Different letters indicate significant differences ($p < 0.05$; Student's t-test). All bars/points show averages and standard deviations (n=3). Source data are provided as a Source Data file.



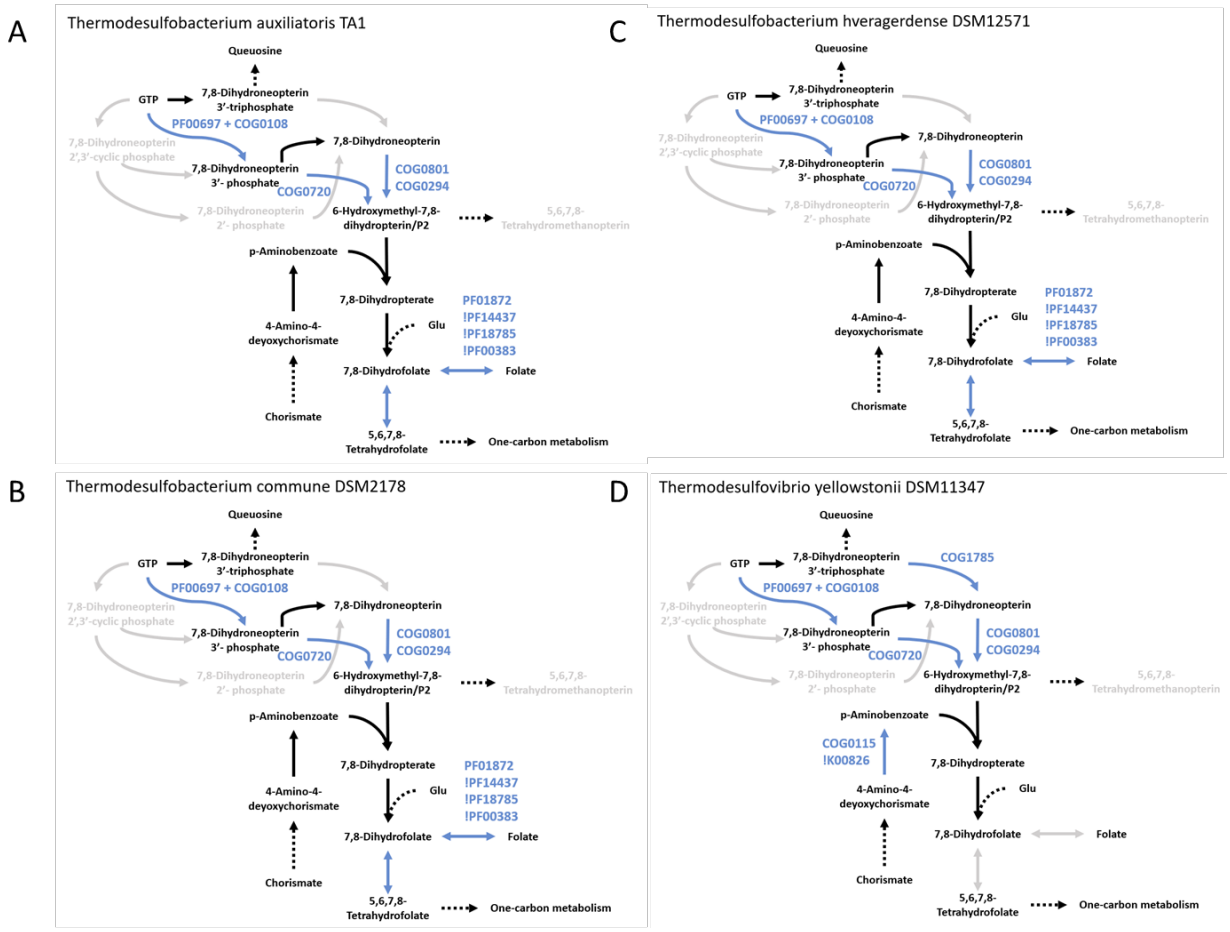
Supplementary Fig. 10. Growth of *C. saccharofermentans* GBS^T in the presence or absence of yeast extract and various autoclaved vitamins. Circles show individual replicates and bars show mean and standard deviation (N=3). Letters indicate significantly different ($p < 0.05$) means based on ANOVA and Tukey's post-hoc tests. Source data are provided as a Source Data file.



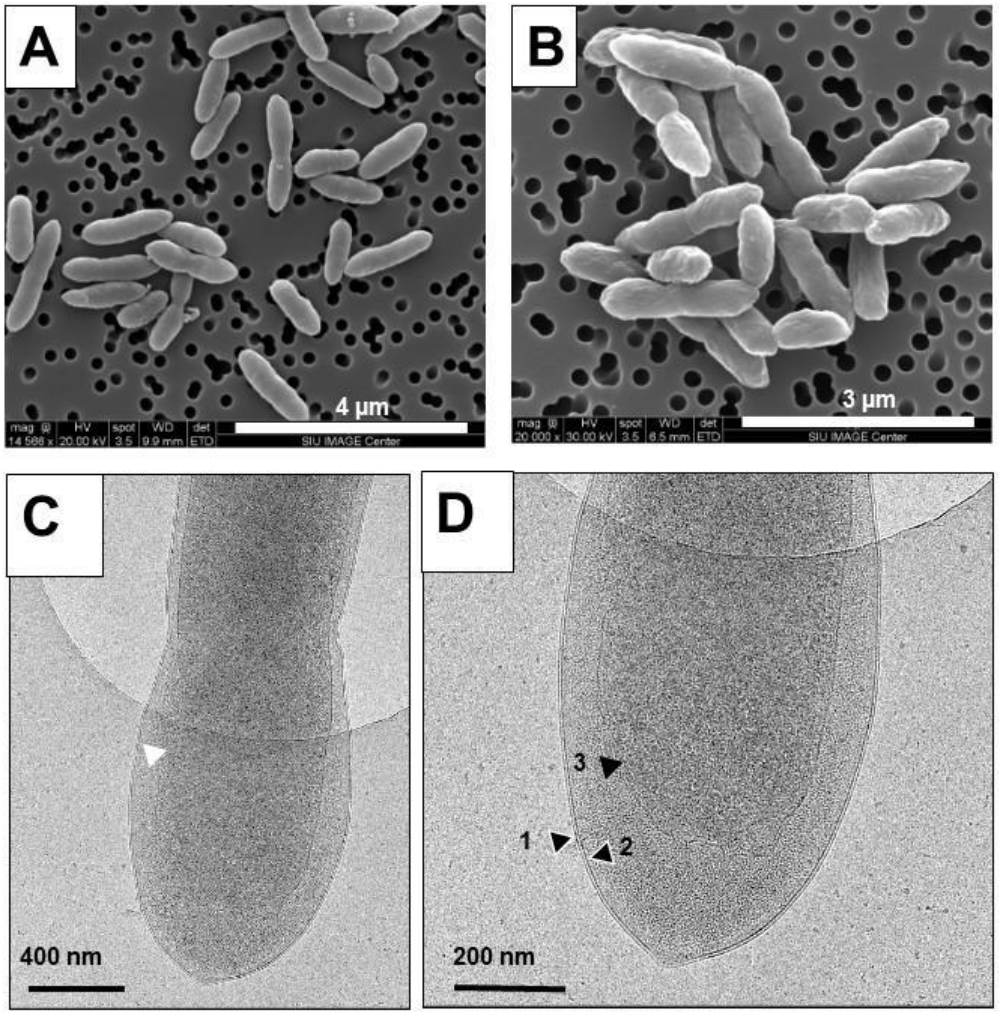
Supplementary Fig. 11. Growth of *C. saccharofermentans* GBS^T at different concentrations of folate. Different concentrations of folate, decreasing from the standard 20 ng mL^{-1}, were added along with standard concentrations of riboflavin, biotin, pantothenate, and vitamin B12 used in the autoclaved vitamin mix. Circles show individual replicates and bars show mean and standard deviation (N=3). Letters indicate significantly different ($p < 0.05$) means at the final (17 day) time point based on ANOVA and Tukey's post-hoc tests. Source data are provided as a Source Data file.



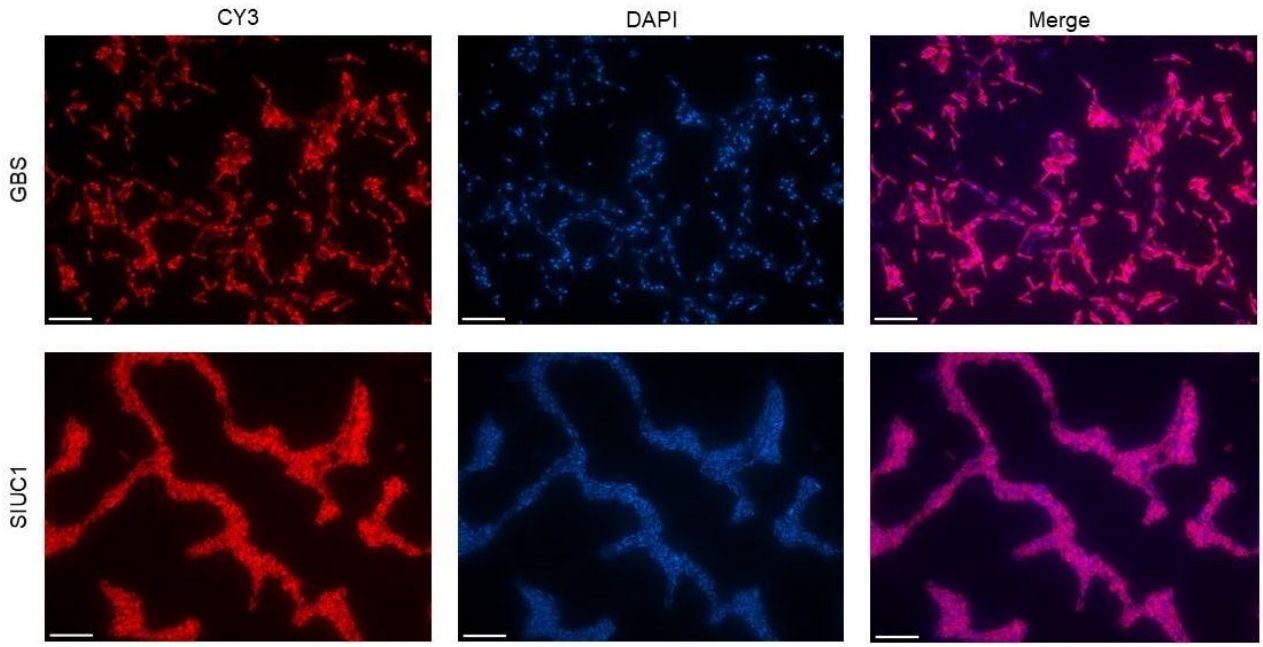
Supplementary Fig. 12. Folate biosynthetic pathway in *Caldatribacterium saccharofermentans* GBS^T. Folate biosynthetic pathway showing incompleteness of pterin and pABA branches but presence folylpolyglutamyl synthase (FPGS) necessary for glutamylation reactions that would be necessary for use of transported folate. The same enzymes are present in *Caldatribacterium inferamans* SUIC1^T. Black, annotated enzymes; grey, not annotated. Annotations are provided in **Supplementary Data 4**.



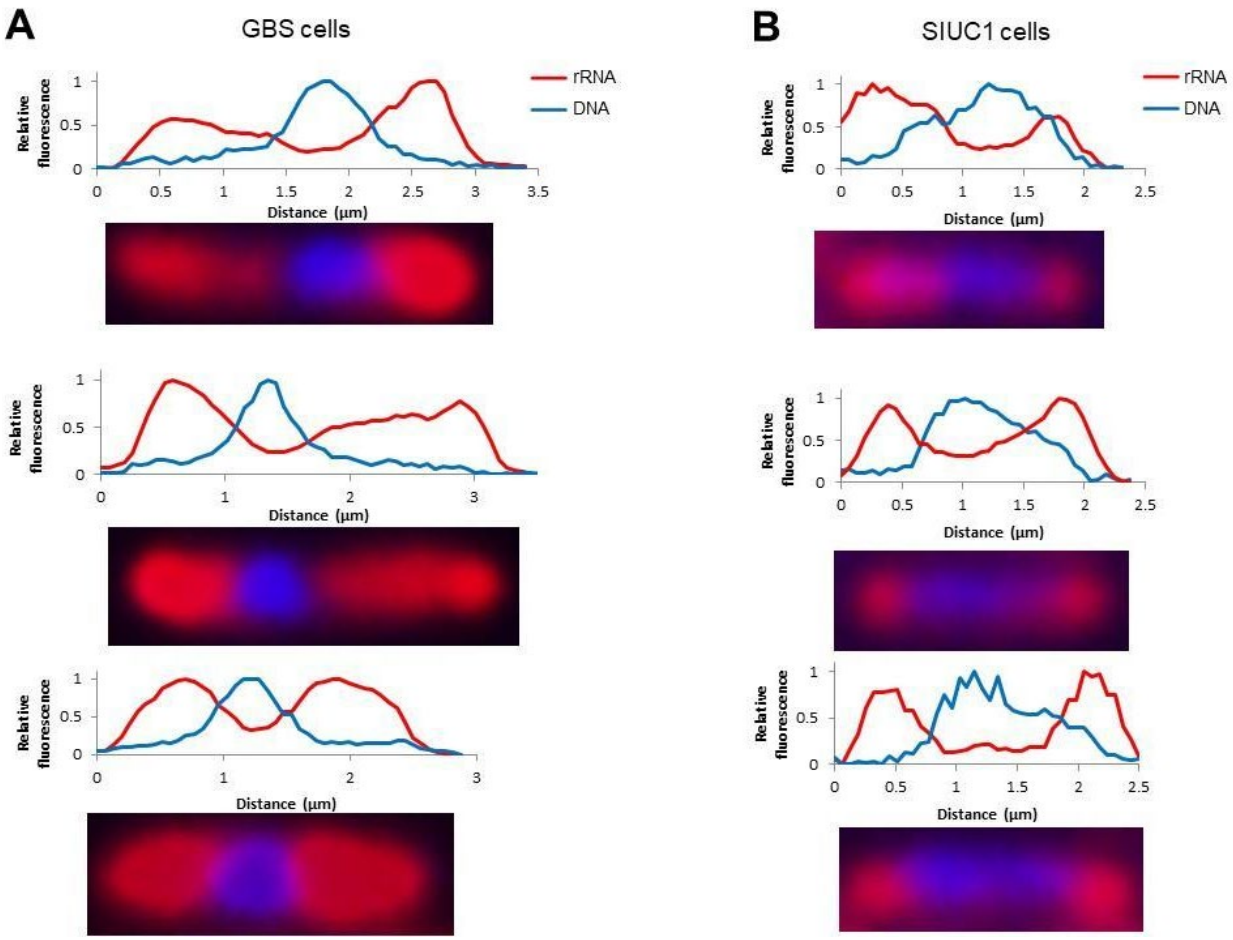
Supplementary Fig. 13. Folate biosynthetic pathway in SRBs. Folate biosynthetic pathway showing completeness of folate biosynthetic pathways in the SRBs used in the experiments. Black, annotated enzymes; blue, annotated using COG or pfam; grey, not annotated. Blue pfams preceded by exclamation points are absent, consistent with our annotation criteria (see Methods). KO annotations for folate and other vitamin synthesis pathways are provided in **Supplementary Data 4**.



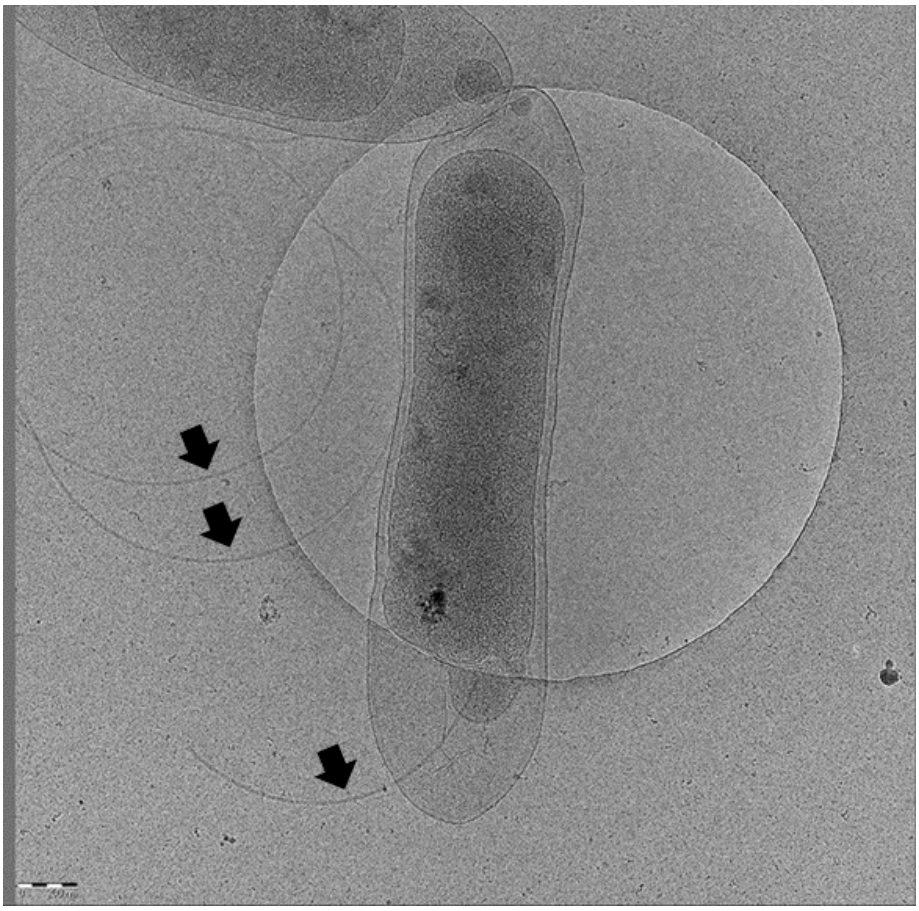
Supplementary Fig. 14. Electron micrographs of *C. saccharofermentans* GBS^T and *C. inferamans* SIUC1^T. (A) Scanning electron micrograph (SEM) of *C. inferamans* SIUC1^T and (B) SEM of *C. saccharofermentans* GBS^T. (C) Cryo-electron micrograph of GBS^T cell with white arrow indicating denser inner region; (D) Close-up of panel C with dark arrows indicating 1. outer, 2. middle, and 3. inner lipid-like membrane layers. Scale bars indicated in each panel. The images are representative of ~20 fields examined.



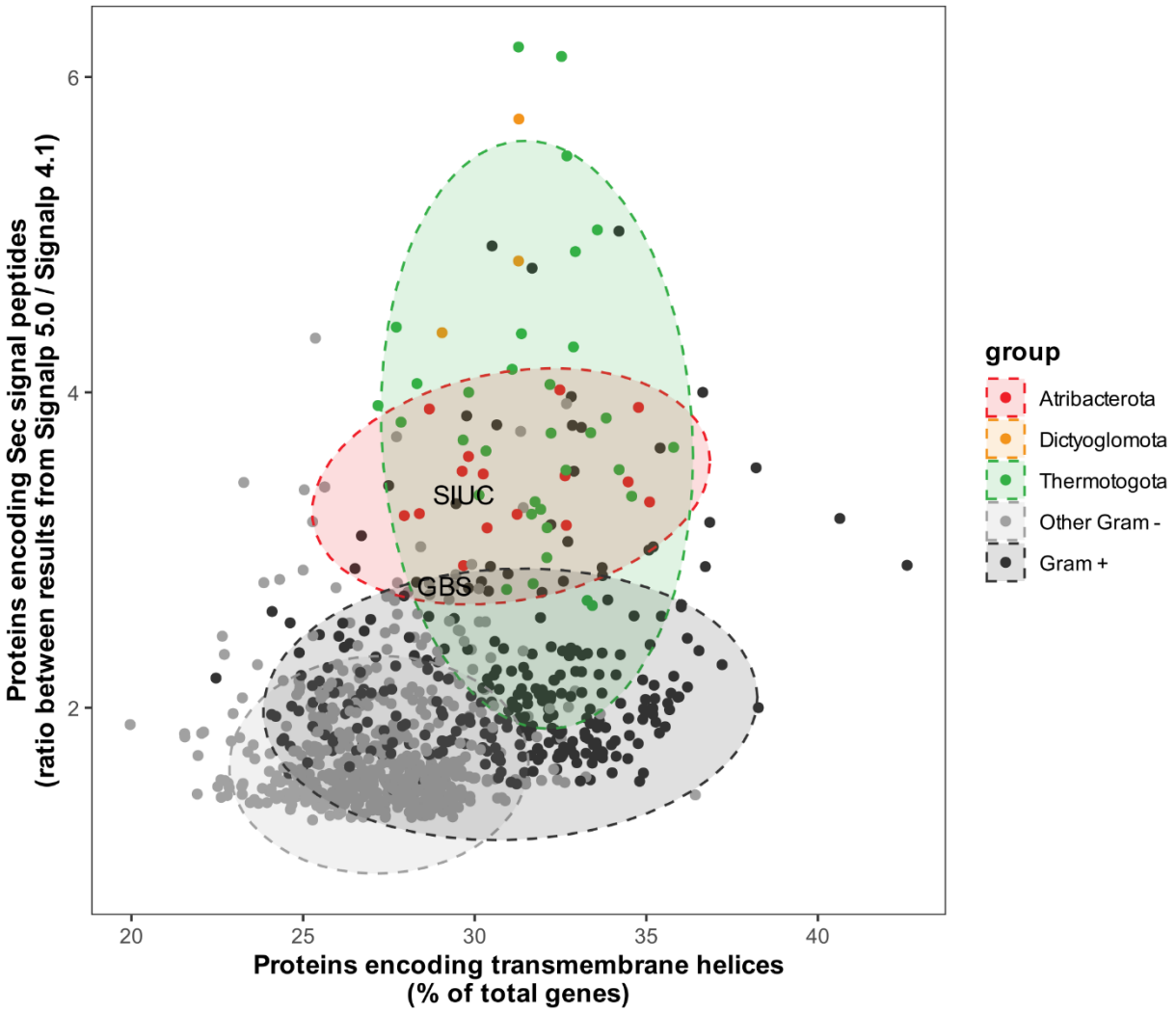
Supplementary Fig. 15. Fluorescence microscopy of GBS^T and SIUC1^T showing rRNA and DNA staining and localization. Fixed cells were hybridized with *Caldatribacterium*-specific rRNA probe and counterstained with DAPI. Scale bars are 10 µm. The images are representative of ~20 fields examined.



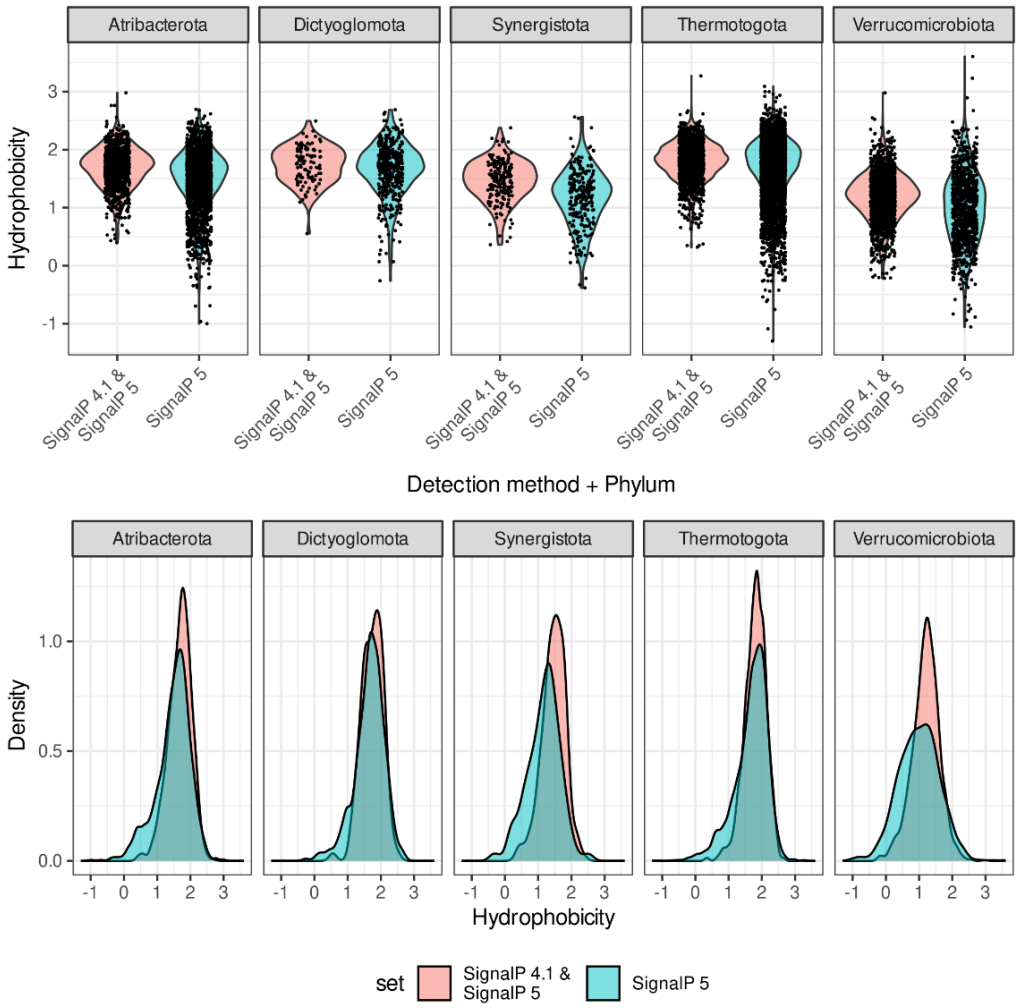
Supplementary Fig. 16. Profiles of localization of DNA and rRNA in cells of strains GBS^T and SIUC1^T. (A) GBS^T and (B) SIUC1^T subjected to FISH with a Cy3-labeled probe targeting *Caldatribacterium* rRNA and DAPI counterstain. The images and profiles are representative of ~20 cells examined for each strain.



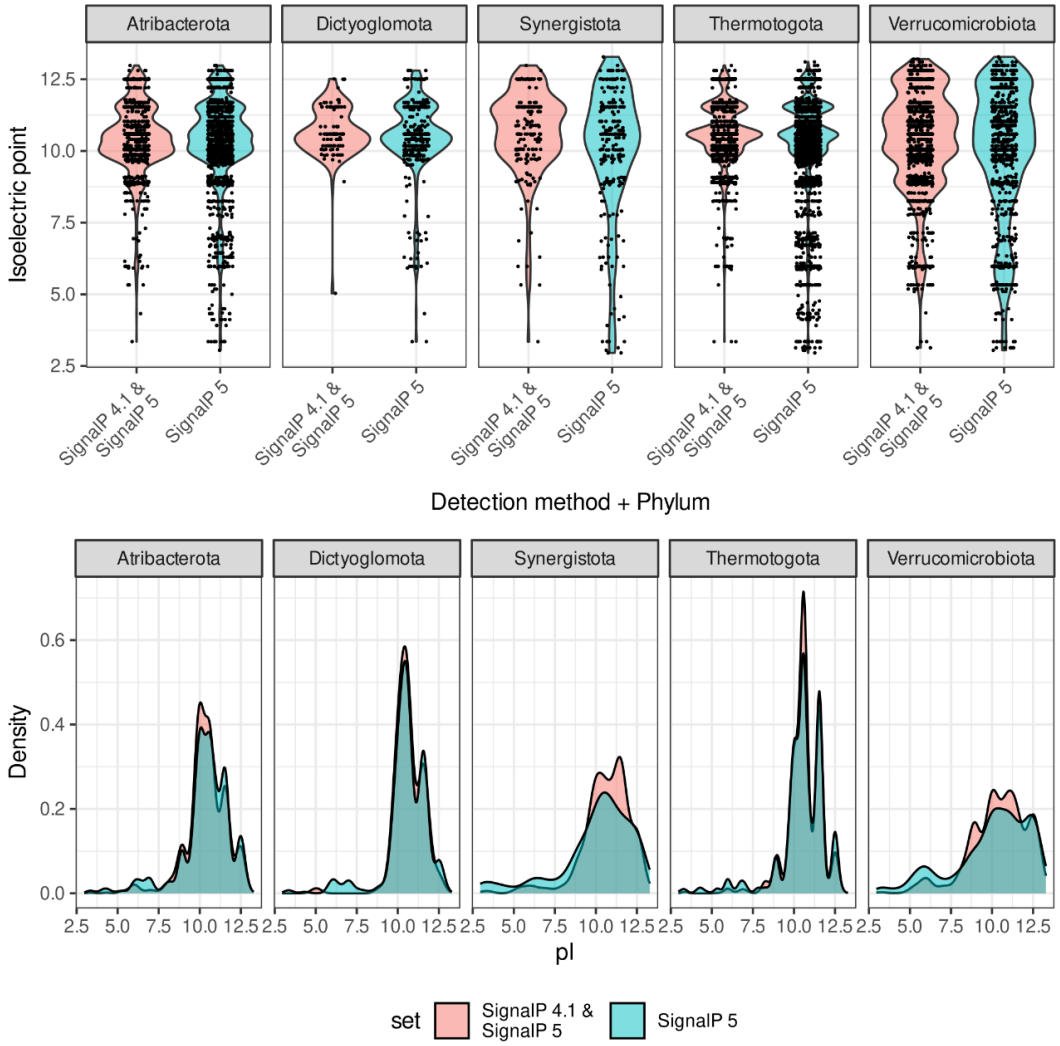
Supplementary Fig. 17. Detached flagella in cryo-EM image of SIUC1^T. Several detached flagella are shown (black arrows). Detached flagella were visible in 11 of 34 imaged fields from two different grids. Attached flagella were not observed. Flagella were not observed in cryo-EMs of strain GBS^T. The images and profiles are representative of >15 detached flagella. Raw data on measurements of flagella are available in **Supplementary Data 8**.



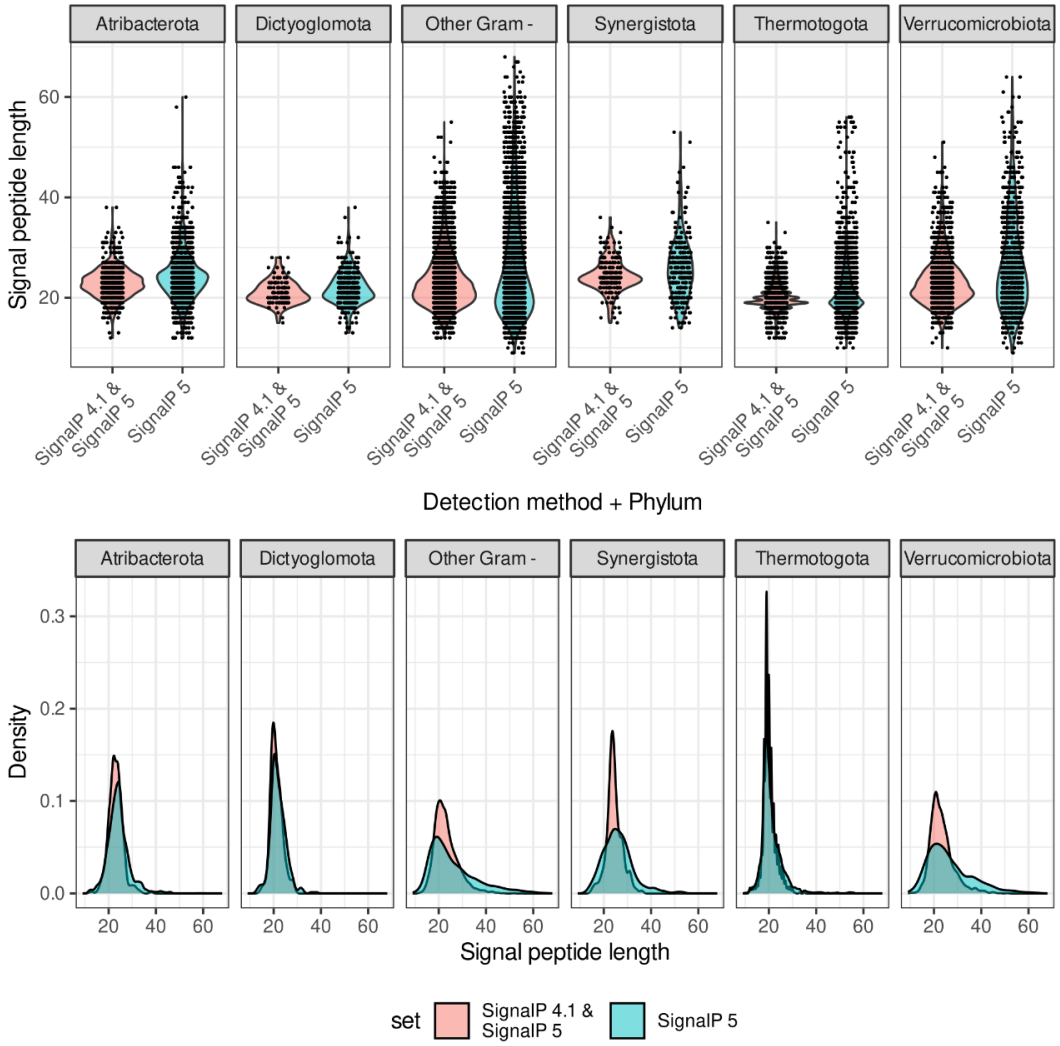
Supplementary Fig. 18. Unique profiles of predicted secreted and membrane proteins in *Atribacterota*. Plot based on Katayama et al., 2020. *Atribacterota* proteomes have a higher ratio of Sec sequences predicted by SignalP-5.0 over SignalP-4.1 compared to other Gram-negative bacteria (P = 1.87e-6) and Gram-positive bacteria (P = 1.08e-6), but not compared to *Thermotogota* (P = 0.416) or *Dictyoglomota* (P = 0.136). *Thermotogota* and *Dictyoglomota* were noted as outliers previously (Katayama et al., 2020). *Atribacterota* proteomes also have a higher percentage of transmembrane proteins compared to other Gram-negative bacteria (P = 6.51e-7), but not Gram-positive bacteria (P = 0.710), *Thermotogota* (P = 0.548), or *Dictyoglomota* (p = 0.722). Statistical significance was assessed via Wald test. Source data are provided as a Source Data file.



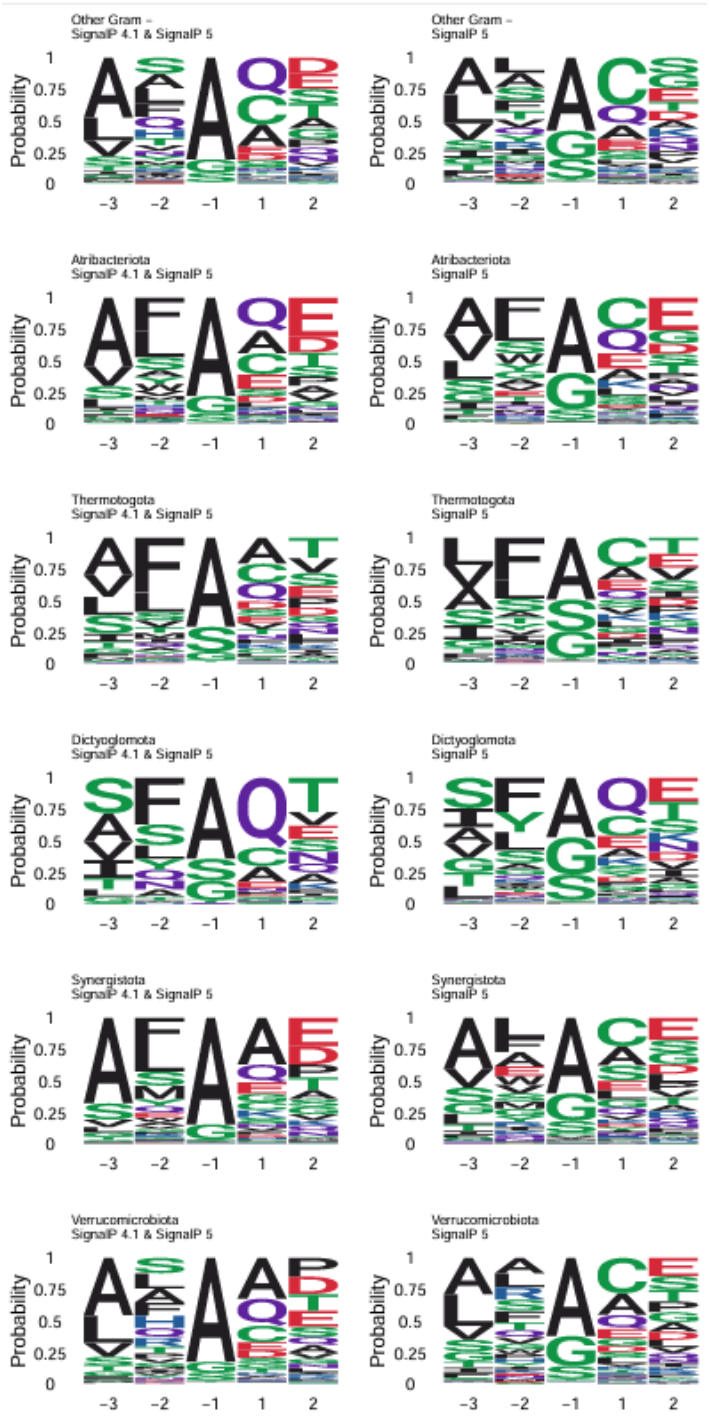
Supplementary Fig 19. SignalP-5 improves predictions of *Atribacterota* secreted proteins with signal peptides that are less hydrophobic. Top, hydrophobicity of signal peptides predicted by SignalP-4.1 and SignalP-5 versus those predicted only by SignalP-5. Differences between SignalP versions were detected. Bottom, histograms showing the distributions of predicted hydrophobicity for signal peptides predicted by SignalP-4.1 and SignalP-5 versus those predicted only by SignalP-5. *Atribacterota* signal peptides annotated by SignalP-4.1 and SignalP-5 were more hydrophobic than those annotated by SignalP-5 alone (difference = 0.21 Kyte-Doolittle units, $p < 1e\text{-}7$; permutation test). The predicted hydrophobicity of *Atribacterota* signal peptides detected only by SignalP-5 were lower than those from *Dictyoglomota* (difference = 0.17 Kyte-Doolittle units, $p = 0$; ANOVA + Tukey's HSD) and *Thermotogota* (0.19 Kyte-Doolittle units, $p < 1e\text{-}7$), but were higher than those from *Synergistota* (0.31 Kyte-Doolittle units, $p < 1e\text{-}7$), and *Verrucomicrobiota* (0.47 Kyte-Doolittle units, $p < 1e\text{-}7$). The predicted hydrophobicity of *Atribacterota* signal peptides detected by both SignalP-5 and SignalP-4.1 was lower than *Thermotogota* (0.11 Kyte-Doolittle units, $p = 0$), higher than those from *Synergistota* (0.22 Kyte-Doolittle units, $p = 0$) and *Verrucomicrobiota* (0.48 Kyte-Doolittle units, $p < 1e\text{-}7$), but did not differ from *Dictyoglomota* (0.05 Kyte-Doolittle units, $p = 0.65$). Source data are provided as a Source Data file.



Supplementary Fig 20. SignalP-5 improves predictions of *Atribacterota* secreted proteins with signal peptides with a lower pI. Top, predicted isoelectric point (pI) of signal peptides predicted by SignalP-4.1 and SignalP-5 versus those predicted only by SignalP-5. Bottom, histograms showing the distributions of pI values for signal peptides predicted by SignalP-4.1 and SignalP-5 versus those predicted only by SignalP-5. *Atribacterota* signal peptides annotated by SignalP-4.1 and SignalP-5 were higher in pI compared to those annotated by only SignalP-5 (difference = 0.21 pH units, p = 0; permutation test). The predicted pI of *Atribacterota* signal peptides detected only by SignalP-5 were lower than those from *Thermotogota* (difference = 0.12 pH units, p = 0.031; ANOVA + Tukey's HSD), higher than those from *Verrucomicrobiota* (0.32 pH, p = 8e-6), trended higher than those from *Synergistota* (0.30 pH, p = 0.053), and were not different from *Dictyoglomota* (0.15 pH, p = 0.39). The predicted pI of *Atribacterota* signal peptides detected by both SignalP-5 and SignalP-4.1 was lower than *Thermotogota* (0.21 pH, p = 0.0037), and not significantly different from *Synergistota* (0.21 pH, p = 0.331), *Verrucomicrobiota* (0.04 pH, p = 0.973), or *Dictyoglomota* (0.25 pH, p = 0.397). Source data are provided as a Source Data file.



Supplementary Fig 21. SignalP-5 improves predictions of *Atribacterota* secreted proteins with longer signal peptides. Top, length of signal peptides predicted by SignalP-4.1 and SignalP-5 versus those predicted only by SignalP-5. Bottom, histograms showing the distributions of the lengths of signal peptides predicted by SignalP-4.1 and SignalP-5 versus those predicted only by SignalP-5. Signal peptides annotated by SignalP-4.1 and SignalP-5 were shorter than those annotated by SignalP-5 (difference = 1.18 AA residues, p = 0; permutation test). *Atribacterota* signal peptides detected only by SignalP-5 were shorter than those from *Dictyoglomota* (difference = 2.27 AA residues, p = 1E-7; ANOVA + Tukey's HSD) and *Thermotogota* (3.00 AA, p = 1E-6), and longer than those from *Synergistota* (1.80 AA, p = 0.003), *Verrucomicrobiota* (difference = 2.4 AA, p = 0), or other Gram-negative bacteria (1.83 AA, p = 0). *Atribacterota* signal peptides detected by both SignalP-5 and SignalP-4.1 were shorter than those from *Dictyoglomota* (1.81 AA, p = 0.002) and *Thermotogota* (2.73 AA, p = 0), longer than those from *Synergistota* (1.46 AA, p = 0.001), and did not differ in length from *Verrucomicrobiota* (difference = 0.34 AA, p = 0.494) or other Gram-negative bacteria (0.12 AA, p = 0.975). Source data are provided as a Source Data file.



Supplementary Fig 22. Signal peptidase recognition sites in *Atribacterota*. Plot showing probabilities of amino acids at the signal peptidase recognition sites for selected phyla predicted by SignalP-4.1 and SignalP-5 versus those predicted only by SignalP-5. Source data are provided as a Source Data file.

REFERENCES

1. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**, 5315-5316 (2022).

2. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* **50**, D785-D794 (2022).

3. Konstantinidis KT, Rosselló-Móra R, Amann R. Uncultivated microbes in need of their own taxonomy. *ISME J* **11**, 2399-2406 (2017).

4. Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**, 587-589 (2017).

5. Minh BQ*, et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* **37**, 1530-1534 (2020).

6. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res* **52**, W78-W82 (2024).

7. Jiao J-Y*, et al.* Cultivation of novel *Atribacterota* from oil well provides new insight into their diversity, ecology, and evolution in anoxic, carbon-rich environments. *Microbiome* **12**, 123 (2024).

8. Dodsworth JA*, et al.* Single-cell and metagenomic analyses indicate a fermentative and saccharolytic lifestyle for members of the OP9 lineage. *Nat Commun* **4**, 1854 (2013).

9. Peacock JP*, et al.* Pyrosequencing reveals high-temperature cellulolytic microbial consortia in Great Boiling Spring after in situ lignocellulose enrichment. *PLoS One* **8**, e59927 (2013).

10. Costa KC, Navarro JB, Shock EL, Zhang CL, Soukup D, Hedlund BP. Microbiology and geochemistry of great boiling and mud hot springs in the United States Great Basin. *Extremophiles* **13**, 447-459 (2009).

11. Bird JT*, et al.* Uncultured microbial phyla suggest mechanisms for multi-thousand-year subsistence in Baltic Sea sediments. *MBio* **10**, 10.1128/mbio. 02376-02318 (2019).

12. Vuillemin A*, et al.* Atribacteria Reproducing over Millions of Years in the Atlantic Abyssal Subseafloor. *mBio* **11**, 10.1128/mbio.01937-01920 (2020).

13. Gluch MF, Typke D, Baumeister W. Motility and thermotactic responses of Thermotoga maritima. *J Bacteriol* **177**, 5473-5479 (1995).

14. Sexton DL, Hashimi A, Beskrovnaya P, Sibanda L, Huan T, Tocheva EI. The cell envelope of *Thermotogae* suggests a mechanism for outer membrane biogenesis. *Proc Natl Acad Sci* **120**, e2303275120 (2023).

15. Tran BM, Punter CM, Linnik D, Iyer A, Poolman B. Single-protein Diffusion in the Periplasm of *Escherichia coli*. *J Mol Biol* **436**, 168420 (2024).

16. Driessen AJ, Albers SV. Membrane adaptations of (hyper) thermophiles to high temperatures. *Physiology and biochemistry of extremophiles*, 104-116 (2007).

17. Koga Y. Thermal adaptation of the archaeal and bacterial lipid membranes. *Archaea* **2012**, 789652 (2012).