Diversity in CO_2 concentrating mechanisms among chemolithoautotrophs from genera *Hydrogenovibrio, Thiomicrorhabdus,* and *Thiomicrospira,* ubiquitous in sulfidic habitats worldwide

Kathleen M. Scott[1#], Juliana M. Leonard[1], Rich Boden[2], Dale Chaput[3], Clare Dennison[1], Edward Haller[1], Tara L. Harmer[4], Abigail Anderson[1], Tiffany Arnold[1], Samantha Budenstein[1], Rikki Brown[1], Juan Brand[1], Jacob Byers[1], Jeanette Calarco[1], Timothy Campbell[1], Erica Carter[1], Max Chase[1], Montana Cole[1], Deandra Dwyer[1], Jonathon Grasham[1], Christopher Hanni[1], Ashlee Hazle[1], Cody Johnson[1], Ryan Johnson[1], Brandi Kirby[1], Katherine Lewis[1], Brianna Neumann[1], Tracy Nguyen[1], Jonathon Nino Charari[1], Ooreoluwa Morakinyo[1], Bengt Olsson[1], Shanetta Roundtree[1], Emily Skjerve[1], Ashley Ubaldini[1], and Robert Whittaker[1].

[1]Department of Integrative Biology, University of South Florida, Tampa, Florida, United States of America

[2]School of Biological & Marine Sciences, University of Plymouth, Drake Circus, Plymouth, UK; Sustainable Earth Institute, University of Plymouth, Drake Circus, Plymouth, UK,

[3]Proteomics and Mass Spectrometry Facility, University of South Florida, Tampa, Florida, United States of America,

[4]Biology Program, Stockton University, Galloway, New Jersey, United States of America

[#]Address correspondences to Kathleen M. Scott, kmscott@usf.edu

Running title: Diverse CCMs in sulfur-oxidizing autotrophs

ABSTRACT Members of *Hydrogenovibrio, Thiomicrospira* and *Thiomicrorhabdus* fix carbon at hydrothermal vents, coastal sediments, hypersaline lakes, and other sulfidic habitats. The genome sequences of these ubiquitous and prolific chemolithoautotrophs suggest a surprising diversity of mechanisms for dissolved inorganic carbon (DIC) uptake and fixation; these mechanisms are verified here. Carboxysomes are apparent in transmission electron micrographs of most of these organisms; lack of carboxysomes in *Thiomicrorhabdus* sp. Milos T2 and *Tmr. arctica*, and an inability to grow under low DIC conditions by *Thiomicrorhabdus* sp. Milos T2 are consistent with an absence of carboxysome loci in their genomes. For the remaining organisms, genes encoding potential DIC transporters from four evolutionarily distinct families (Tcr0853/0854, Chr, SbtA, SulP) are located downstream of carboxysome loci. Transporter genes collocated with carboxysome loci, as well as some homologs located elsewhere on the chromosomes, had elevated transcript levels under low DIC conditions, as assayed by qRT-PCR. DIC uptake was measureable via silicone oil centrifugation when a representative of each of the four types of transporter was expressed in *Escherichia coli.* Expression of these genes in carbonic anhydrase-deficient *E. coli* EDCM636 enabled it to grow under low DIC conditions, consistent with DIC transport by these proteins. The results from this study expand the range of DIC transporters within the SbtA and SulP transporter families, verify DIC uptake by transporters encoded by Tcr_0853 and Tcr_0854 and their homologs, and introduce DIC as a potential substrate for transporters from the Chr family.

IMPORTANCE Autotrophic organisms take up and fix DIC, introducing carbon into the biological component of the global carbon cycle. The mechanisms for DIC uptake and

fixation by autotrophic *Bacteria* and *Archaea* are likely to be diverse, but have only been well-characterized among "*Cyanobacteria".* Based on genome sequences, members of *Hydrogenovibrio, Thiomicrospira* and *Thiomicrorhabdus* have a variety of mechanisms for DIC uptake and fixation. We verified that most of these organisms are capable of growing under low DIC conditions, when they upregulate carboxysome loci and transporter genes collocated with these loci on their chromosomes. When these genes, which fall into four evolutionarily independent families of transporters, are expressed in *E. coli,* DIC transport is detected. This expansion in known DIC transporters across four families, from organisms from a variety of environments, provides insight into the ecophysiology of autotrophs, as well as a toolkit for engineering microorganisms for carbon-neutral biochemistries of industrial importance.

KEYWORDS CO_2 concentrating mechanism, chemolithoautoroph, autotroph, carbon fixation

(INTRODUCTION)

Autotrophic members of domains *Bacteria* and *Archaea* are responsible for introducing carbon into the biological portion of the global carbon cycle in virtually any habitat with sufficient light or chemical energy to power the process of carbon fixation. They use CO_2 from the air, or dissolved inorganic carbon (DIC = CO_2 + HCO_3^- + CO_3^{2-}), if aquatic, as their carbon source, and have a variety of mechanisms to compensate for variability in the availabilities of these compounds.

CO_2-concentrating mechanisms (CCMs) are one such mechanism, and have been particularly well-studied among members of the phylum "*Cyanobacteria".* In these organisms, active transport of HCO_3^- elevates its intracellular concentration. Cytoplasmic HCO_3^- enters carboxysomes. These protein-bound microcompartments contain the enzymes carbonic anhydrase (EC 4.2.1.1), and ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO, EC 4.1.1.39). These enzymes act together to dehydrate some of the HCO_3^- to form CO_2, and use the CO_2 to carboxylate ribulose bisphosphate, leading to the formation of 3-phosphoglycerate for biosynthesis (1-3). The HCO_3^- transporters characterized in members of phylum "*Cyanobacteria"* fall into three evolutionarily independent lineages: BCT1, an ABC transporter (4); BicA, a member of the SulP family of transporters (5); and SbtA (6). Loss of cytoplasmic DIC is minimized by conversion of cytoplasmic CO_2 to HCO_3^- via membrane-associated carbonic anhydrases, which couple CO_2 hydration to redox reactions (7, 8). This arsenal of DIC transporters and traps is distributed among members of "*Cyanobacteria"* based on their habitats; those inhabiting freshwater and sediments, in which DIC concentration and composition can vary most greatly (e.g., due to pH differences) tend

to carry a variety of these complexes, while those inhabiting the open ocean, where DIC concentrations and pH values are subject to much less variation, tend to carry a more limited subset (9). For organisms carrying a variety of transporters and traps, these complexes are differentially regulated in a manner consistent with differences in their parameters (e.g., higher-affinity transporters BCT1 and SbtA are favored when HCO_3^- concentrations are particularly low (1)).

CCMs are not well-characterized among autotrophs from the many other phyla of *Bacteria* and *Archaea* with autotrophic members. Carboxysomes are present in many autotrophic members of *Alpha-, Beta-,* and *Gammaproteobacteria,* and their structure and function have been well-characterized for *Halothiobacillus neapolitanus* Parker X^T from the *Chromatiales* of the *Gammaproteobacteria* (10-12). DIC uptake has only been studied in detail for *Hydrogenovibrio crunogenus*, a sulfur-oxidizing chemolithoautotroph from the *Thiotrichales* of the *Gammaproteobacteria,* isolated from deep-sea hydrothermal vents (13). This organism generates elevated intracellular DIC concentrations in an energy-dependent manner (14, 15), and has carboxysomes (16), which likely facilitates its ability to grow rapidly under low-DIC conditions (14). Random and site-directed mutagenesis of gene loci Tcr_0853 and Tcr_0854*,* which are located downstream of the carboxysome locus in this organism, result in a high-CO_2 requiring phenotype, and loss of an ability to generate high intracellular concentrations of DIC, suggesting that these genes encode a two-component DIC transporter (17). Homologs of these genes are common among autotrophic *Bacteria,* and one of them (Tcr_0854*)* is from a PFAM without prior biochemical characterization (PFAM10070; (17).

Several members of *Thiomicrospira, Thiomicrorhabdus,* and *Hydrogenovibrio,* organisms taxonomically affiliated with *H. crunogenus* (18), have had their genomes sequenced. Taxa were selected for sequencing to represent both the taxonomic breadth of these genera, as well as the range of habitats from which these organisms have been isolated, including shallow and deep-sea hydrothermal vents, coastal sediments, and soda and salt lakes (19). Despite the rather narrow taxonomic range of the organisms sequenced, a surprising diversity in mechanisms for DIC uptake and fixation was suggested from the genome data. Genome sequences of some members of *Thiomicrorhabdus* lack carboxysome loci altogether, suggesting the absence of a CCM. For members of *Thiomicrospira,* genes encoding carboxysome components are present, but those encoding carboxysomal carbonic anhydrase are lacking. Instead, they each carry a gene in its place without apparent homologs beyond this genus (19), raising the possibility that these genes might encode a novel form of carboxysomal carbonic anhydrase. In all cases, when present, carboxysome loci are followed by genes encoding transporters from four evolutionarily distinct families. Carboxysome locus-associated genes encoding potential transporters include homologs to those encoding the potential DIC transporter in *H. crunogenus* (Tcr_0853*,* Tcr_0854), and members of the SulP and SbtA families distantly related to those known to transport HCO_3^- in members of "*Cyanobacteria".* Also included are members of the Chr family, which is widely distributed among prokaryotes. The two biochemically characterized members of this family confer resistance to chromate by extruding this anion, perhaps functioning as a chromate/sulfate antiporter (20).

The unexpected diversity in mechanisms for DIC uptake and fixation suggested by genome data from members of *Thiomicrospira, Thiomicrorhabdus,* and *Hydrogenovibrio*, was verified here. Carboxysome presence or absence was confirmed via transmission electron microscopy. To determine whether the genes encoding potential DIC transporters might facilitate growth under low-DIC conditions, their transcription patterns were monitored, and representative members of all four potential DIC transporter families were heterologously expressed in *E. coli* to verify an ability to transport DIC.

RESULTS

Genome context of carboxysome loci, and phylogenetic analysis of genes encoding potential DIC transporters. Carboxysome loci are present in the genomes of most of the organisms studied here (available at Integrated Microbial Genomes and Microbiomes https://img.jgi.doe.gov/). The genome sequences of *Tmr. arctica* and *Thiomicrorhabdus* sp. Milos-T2 lack carboxysome loci (19); either these loci are absent, or they are present in a portion of the genome which has yet to be sequenced. Genomes from all four sequenced members of *Thiomicrorhabdus* were scrutinized for evidence of genome rearrangement in the region associated with the carboxysome locus. For *Tmr. frisia* KP2 and *Tmr. chilensis*, genome synteny was conserved upstream and downstream of the carboxysome locus (Fig. 1). These conserved regions were also present in *Tmr. arctica* and *Thiomicrorhabdus* sp. Milos-T2, but without the intervening carboxysome locus. These data are consistent with carboxysome locus loss in these two taxa.

For the other organisms, phylogenetic analyses were conducted on genes encoding potential DIC transporters from the regions immediately downstream from the carboxysome loci, along with homologs to these genes present elsewhere in these genomes and others (Fig. 2; Fig. 3; Figs. S1-S5 depict sequence logos derived from alignments of these genes). For all four types of potential transporters, genes from the organisms studied here fell into multiple, distinct and distant clades: 2 clades for homologs to Tcr_0853 and 0854, 3 clades for Chr, 2 clades for SbtA, and 4 clades for SulP. Homologs of SulP, SbtA, and the two-component transporter from *H. crunogenus* are often collocated with carboxylases and other enzymes that consume DIC (Fig. 2; Fig. 3) which suggests a role in DIC uptake for them.

Growth under low-DIC conditions, and DIC concentrations *in situ*. All taxa tested here grew under high-DIC conditions; all but *Thiomicrorhabdus* sp. Milos-T2 were capable of growth under low-DIC conditions (2 mM DIC under ~400 ppm CO_2 ambient headspace; Fig. 4). *Tmr. arctica* lacks genes encoding carboxysomes; therefore, growth under low-DIC conditions was unexpected, motivating further consideration of its habitat. CO_2 concentrations estimated for the marine Arctic sediments from which *Tmr. arctica* was isolated are lower than those at the hydrothermal vent habitat from which *H. crunogenus* was isolated (Table 1).

Carboxysome presence and differential expression. When cells were incubated under low-DIC conditions, carboxysomes were apparent in transmission electron micrographs of all taxa whose genomes encode these microcompartments, and were absent in *Tmr. arctica* and *Thiomicrorhabdus* sp. Milos-T2 (Fig. 5).

Transcripts from genes present in carboxysome loci were more abundant when cells were cultivated under low-DIC conditions (Table 2), which is consistent with the role of carboxysomes in CCMs to facilitate growth under low-DIC conditions (21). Carboxysome-associated genes did not have as large a change in transcript abundance in *Tms. pelophila,* and transcripts from the gene N746DRAFT_0321 (*hyp(csoS2),* Table 2), were undetectable under low-DIC conditions.

Response of transporter transcript abundances to DIC concentration. Transcript levels from many genes encoding potential DIC transporters were significantly different when cells were grown under low-DIC versus high-DIC conditions (Table 2; two-tailed *t*-test, $\alpha < 0.05$, $-\Delta\Delta C_t = 0$). Many of these genes were upregulated under low-DIC conditions, with $-\Delta\Delta C_t > 1$, indicating that transcript levels were at least doubled under low-DIC conditions (Table 2; one-tailed *t*-test, $\alpha < 0.05$).

Under low-DIC conditions, all genes assayed here encoding members of two distinct clusters within the SbtA family, and two clusters of homologs of the Tcr_0853/0854-encoded transporter from *H. crunogenus* (17), were upregulated, whether adjacent to carboxysome loci or not (Table 2). When genes encoding members of the Chr and SulP families were adjacent to carboxysome loci, they were also upregulated under low-DIC conditons (Table 2). When located elsewhere on the chromosome, some members of the SulP family were upregulated, but members of Chr not associated with carboxysomal loci were not.

DIC uptake activity of heterologously expressed transporters. Genes encoding members of all four families of potential DIC transporters were selected for

heterologous expression based on collocation with the carboxysome locus and upregulation under low-DIC conditions (Table 2): the two-component transporter from *H. crunogenus* XCL-2 (Tcr_0853*,* Tcr_0854), members of Chr and SulP transporter families from *H. thermophilus* JR2, and a member of the SbtA transporter family from *Tmr. frisia* KP2. Mass spectrometric analysis of proteins from membranes from *E. coli* constructs expressing these transporters verified their expression (Table 3). Signal intensity was always low for the protein product of Tcr_0853, which may reflect that this protein is likely to be particularly hydrophobic (11 predicted membrane-spanning alpha helices), and therefore more difficult to solubilize, digest, elute from the C_{18} column used to resolve the peptides, and ionize for mass spectrometry.

E. coli expressing these genes were able to generate elevated intracellular DIC concentrations (Fig. 6; Fig. S6). Intracellular DIC was similar to extracellular when putative DIC transporters were oriented in reverse relative to the T7 promotor driving their expression. When genes were correctly oriented relative to the T7 promoter, intracellular DIC concentrations were higher than when in reverse orientation (Fig. 6; for *sbtA*, *chr*, *sulP*: two-tailed *t*-test, F versus R orientations, $\alpha < 0.05$; for 8534 (F), 8534(R), 853(F), 854(F): ANOVA, with post-hoc multiple comparisons via Scheffe, Bonferroni, and Tukey tests; $\alpha < 0.05$ for 8534 (F)). Intracellular DIC was particularly high for cells expressing SbtA. The presence of both Tcr_0853 and Tcr_0854 were necessary for the accumulation of intracellular DIC, suggesting that the gene products of both are required for DIC uptake, and that they may form a two-subunit transporter. The presence of *chr* in the forward orientation relative to the promotor did result in

223 elevated DIC concentrations compared to when it was in the reverse orientation, but
224 intracellular concentrations were lower than for the other transporters (Fig. 6, Fig. S6).

225 Expression of all four types of transporter genes in *E. coli* EDCM636 enabled this
226 carbonic anhydrase-deficient strain (22) to grow under an atmosphere of ~400 ppm CO_2
227 (Fig. 7). Cells expressing both Tcr_0853 and Tcr_0854 grew more rapidly than those
228 expressing either of these genes individually, in reverse orientation, or in the absence of
229 IPTG (Fig. 7A). Cells expressing *chr* only grew when this gene was in the forward
230 orientation relative to the T7 promotor. This growth was preceded by a very long lag
231 period, perhaps due to the very low levels of DIC transport measured in cells expressing
232 this gene (Fig. 6). In replicate experiments, it was not clear whether the presence of
233 IPTG stimulated growth (Fig. 7B, 7C). Cells expressing *sbt* and *sulP* grew most rapidly
234 when in the forward orientation relative to the T7 promotor and when IPTG was added
235 to the growth medium (Fig. 7D, E). Growth in the absence of IPTG may have been due
236 to background expression of T7 RNA polymerase (23).

237

238 **DISCUSSION**

239 Strategies for coping with growth under low-DIC conditions are quite diverse
240 among members of *Thiomicrospira, Thiomicrorhabdus,* and *Hydrogenovibrio.* While
241 most strains queried here appear to have CCMs, two do not. For those taxa that do
242 have CCMs, there are variations in carboxysome loci and transporter genes associated
243 with these loci that suggests a surprising heterogeneity among the CCMs of these
244 organisms.

Tmr. arctica and *Thiomicrorhabdus* sp. Milos-T2 do not appear to have CCMs, based on an absence of carboxysome loci and homologs to most of the transporter genes associated with these loci. The inability of *Thiomicrorhabdus* sp. Milos-T2 to grow under low-DIC conditions is consistent with these genome traits. In contrast, growth by *Tmr. arctica* under low-DIC conditions was surprising. This organism may have a novel mechanism for growing under low-DIC conditions. Alternatively, its growth may be facilitated by higher CO_2 concentrations that result from the lower temperatures at which this organism grows, since lower temperatures increase the solubility of CO_2, and also increase the pK_a of bicarbonate (24). However, these physical and chemical factors do not appear likely to result in particularly high concentrations of CO_2 in the habitat from which *Tmr. arctica* was isolated (Table 1). Perhaps the lower maximum specific growth rates observed for this psychrophilic organism (25), may render carboxysomes and CCMs unnecessary.

The branching order predicted from supertrees constructed for members of *Thiomicrospira, Thiomicrorhabdus,* and *Hydrogenovibrio* (18, 19); Fig. 1) suggests that the loss of carboxysome loci may have occurred independently in the lineages leading to *Tmr. arctica* and *Thiomicrorhabdus* sp. Milos-T2. CCMs may not have provided a selective advantage for these organisms. Given the size of carboxysomes, as well as their abundance when expressed (21), carboxysome loss would provide an energetic advantage for cells growing in habitats with consistently elevated concentrations of CO_2. *Thiomicrorhabdus* sp. Milos-T2 was cultured from the same hydrothermal vent system as *Hydrogenovibrio* sp. Milos-T1 (26), which has a carboxysome locus. Perhaps they inhabit different niches with different CO_2 abundances in this system.

Carboxysome loci in members of *Thiomicrospira* are unusual in their lack of genes encoding carboxysomal carbonic anhydrase (*csoSCA)*, distinguishing them from those present in members of *Thiomicrorhabdus* and *Hydrogenovibrio,* as well as many other members of the *"Protoebacteria"* (27). Orthologs to genes encoding carboxysomal carbonic anhydrase are entirely absent from their genomes. In *Tms. pelophila,* the locus N746DRAFT_0321 (IMG gene object ID 2568509999), located at the position in the carboxysomal locus usually occupied by *csoSCA,* does not appear to be transcribed under low-DIC conditions (Table 2, *hyp(csoS2*). Based on this transcription pattern, it is unlikely that this locus encodes a protein that could fulfil the role of carboxysomal carbonic anhydrase under low-DIC conditions. It is possible that the carboxysomes in these organisms function without carbonic anhydrase activity.

Heterologously expressed members of all four transporter families associated with carboxysome loci are capable of DIC uptake (Fig. 6, Fig. 7). These measurements verify DIC uptake by the proteins encoded by Tcr_0853 and Tcr_0854. They expand DIC uptake by SbtA-family transporters to the product of the *sbtA* gene from *Tmr. frisia* KP2, whose predicted amino acid sequence is only 28 to 29% identical to biochemically characterized SbtA-family bicarbonate transporters from "*Cyanobacteria".* Likewise, it also broadens the known distribution of DIC transporters within the SulP-family transporters; the SulP protein from *Hydrogenovibrio thermophilus* JR2 is only 23% identical to the SulP-family bicarbonate transporter from *E. coli,* and only 19 - 21% identical to SulP-family bicarbonate transporters from "*Cyanobacteria".* It also adds DIC transport as a potential function for members of Chr. Perhaps these transporters have differences in affinities for DIC, transport different forms of DIC (CO_2, HCO_3^-, CO_3^{2-}), or

have different mechanisms for transport (e.g., symport with cations; antiport with anions), which provide advantages for their activities under specific growth conditions.

DIC uptake by some transporters from "*Cyanobacteria"* cannot be successfully assayed by silicone oil centrifugation and complementation of growth of carbonic anhydrase-deficient *E. coli* (28). Perhaps shared membership with *E. coli* in the class *Gammaproteobacteria,* or the stronger T7 promoter used for expression here, were responsible for successful heterologous expression and activity. Successful heterologous expression of the genes studied here bodes well for their potential use in constructing organisms capable of synthesizing industrially relevant precursor compounds from CO_2 and HCO_3^-.

Multiple clades of each transporter family are present among the organisms studied here. Among these organisms, homologs to Tcr_0853 and 0854, as well as SbtA, fall into two clades each. In both cases, most of the genes fall together into a single well-supported clade. Genes outside of this clade fall among those present in distantly related members of *"Proteobacteria",* and organisms carrying these genes also carry a copy falling within the clade. These 'extra copies' could have been relatively recently acquired via horizontal gene transfer. Representatives from both clades from both transporter families all have elevated transcript levels when cells are grown under low-DIC conditions. Furthermore, genes encoding these two types of transporters are usually present adjacent to genes encoding CO_2-metabolizing enzymes (Fig. 2; Fig. 3); this colocation, as well as upregulation under low CO_2 conditions suggests that members of these transporter families may predominantly transport DIC.

The only members of the Chr and SulP families to have elevated transcript levels under low-DIC conditions fell within a single clade of each family (Table 2; Fig. 2; Fig. 3). The members of the Chr and SulP families tested here whose transcript levels were not sensitive to DIC concentrations and are not collocated with carboxysome genes may not play a role in CCMs, and may instead be either constitutively expressed DIC transporters, or transport sulfate or other cations. Few members of these transporter families are collocated with genes encoding CO_2-metabolizing enzymes (Fig. 2; Fig. 3), suggesting that roles in DIC uptake may be less widespread in these transporter families. However, it is important to note that a SulP-family transporter has been implicated in DIC uptake in *E. coli* (5, 29).

The SulP-family transporters studied here have domains that distinguish them from other members of this transporter family. Similar to the SulP transporters present in ”*Cyanobacteria",* they lack the β-carbonic anhydrase domain found in some members of this family. When present in other organisms, this carbonic anhydrase domain is located on the cytoplasmic side of the cell membrane. This absence of a carbonic anhydrase domain is consistent with transporting HCO_3^- into the cytoplasm to generate elevated DIC concentrations there; if it were present, a carbonic anhydrase domain would convert the transported HCO_3^- into CO_2 at the cell membrane, where it would diffuse out of the cell (30). Unlike those present in *"Cyanobacteria",* the SulP-family transporters found in the organisms studied here lack the carboxy-terminal STAS domain typically found in SulP proteins. The STAS domain is hypothesized to regulate the activities of these transporters (31, 32). HCO_3^- transporters in *"Cyanobacteria"* are post-translationally regulated, and inactive in the dark (reviewed in (33). Though this

particular mechanism, STAS-domain mediated post-translational regulation, is absent for the SulP proteins studied here, it is still possible that the CCMs of these organisms could be post-translationally regulated, which would provide them with an advantage given the temporal heterogeneity of the habitats from which they were isolated (e.g., hydrothermal vents (34).

The presence of multiple DIC transporter genes in many of the organisms studied here, and their variability in collocation with the carboxysome locus, provide a basis for a model of DIC transporter gene acquisition, loss, and changes in chromosome location. Based on its chromosomal location in members of *Thiomicrospira, Thiomicrorhabdus,* and *Hydrogenovibrio*, it seems likely that an SbtA-family transporter was collocated with the carboxysome locus in the shared ancestor of these three genera (Fig. 8). In this scenario, members of the other three transporter families were encoded elsewhere in the genomes of all three genera, and displaced members of SbtA in many taxa (e.g., orthologs to Tcr_0853 and 0854 (853&4-I); *chr-I;* Fig. 8). Acquisition of a second copy of homologs to Tcr_0853 and 0854 by the lineage leading to *Hydrogenovibrio sp.* Milos-T1 was likely via horizontal gene transfer, given its placement among genes from more distantly related organisms (Fig. 2A). The selective advantage of placing DIC transporter genes adjacent to those encoding carboxysome genes is not apparent, as transporter genes positioned elsewhere on the chromosome are also upregulated under low DIC conditions (e.g., Tcr_0853 and 0854 homologs in *H. thermophilus* JR-2 and MA2-6, *Hydrogenovibrio* sp. Milos-T1, and *Tms. pelophila;* Table 2). Functional characterization of these different types of transporters, as well as a

358 more detailed examination of the growth conditions under which they are upregulated,
359 may clarify the forces driving selective pressure for their positions on the chromosome.

360 **MATERIALS AND METHODS**

361 **Phylogenetic analysis of genes encoding potential DIC transporters.** Transporter
362 genes and their homologs were collected from the Integrated Microbial Genomes and
363 Microbiomes (IMG/M) database (35). Amino acid sequences predicted from the genes
364 were aligned via MUSCLE (36), and sequence logos were generated from the
365 alignments via Weblogo (http://weblogo.berkeley.edu/ (37). Alignments were refined via
366 GBLOCKS with stringent criteria (38). Phylogenetic trees were constructed in PhyML
367 3.0 (39) using Maximum Likelihood (ML) analysis. Smart Model Selection (SMS) in
368 PhyML 3.0 (40) was used to evaluate best-fit models of evolution (853/854: LG +G+I+F,
369 G = 0.888, I= 0.144; Chr: LG +G +F, G = 0.869; SbtA: LG + G + I + F, G = 1.043; I =
370 0.072; SulP: LG +G +F, G = 1.365; LG = Le and Gascuel model; G = gamma
371 distribution parameter; I = proportion of invariant sites; F = amino acid frequencies
372 estimated from the sequences (41)). Results were assessed using 1000 bootstrap
373 replicates, and the consensus tree was visualized using FigTree (Version 1.4.3; (42)).

374 **Growth under low and high-DIC conditions**. Organisms were obtained from
375 the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ):
376 *Hydrogenovibrio crunogenus* XCL-2 DSM 25203, *Hydrogenovibrio thermophilus JR2*
377 DSM 25194, *Hydrogenovibrio thermophilus* MA2-6 DSM 13155, *Hydrogenovibrio*
378 *halophilus* DSM 15072^T, *Hydrogenovibrio kuenenii* DSM 12350*, Hydrogenovibrio*
379 *marinus* DSM 11271^T, *Hydrogenovibrio* sp. Milos-T1 DSM 13190, *Thiomicrorhabdus*

frisia Kp2 DSM 25197, *Thiomicrorhabdus sp.* Milos-T2 DSM 13229, *Thiomicrorhabdus chilensis* DSM 12352, *Thiomicrorhabdus arctica* DSM 13458, *Thiomicrospira pelophila* DSM 1534^{T}.

To test for ability to grow under low-DIC conditions, organisms were cultivated at 20°C in thiosulfate-supplemented artificial seawater (TASW; (14), pH 7.5, 15 µg/L vitamin B-12). For *H. halophilus,* NaCl was raised to 1.5 M (43), and for *Tmr. arctica,* the culture temperature was 10°C. High DIC cultures (50 mM $NaHCO_3$, 5% headspace CO_2) in TASW were used to inoculate paired flasks containing high DIC and low-DIC medium (2 mM $NaHCO_3$, 0.04% ambient air headspace CO_2; one low-DIC and high-DIC culture per strain). Turbidity was monitored at 600 nm.

Estimation of DIC concentrations present in microorganism habitats. To estimate DIC concentrations in Arctic sediments for comparison with those present at hydrothermal vents, DIC speciation was modelled in PHREEQC Interactive 3.3.12 (US Geological Survey, (44)) using the Lawrence Livermore National Laboratory database (llnl.dat, based in part on the EQ3/6 model (45)). Seawater inorganic ion composition was based on that of (46). For the Arctic samples, the initial pH was 8.22 (2.32 mM DIC), and the model was run on the basis of water without air equilibration or a gas-phase present, at 0.01 or 6°C, and 17.7 atm (168 m depth). For vent samples, the initial pH was either 7.20 (2.7 mM DIC) or 5.6 (7.1 mM DIC) and 206.8 atm (2,075 m depth).

Transmission electron microscopy of carboxysomes. Cells were cultivated to verify carboxysome presence in these taxa. Since *Thiomicrorhabdus* sp. Milos-T2 did not grow under low-DIC conditions (see results), a two-stage process was used to

induce carboxysome synthesis in all organisms tested (16). First, cells were cultivated under high-DIC conditions (see above). Cells were harvested from these cultures via centrifugation, and resuspended in low-DIC TASW medium. After incubating in low-DIC medium overnight, cells were centrifuged, preserved with 2.5% glutaraldehyde, and prepared for transmission electron microscopy as in (16).

qRT-PCR assay of transcript abundances from genes encoding carboxysome components and potential DIC transporters. To determine whether carboxysome-associated transporter gene transcripts were more abundant when cells were grown under low-DIC conditions, taxa that were amenable to cultivation in chemostats were grown under two conditions: DIC limitation (low-DIC), and NH_3 limitation (high-DIC; (14). Cells were harvested via centrifugation (10,000 × *g*, 5 min, 4°C), flash-frozen with liquid nitrogen, and stored at -80°C for subsequent RNA extraction using the Ambion RiboPure-Bacteria Kit (17). Primers were designed to target genes encoding carboxysome components (*csoS2; csoS3;* positive control for CCM induction), citrate synthase (calibrator for the $2^{-\Delta\Delta Ct}$ method; (47), and transporters (Table 4), and qRT-PCR assays were implemented in an Applied Biosystems Step One real-time PCR system, using QuantiTect SYBR Green RT-PCR (Qiagen, Inc.) as described in (17).

Heterologous expression of potential DIC transporters. Genes representing all four families of potential DIC transporters were selected from *H. crunogenus* XCL-2 (Tcr_0853*,* Tcr_0854), *H. thermophilus* JR2 (Chr and SulP), and *Tmr. frisia* KP2 (SbtA). Genomic DNA was purified as in (19), and PCR primers were designed to amplify genes and heterologously express them with native amino and carboxy termini (Table

5). High-fidelity Platinum SuperFi DNA polymerase (Invitrogen; Carlsbad, CA) was used as recommended by manufacturer. PCR products were cloned into pET101/D-TOPO vector, and transformed into OneShot TOP10 competent cells (Invitrogen; Carlsbad, CA). Plasmids were purified from transformed cells, and the sequence and orientation of the target genes was verified (Macrogen USA; Rockville, MD). Most constructs were oriented in the forward direction relative to the vector T7 promoter. For constructs with SbtA and both Tcr_0853 and Tcr_0854, some clones had these genes in reverse orientation relative to the T7 promoter. These were used as negative controls. Constructs with *chr* and *sulP* genes in reverse orientation were generated using PCR primers that would orient them as such relative to the promoter (Table 5). Plasmids were transformed into *E. coli* strain Lemo21(DE3) (New England Biolabs, Ipswich, MA), which has been optimized for membrane transporter expression by modulating T7 RNA polymerase activity with a rhamnose-inducible inhibitor for this enzyme (48, 49). Expression, as assayed by DIC uptake (see description of silicone oil centrifugation below), was optimized by growing each strain in the presence of a range of rhamnose concentrations (0 – 2 mM), inducing gene expression with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and harvesting at a range of times (4-24 hours post-induction).

E. coli carrying genes encoding potential DIC transporters were cultivated on a gyrotary shaker (150 rpm, 37°C) in lysogeny broth supplemented with 100 mg/L ampicillin and 30 mg/L chloramphenicol. Since rhamnose addition was not found in pilot experiments to enhance expression of the target genes, it was not added to the growth media. When OD_{600} reached 0.5 – 1.2, IPTG was added (0.4 mM), and cells

were cultivated for another four hours at 30°C. Cells were harvested (10,000 × g 5 min, 4°C). For proteomic analysis, cells were stored at -20°C. For DIC uptake assays, cells were resuspended in fresh medium (OD_{600} ~50 0.4 mM IPTG) and stored on ice until use, within 1 hr of harvest.

Membrane preparations and proteomics. To verify heterologous expression of the potential DIC transporter genes, membranes were prepared from *E. coli* grown as described above. Pellets from 20 ml cultures were thawed and resuspended in 10 ml membrane buffer (50 mM TRIS pH 8, 10 mM EDTA) supplemented with chicken egg lysozyme (0.1 mg/ml), and incubated for 30 min at 20°C. Lysate was sonicated on ice for 15 sec to decrease viscousity, and centrifuged to remove debris and intact cells (6,000 × *g*, 30 min, 4°C). Supernatant was centrifuged to pellet the membranes (75,000 × *g*, 30 min, 4°C). Pellets (membranes) were rinsed twice with membrane buffer.

Membrane pellets were then resuspended in SDS-PAGE sample buffer. Instead of heating to 95°C, which can cause membrane proteins to aggregate (50), samples were incubated at 37°C for 1 hr to facilitate dissolution before subjecting them to SDS-PAGE (51). Coomassie-stained gel fragments were excised from the molecular weight region corresponding to those predicted from the amino acid sequence of the target protein, and processed as described previously (17).

Peptides were separated using a 50cm C_{18} reversed-phase-HPLC column on an Ultimate3000 UHPLC system (Thermo Fisher Scientific) with a 60 minute gradient (4-40% acetonitrile with 0.1% formic acid) and analyzed on a hybrid quadrupole-Orbitrap instrument (Q Exactive Plus, Thermo Fisher Scientific) using data-dependent acquisition, where the top 10 most abundant ions were selected for MS/MS analysis in

471 the linear ion trap. Raw data files were processed in MaxQuant (version 1.6.1.0,
472 www.maxquant.org) and searched against the *H. crunogenus* Uniprot proteome, which
473 had been modified to also include the amino acid sequences predicted from the *H.*
474 *thermophilus* JR2 and *Tmr. frisia* KP2 genes that were cloned, using search parameters
475 and filtering criteria as in (17).

476 **DIC uptake activity of heterologously expressed transporters.** As in (28),
477 two approaches were taken to determine whether the heterologously expressed
478 transporters were capable of DIC uptake: silicone oil centrifugation and
479 complementation of growth of carbonic anhydrase-deficient *E. coli.* Silicone oil
480 centrifugation was used to assay DIC uptake as described (14). 10 µl portions of
481 suspended cells were added to 200 µl lysogeny broth, 50 mM HEPES pH 8, 0.25 mM
482 DI^{14}C (2 mCi/ml NaH^{14}CO$_3$, 15 mCi/mmol; MP Biomedicals, Inc., Irvine, CA). These
483 200 µl suspensions were layered on top of microcentrifuge tubes preloaded with a
484 dense killing solution overlain by silicone oil (14). Timecourses were run to determine
485 how long to incubate the cells before centrifugation into the killing solution to assay
486 DI^{14}C uptake. Based on these pilot experiments, incubations of 90 sec were used. At
487 90 sec, microcentrifuge tubes were centrifuged at maximum speed (14,000 × *g*) for 30
488 sec before processing as described in (14). Cell-free controls were run in parallel with
489 the samples, and ^{14}C counts from these controls (^{14}C accumulation in the killing solution
490 due to e.g., $^{14}CO_2$ diffusion) were subtracted from counts measured when cells were
491 present. Cell volumes (cytoplasm plus periplasm) were determined via silicone oil
492 centrifugation by incubating cells in the presence of tritiated water (3 µci/ml, Amersham
493 Biosciences, Little Chalfont, UK). Cytoplasm volumes were calculated from cell

volumes by assuming they were 92% of cell volume (52). 2-tailed t-tests were used to determine whether DIC concentrations differed in cells expressing genes in forward versus reverse orientation relative to the T7 promoter. For cells expressing Tcr_0853, Tcr_0854, or both, differences in intracellular DIC concentrations were tested for significance with ANOVA, using Scheffe, Bonferroni, and Tukey tests for post-hoc multiple comparisons. Statistical tests were implemented in IBM SPSS Statistics version 24.

E. coli EDCM636 is only capable of growth under high DIC conditions due to disruption of its β-carbonic anhydrase gene with a kanamycin resistance cartridge (22). When genes encoding DIC transporters from "*Cyanobacteria"* are expressed by this strain, it is capable of growing under ambient atmosphere (~400 ppm CO_2; (28). A culture of this strain was obtained from the Coli Genetic Stock Center at Yale University to screen transporters for DIC uptake activity. Unlike (28), in which DIC transporter expression was driven by the *lac* promoter, target gene expression in this study was driven by the T7 promoter. Since the transporter genes carried on pET101/D-TOPO (described above) require T7 RNA polymerase for expression, it was necessary to introduce a derivative of plasmid pAR1219 carrying an IPTG-inducible copy of the gene encoding this enzyme (53) into *E. coli* EDCM636. Since both pET101/D-TOPO and pAR1219 confer resistance to ampicillin, it was necessary to modify pAR1219 beforehand by interrupting its beta lactamase gene with a trimethoprim resistance cartridge using the EZ-Tn5 <DHFR-1> insertion kit (Epicentre). Chemically competent *E. coli* EDCM636 were transformed first with modified pAR1219 (conferring trimethoprim resistance, but not ampicillin resistance). These cells were subsequently

transformed with pET101/D-TOPO plasmids carrying candidate DIC transporters, and screened for an ability to grow under low DIC conditions. Lysogeny broth (100 mg/L ampicillin, 50 mg/L kanamycin, 50 mg/L trimethoprim) was inoculated with each strain, and 5-fold serial dilutions were prepared for each. These overnight cultures were grown overnight under a headspace of 5% CO_2. Since pseudo-revertants capable of growing under low-DIC conditions are frequent in *E. coli* EDCM636 (22, 28), the highest-titer overnight culture was divided into two equal portions. One portion was incubated under 5% CO_2 with the other cultures, while the other was incubated under ambient air overnight. The next morning, the absence of growth for the culture incubated under ambient air was verified. The optical density of the serial dilutions incubated under 5% CO_2 was measured, and cultures with OD_{600} = 0.3 – 0.5 were selected to be used as inocula. These cultures were split into two portions, one was brought to 0.4 mM IPTG, and all were incubated another hour under 5% CO_2. Each was then added 1:100 v/v to three portions of fresh lysogeny broth supplemented with the antibiotics described above, plus 0.4 mM IPTG as appropriate for the experiment. These cultures were incubated (30°C, 150 rpm) under ambient atmosphere and monitored for growth spectrophotometrically at 600 nm.

SUPPLEMENTAL MATERIAL

FIG S1 Sequence logo from the alignment of Tcr_0853 homologs used to construct the phylogenetic tree in Fig. 2A.

FIG S2 Sequence logo from the alignment of Tcr_0854 homologs used to construct the phylogenetic tree in Fig. 2A.

FIG S3 Sequence logo from the alignment of Chr family transporters used to construct the phylogenetic tree in Fig. 2B.

FIG S4 Sequence logo from the alignment of SbtA family transporters used to construct the phylogenetic tree in Fig. 3A.

FIG S5 Sequence logo from the alignment of SulP family transporters used to construct the phylogenetic tree in Fig. 3B.

FIG S6 Intracellular DIC accumulation by *E coli* expressing potential DIC transporter genes A) Tcr_0853, Tcr_0854, *sulP,* or *chr*, and B) *sbtA* from pilot experiments.

ACKNOWLEDGMENTS

We are grateful for support from the National Science Foundation ((NSF-IOS-1257532 to K.M.S.) and the University of South Florida Pilot Studies Awards program. We are also grateful to Gordon Cannon for discussion and comments on an earlier version of this manuscript, as well as the insightful comments of the anonymous reviewers of this manuscript. Author contribution statement: KMS designed the study; all authors conducted experiments, and contributed to data analysis and writing the manuscript.

REFERENCES

1. **Price GD.** 2011. Inorganic carbon transporters of the cyanobacterial CO_2 concentrating mechanism. Photosynth Res **109:**47-57.
2. **Whitehead L, Long BM, Price GD, Badger MR.** 2014. Comparing the in vivo function of alpha-carboxysomes and beta-carboxysomes in two model cyanobacteria. Plant Physiol **165:**398-411.
3. **Kerfeld CA, Melnicki MR.** 2016. Assembly, function and evolution of cyanobacterial carboxysomes. Curr Opin Plant Biol **31:**66-75.
4. **Omata T, Takahashi Y, Yamaguchi O, Nishimura T.** 2002. Structure, function, and regulation of the cyanobacterial high-affinity bicarbonate transporter, BCT1. Funct Plant Biol **29:**151-159.
5. **Price G, Woodger F, Badger M, Howitt S, Tucker L.** 2004. Identification of a SulP-type bicarbonate transporter in marine cyanobacteria. Proc Natl Acad Sci USA **101:**18228-18233.
6. **Shibata M, Katoh H, Sonoda M, Ohkawa H, Shimoyama M, Fukuzawa H, Kaplan A, Ogawa T.** 2002. Genes essential to sodium-dependent bicarbonate transport in cyanobacteria. J Biol Chem **277:**18658-18664.
7. **Shibata M, Ohkawa H, Kaneko T, Fukuzawa H, tabata S, Kaplan A, Ogawa T.** 2001. Distinct constitutive and low-CO_2-induced CO_2 uptake systems in cyanobacteria: Genes involved and their phylogenetic relationship with homologous genes in other organisms. Proc Natl Acad Sci **98:**11789-11794.
8. **Han X, Sun N, Xu M, Mi H.** 2017. Co-ordination of NDH and Cup proteins in CO_2 uptake in cyanobacterium *Synechocystis* sp. PCC 6803. J Exp Bot **68:**3869-3877.
9. **Badger MR, Price GD, Long BM, Woodger FJ.** 2006. The environmental plasticity and ecological genomics of the cyanobacterial CO_2 concentrating mechanism. J Exp Bot **57:**249-265.
10. **Heinhorst S, Williams E, Cai F, Murin D, Shively J, Cannon G.** 2006. Characterization of the carboxysomal carbonic anhydrase CsoCSA from *Halothiobacillus neapolitanus*. J Bacteriol **188:**8087-8094.
11. **Dou Z, Heinhorst S, Williams E, Murin E, Shively J, Cannon G.** 2008. CO_2 fixation kinetics of *Halothiobacillus neapolitanus* mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO_2. J Biol Chem **283:**10377-10384.
12. **Sutter M, Roberts EW, Gonzalez RC, Bates C, Dawoud S, Landry K, Cannon GC, Heinhorst S, Kerfeld CA.** 2015. Structural characterization of a newly identified component of alpha-carboxysomes: The AAA plus domain protein CsoCbbQ. Sci Rep **5:**16243.
13. **Jannasch H, Wirsen C, Nelson D, Robertson L.** 1985. *Thiomicrospira crunogena* sp. nov., a colorless, sulfur-oxidizing bacterium from a deep-sea hydrothermal vent. Int J Syst Bacteriol **35:**422-424.
14. **Dobrinski KP, Longo DL, Scott KM.** 2005. A hydrothermal vent chemolithoautotroph with a carbon concentrating mechanism. J Bacteriol **187:**5761-5766.
15. **Menning KJ, Menon BB, Fox G, Scott UMLKM.** 2016. Dissolved inorganic carbon uptake in *Thiomicrospira crunogena* XCL-2 is Δp- and ATP-sensitive and enhances RubisCO-mediated carbon fixation. Arch Microbiol **198:**149-159.
16. **Dobrinski KP, Enkemann SA, Yoder SJ, Haller E, Scott KM.** 2012. Transcription response of the sulfur chemolithoautotroph *Thiomicrospira crunogena* to dissolved inorganic carbon limitation. J Bacteriol **194:**2074-2081.
17. **Mangiapia M, MCBL U, Brown T-RW, Chaput D, Haller E, Harmer TL, Hashemy Z, Keeley R, Leonard J, Mancera P, Nicholson D, Stevens S, Wanjugi P, Zabinski T, Pan C, Scott KM.** 2017. Proteomic and mutant analysis of the CO_2 concentrating mechanism of hydrothermal vent chemolithoautotroph *Thiomicrospira crunogena*. J Bacteriol **199:**e00871-00816.

18. **Boden R, Scott KM, Williams J, Russel S, Antonen K, Rae AW, Hutt LP.** 2017. An evaluation of *Thiomicrospira, Hydrogenovibrio* and *Thioalkalimicrobium*: Reclassification of four species of *Thiomicrospira* to each *Thiomicrorhabdus* gen. nov. and *Hydrogenovibrio*, and reclassification of all four species of *Thioalkalimicrobium* to *Thiomicrospira*. Int J Syst Evol Micr **67:**1140-1151.
19. **Scott KM, Williams J, Porter CMB, Russel S, Harmer TL, Paul JH, Antonen KM, Bridges MK, Camper GJ, Campla CK, Casella LG, Chase E, Conrad JW, Cruz MC, Dunlap DS, Duran L, Fahsbender EM, Goldsmith DB, Keeley RF, Kondoff MR, Kussy BI, Lane MK, Lawler S, Leigh BA, Lewis C, Lostal LM, Marking D, Mancera PA, McClenthan EC, McIntyre EA, Mine JA, Modi S, Moore BD, Morgan WA, Nelson KM, Nguyen KN, Ogburn N, Parrino DG, Pedapudi AD, Pelham RP, Preece AM, Rampersad EA, Richardson JC, Rodgers CM, Schaffer BL, Sheridan NE, Solone MR, Staley ZR, Tabuchi M, Waide RJ, Wanjugi PW, Young S, Clum A, Daum C, Huntemann M, Ivanova N, Kyrpides N, Mikhailova N, Palaniappan K, Pillay M, Reddy TBK, Shapiro N, Stamatis D, Varghese N, Woyke T, Boden R, Freyermuth SK, Kerfeld CA.** 2018. Genomes of ubiquitous marine and hypersaline *Hydrogenovibrio, Thiomicrorhabdus*, and *Thiomicrospira* spp. encode a diversity of mechanisms to sustain chemolithoautotrophy in heterogeneous environments. Environ Microbiol **20:**2686 – 2708
20. **Nies DH, Koch S, Wachi S, Peitzsch N, Saier MH, Jr.** 1998. CHR, a novel family of prokaryotic proton motive force-driven transporters probably containing chromate/sulfate antiporters. J Bacteriol **180:**5799-5802.
21. **Kerfeld CA, Heinhorst S, Cannon GC.** 2010. Bacterial microcompartments. Ann Rev Microbiol **64:**391-408.
22. **Merlin C, Masters M, McAteer S, Coulson A.** 2003. Why Is Carbonic Anhydrase Essential to *Escherichia coli*? J Bacteriol **185:**6415-6424.
23. **Rosano GL, Ceccarelli EA.** 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. Front Microbiol **5:**172.
24. **Zeebe RE, Wolf-Gladrow D.** 2003. CO_2 in seawater: Equilibrium, kinetics, isotopes. Elsevier, New York.
25. **Knittel K, Kuever J, Meyerdierks A, Meinke R, Amann R, Brinkhoff T.** 2005. *Thiomicrospira arctica* sp. nov. and *Thiomicrospira psychrophila* sp. nov., psychrophilic, obligately chemolithoautotrophic, sulfur-oxidizing bacteria isolated from marine Arctic sediments. Int J Syst Evol Microbiol **55:**781-786.
26. **Brinkhoff T, Sievert SM, Kuever J, Muyzer G.** 1999. Distribution and diversity of sulfur-oxidizing *Thiomicrospira* spp. at a shallow-water hydrothermal vent in the Aegean Sea (Milos, Greece). Appl Environ Microb **65:**3843-3849.
27. **Axen SD, Erbilgin O, Kerfeld CA.** 2014. A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. PLoS Comput Biol **10:** e1003898.
28. **Du J, Förster B, Rourke L, Howitt SM, Price GD.** 2014. Characterisation of cyanobacterial bicarbonate transporters in *E. coli* shows that SbtA homologs are functional in this heterologous expression system. PLoS ONE **9:**e115905.
29. **Babu M, Greenblatt JF, Emili A, Strynadka NCJ, Reithmeier RAF, Moraes TF.** 2010. Structure of a SLC26 anion transporter STAS domain in complex with acyl carrier protein: Implications for *E. coli* YchM in fatty acid metabolism. Structure **18:**1450-1462.
30. **Price GD, Howitt SM.** 2011. The cyanobacterial bicarbonate transporter BicA: its physiological role and the implications of structural similarities with human SLC26 transporters. Biochemistry and cell biology = Biochimie et biologie cellulaire **89:**178-188.
31. **Aravind L, Koonin EV.** 2000. The STAS domain - a link between anion transporters and antisigma-factor antagonists. Current biology : CB **10:**R53-55.

32. **Shelden MC, Howitt SM, Price GD.** 2010. Membrane topology of the cyanobacterial bicarbonate transporter, BicA, a member of the SulP (SLC26A) family. Mol Membr Biol **27:**12-22.
33. **Price GD, Badger MR, Woodger FJ, Long BM.** 2009. Advances in understanding the cyanobacterial CO_2-concentrating-mechanism (CCM): functional components,Ci transporters, diversity, genetic regulation and prospects for engineering into plants. J Exp Bot **59:**1441-1461.
34. **Le Bris N, Govenar B, Le Gall C, Fisher CR.** 2006. Variability of physico-chemical conditions in 9 degrees 50 ' N EPR diffuse flow vent habitats. Mar. Chem. **98:**167-182.
35. **Markowitz VM, Chen IA, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J, Woyke T, Huntemann M, Anderson I, Billis K, Varghese N, Mavromatis K, Pati A, Ivanova NN, Kyrpides NC.** 2014. IMG 4 version of the integrated microbial genomes comparative analysis system. Nucl Acids Res **42:**D560-D567.
36. **Edgar RC.** 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucl Acids Res **32:**1792-1797.
37. **Crooks GE, Hon G, Chandonia JM, Brenner SE.** 2004. WebLogo: A sequence logo generator. Genome Res **14:**1188-1190.
38. **Talavera G, Castresana J.** 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic biology **56:**564-577.
39. **Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O.** 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology **59:**307-321.
40. **Lefort V, Longueville JE, Gascuel O.** 2017. SMS: Smart model selection in PhyML. Mol Bio Evol **34:**2422-2424.
41. **Le SQ, Gascuel O.** 2008. An improved general amino acid replacement matrix. Mol Biol Evol **25:**1307-1320.
42. **Rambaut A.** 2016. FigTree: Tree Figure Drawing Tool Version 1.4.3, CoInstitute of Evolutionary Biology, University of Edinburgh.
43. **Sorokin DY, Tourova TP, Kolganova TV, Spiridonova EM, Berg IA, Muyzer G.** 2006. *Thiomicrospira halphila* sp. nov., a moderately halophilic, obligately chemolithoautotrophic, sulfur-oxidizing bacterium from hypersaline lakes. Int J Syst Evol Microbiol **56:**2375-2380.
44. **Parkhurst DL, Appelo CAJ.** 1999. User's guide to PHREEQC (Version 2) : A computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations. Report 99-4259.
45. **Wolery T.** 1979. Calculation of chemical equilibrium between aqueous solution and minerals: the EQ3/6 software package. [In FORTRAN extended 4. 6 for CDC6600 and 7600].
46. **Nordstrom DK, Plummer LN, Wigley TML, Wolery TJ, Ball JW, Jenne EA, Bassett RL, Crerar DA, Florence TM, Fritz B, Hoffman M, Holdren GR, Lafon GM, Mattigod SV, McDuff RE, Morel F, Reddy MM, Sposito G, Thrailkill J.** 1979. A comparison of computerized chemical models for equilibrium calculations in aqueous systems, p. 857-892, Chemical Modeling in Aqueous Systems, vol. 93. American Chemical Society.
47. **Livak KJ, Schmittgen TD.** 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. Methods **25:**402-408.
48. **Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, Högbom M, van Wijk KJ, Slotboom DJ, Persson JO, de Gier J-W.** 2008. Tuning *Escherichia coli* for membrane protein overexpression. Proc Nat Acad Sci USA **105:**14371-14376.
49. **Schlegel S, Löfblom J, Lee C, Hjelm A, Klepsch M, Strous M, Drew D, Slotboom DJ, de Gier J-W.** 2012. Optimizing membrane protein overexpression in the *Escherichia coli* strain Lemo21(DE3). J Mol Biol **423:**648-659.

50. **Sagné C, Isambert MF, Henry JP, Gasnier B.** 1996. SDS-resistant aggregation of membrane proteins: application to the purification of the vesicular monoamine transporter. Biochem J **316:**825-831.
51. **Sambrook J, Russell DW.** 2001. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
52. **Oliver DB.** 1996. Periplasm. *In* Neidhardt FC (ed.), *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, vol. 1. ASM Press, Washington, DC.
53. **Davanloo P, Rosenberg AH, Dunn JJ, Studier FW.** 1984. Cloning and expression of the gene for bacteriophage T7 RNA polymerase. Proc Natl Acad Sci U S A **81:**2035-2039.
54. **Cervantes C, Ohtake H, Chu L, Misra TK, Silver S.** 1990. Cloning, nucleotide sequence, and expression of the chromate resistance determinant of Pseudomonas aeruginosa plasmid pUM505. J Bacteriol **172:**287-291.
55. **Zolotarev AS, Unnikrishnan M, Shmukler BE, Clark JS, Vandorpe DH, Grigorieff N, Rubin EJ, Alper SL.** 2008. Increased sulfate uptake by *E. coli* overexpressing the SLC26-related SulP protein Rv1739c from *Mycobacterium tuberculosis*. Comparative biochemistry and physiology. Part A, Molecular & integrative physiology **149:**255-266.
56. **Weissgerber TL, Milic NM, Winham SJ, Garovic VD.** 2015. Beyond bar and line graphs: time for a new data presentation paradigm. PLoS Biol **13:**e1002128.
57. **Goffredi SK, Childress JJ, Desaulniers NT, Lee RW, Lallier FH, Hammond D.** 1997. Inorganic carbon acquisition by the hydrothermal vent tubeworm *Riftia pachyptila* depends upon high external P-CO_2 and upon proton-equivalent ion transport by the worm. J Exp Biol **200:**883-896.
58. **Ahmad A, Barry JP, Nelson DC.** 1999. Phylogenetic affinity of a wide, vacuolate, nitrate-accumulating *Beggiatoa* sp. from Monterey Canyon, California, with *Thioploca* spp. Appl Environ Microbiol **65:**270-277.

FIGURE LEGENDS

FIG 1 Carboxysome-associated locus and genome context among members of genus *Thiomicrorhabdus.* Homologous genes are consistently colored among genomes. Black genes are unique to the genome within the region depicted. For *Tmr. chilensis,* dots indicate a region of the scaffold that has not been sequenced. Locus tags depicted are A379DRAFT_1550 – 1580 (*Tmr. frisia* KP2), BS34DRAFT_2186 – 2175 (*Thiomicrorhabdus* sp. Milos-T2), F612DRAFT_1864 – 1855 (*Tmr. arctica*), and B076DRAFT_0150 – 0174 (*Tmr. chilensis*). The phylogenetic tree on the left is a portion of a larger phylogenetic analysis in (19).

FIG 2 Maximum likelihood analysis of homologs of Tcr_0853 and 0854 (A), and members of the Chr (B) transporter family. Borders highlight members of *Thiomicrospira, Hydrogenovibrio,* and *Thiomicrorhabdus;* 'Q+' and 'Q-' indicate genes whose transcripts were assayed via qRT-PCR and found to be upregulated, or not, respectively, under low-DIC conditions (Table 2). 'E' indicates genes heterologously expressed in *E. coli.* Taxon names are preceded by Integrated Microbial Genomes gene object id numbers or GenBank accession numbers, and are also preceded by 'BC' when the gene products have been characterized biochemically (54). When transporter family genes were collocated with genes encoding enzymes that consume or produce DIC, taxon names are preceded by the following abbreviations: CA – carbonic anhydrase; cbbM – form II RubisCO; CS – carboxysome; FDH – formate dehydrogenase; OAOR – oxoacid: acceptor oxidoreductase. In (A), '3sub' indicates that a gene encoding a potential third subunit is present between genes encoding homologs of Tcr_0853 and 0854. Alignments had 420 (A) and 138 (B) positions. Bootstrap values >65% from 1000 resamplings of the alignment are shown, and the trees are unrooted. The scale bar represents the number of substitutions per site.

FIG 3 Maximum likelihood analysis of homologs of SbtA (A) and SulP (B) transporter families. Borders highlight members of *Thiomicrospira, Hydrogenovibrio,* and *Thiomicrorhabdus;* 'Q+' and 'Q-' indicate genes whose transcripts were assayed via qRT-PCR and found to be upregulated, or not, respectively, under low-DIC conditions (Table 2). 'E' indicates genes heterologously expressed in *E. coli.* Taxon names are

preceded by Integrated Microbial Genomes gene object id numbers or GenBank accession numbers, and are also preceded by 'BC' when the gene products have been characterized biochemically (28, 29, 55). When transporter family genes were collocated with genes encoding enzymes that consume or produce DIC, taxon names are preceded by the following abbreviations: CA – carbonic anhydrase; cbbL – form I RubisCO; cbbM – form II RubisCO; CS – carboxysome; FHL – formate hydrogen lyase; OAOR – oxoacid: acceptor oxidoreductase; PUR – purine biosynthesis; PyrC – pyruvate carboxylase.). Alignments had 148 (A), and 160 (B) positions. Bootstrap values >65% from 1000 resamplings of the alignment are shown, and the trees are unrooted. The scale bar represents the number of substitutions per site.

FIG 4 Growth of *Hydrogenovibrio* (A)*, Thiomicrorhabdus,* and *Thiomicrospira* (B) species under low DIC conditions. Single cultures from each species were cultivated under an ambient headspace with 2 mM DIC. *Tmr. arctica* was cultivated at 10ºC; the rest were grown at 20ºC.

FIG 5 Transmission electron micrographs of cells exposed to low-DIC conditions to induce carboxysome synthesis. Carboxysomes are visible as 0.1 µm electron-dark inclusions; when present in the cells, two are indicated per cell with arrows.

FIG 6 Intracellular DIC accumulation by *E coli* expressing potential DIC transporter genes A) Tcr_0853, Tcr_0854, or *chr*, and B) *sbtA* or *sulP*. (F) or (R) following gene

779 names indicates the orientation of the gene (forward or reverse) with respect to the T7
780 promotor. '8534' is a construct carrying both Tcr_0853 and Tcr_0854, while '0853' and
781 '0854' each carry Tcr_0853 or Tcr_0854 respectively. DIC concentrations were
782 measured 8 times for cells from a single culture of each construct, and the median value
783 for each construct is indicated with a short horizontal bar. The concentration of
784 extracellular DIC was 0.25 mM, and the incubation time was 90 sec. Asterisks indicate
785 constructs in which genes in forward orientation accumulated DIC to a significantly
786 higher concentration than when in reverse orientation (α<0.05). Scatterplots of
787 intracellular DIC pools were generated using the template provided in (56).

788

789 **FIG 7** Growth of carbonic anhydrase-deficient *E. coli* EDCM636 under ~400 ppm CO_2
790 when expressing candidate DIC transporters. Cells were cultivated in the presence of
791 0.4 mM IPTG, unless indicated otherwise ('- IPTG'). Genes encoding potential DIC
792 transporters were oriented in forward (F) or reverse (R) orientation relative to the T7
793 promoter driving expression. A. Cells carrying *Tcr_0853* (853), *Tcr_0854* (854), or both
794 (8534). B and C. Replicate experiments for cells expressing *chr* genes. D and E. Cells
795 expressing *sbt* or *sulP* genes, respectively.

796

797 **FIG 8** Model of DIC transporter gene acquisition, loss, and changes in genome
798 location. The sequence of events with the least number of gene gains, losses, and
799 movements within the chromosome is presented, overlaying a ribosome-protein based
800 supertree (19) of the organisms. Asterisks mark clades with 98 – 100 % bootstrap

support. On the left is a possible ancestral carboxysome locus. On the right are the carboxysome loci in these organisms. Numbers in parentheses indicate the number of scaffolds of the draft genome sequence; (C) indicates that the genome sequence is complete. + *gene* = phylogenetic analysis and gene taxonomic distribution suggest acquisition via horizontal gene transfer; - *gene* = gene is absent from sequenced genome; CS = gene is collocated with carboxysome locus; to CS = gene moved from elsewhere in the chromosome, to collocation with carboxysome locus; CCM = CO_2 concentrating mechanism.

812 **TABLE 1** Dissolved inorganic carbon speciation in *Tmr. arctica* and *H. crunogenus*

813 habitats

	Arctic marine sediments[a] (*Tmr. arctica*)	**Hydrothermal vents[a] (*H. crunogenus*)**
Temperature	0.01 – 6.00 °C	2.00 – 20.00 °C
pH	8.22	7.20 - 5.60
Sulfide[b]	0 – 0.3	0 – 0.3
DIC	2.32	2.7 – 7.1
CO_2	0.017 – 0.015	0.194 – 5.157
HCO_3^-	1.062 – 1.344	1.483 – 1.029
$NaHCO_3$	0.464 – 0.627	0.682 – 0.326
$MgHCO_3^+$	0.167 – 0.206	0.232 – 0.153
$CaHCO_3^+$	0.030 – 0.037	0.041 – 0.028
CO_3^{2-}	0.012 – 0.022	0.002 – 0.000
$NaCO_3^-$	0.007 – 0.011	0.001 – 0.000
$MgCO_3$	0.023 – 0.047	0.004 – 0.001
$CaCO_3$	0.007 – 0.015	0.001 – 0.000

814

815 [a]Temperatures, pH, sulfide, and DIC from Arctic marine sediments and hydrothermal

816 vents are based on those present at the locations from which these organisms were

817 isolated (25, 34, 57, 58).

818 [b]Compounds are presented in mM

819 **TABLE 2** Transcript abundances of genes encoding carboxysome components and potential DIC transporters in
820 members of the genera *Hydrogenovibrio, Thiomicrorhabdus,* and *Thiomicrospira*
821

Taxon	Genes[a]	α ($-\Delta\Delta C_t = 0$)[b]	$-\Delta\Delta C_t \pm$ SD[c]	α ($-\Delta\Delta C_t > 1$)[d]	Fold increase (low DIC/high DIC)
Hydrogenovibrio crunogenus XCL-2	*csoS3*		9.7 ± 1.9[e]		823
	853-I (CS)		8.0 ± 1.8[e]		263
	854-I (CS)		8.4 ± 1.6[e]		340
	chr-I	-[f]	0.4 ± 0.7		1.3
	sulP-II	<0.05	0.8 ± 0.2	N/A[g]	1.8
Hydrogenovibrio thermophilus JR2	*csoS3*	<0.005	11.0 ± 0.2	<0.001	1984
	chr-I (CS)	<0.005	6.8 ± 0.2	<0.001	114
	853-I	<0.005	10.2 ± 0.3	<0.001	1181
	854-I	<0.005	8.7 ± 0.1	<0.001	413
	sulP-I	<0.005	6.9 ± 0.1	<0.001	123
	sulP-II	<0.01	-1.1 ± 0.1	N/A	0.5
Hydrogenovibrio thermophilus MA2-6	*csoS3*	<0.005	12.5 ± 0.4	<0.001	5746
	chr-I (CS)	<0.005	6.2 ± 0.4	<0.005	73
	853-I	<0.005	6.4 ± 0.4	<0.001	85
	854-I	<0.005	8.2 ± 0.4	<0.001	289
	sulP-II	-	-0.1 ± 0.5	N/A	0.9
Hydrogenovibrio halophilus	*csoS3*	<0.005	11.3 ± 0.3	<0.001	2517
	sbtA-I (CS)	<0.005	9.4 ± 0.2	<0.001	680
	chr-I	<0.05	-0.5 ± 0.2	N/A	0.7
	sulP-III	<0.05	-0.8 ± 0.2	N/A	0.6

Hydrogenovibrio marinus	*csoS3*	<0.005	8.4 ± 0.4	<0.001	340
	853-I (CS)	<0.005	5.9 ± 0.4	<0.001	61
	854-I (CS)	<0.01	4.1 ± 0.3	<0.005	18
	chr-II	-	-0.9 ± 0.5	N/A	0.5
Hydrogenovibrio sp. Milos-T1	*csoS3*	<0.005	4.8 ± 0.2	<0.001	28
	853-I (CS)	<0.005	3.5 ± 0.2	<0.001	11
	854-I (CS)	<0.01	5.1 ± 0.4	<0.005	33
	853-II	<0.01	2.7 ± 0.2	<0.005	6
	*hyp(*853*)*	<0.005	4.1 ± 0.2	<0.005	17
	854-II	<0.005	4.1 ± 0.3	<0.005	17
	chr-II	-	-0.3 ± 0.2	N/A	0.8
Thiomicrorhabdus frisia Kp2	*csoS3*	<0.005	12.4 ± 0.3	<0.001	5327
	sbtA-I (CS)	<0.05	9.2 ± 0.6	<0.025	597
	sulP-II	<0.05	1.4 ± 0.4	-	2.7
Thiomicrospira pelophila					
	csoS2	<0.025	1.8 ± 0.3	<0.05	3
	hyp(csoS2)	N/A	ND[h]	N/A	
	sulP-I (CS)	<0.025	1.6 ± 0.4	-	3
	sbtA-I (CS)	<0.01	2.2 ± 0.3	<0.01	5
	853-I	<0.005	6.4 ± 0.4	<0.001	99
	854-I	<0.005	6.2 ± 0.4	<0.001	73
	chr-I	-	-0.2 ± 0.3	N/A	1
	sbtA-II	<0.05	10.4 ± 2.7	<0.025	1391
	sulP-III	-	-0.1 ± 0.5	N/A	0.9

[a]Gene abbreviations: *chr* = chromate ion transporter family; *sulP* = sulfate transporter family; 853, 854 = homologs to Tcr_0853, 0854*; hyp(x)* = hypothetical protein adjacent to gene x; *sbtA* = sodium-dependent bicarbonate transporter

825 family. Roman numerals (I, II, III) are consistent with the clades labeled in Fig. 2 and 3. (CS) indicates that the genes are
826 adjacent to the carboxysome locus. IMG gene object ID numbers for all genes targeted here are listed in Table 4.

827 [b]For all species, citrate synthase was used as the calibrator gene.

828 [c]two-tailed *t*-test, n=3 for all except *Tmr. frisia* Kp2 *sbtA (CS),* for which n=2

829 [d]one-tailed *t*-test, n=3 for all except *Tmr. frisia* Kp2 *sbtA (CS),* for which n=2

830 [e](17)

831 [f] -: $\alpha > 0.05$

832 [g]N/A: Not applicable

833 [h]ND: Not detectable; C_t values similar to cDNA-free controls (>30). Primers successfully amplified this target when gDNA
834 was the template.

835

TABLE 3 Detection of putative DIC transporter proteins when heterologously expressed in *E. coli*

Sample[a]	Strain	Protein	Intensity	Unique Peptides	Sequence Coverage (%)
8534	*Hydrogenovibrio crunogenus* XCL-2	Tcr_0853	1.37E+09	5	11.4
		Tcr_0854	7.12E+11	57	76.7
853	*Hydrogenovibrio crunogenus* XCL-2	Tcr_0853	4.30E+07	2	4.4
854	*Hydrogenovibrio crunogenus* XCL-2	Tcr_0854	2.69E+10	35	60.9
Chr	*Hydrogenovibrio thermophilus* JR2	Chr	2.56E+09	5	16.9
SbtA	*Thiomicrorhabdus frisia* Kp2	SbtA	5.49E+09	4	7.5
SulP	*Hydrogenovibrio thermophilus* JR2	SulP	6.21E+08	2	5.4

[a]Samples consisted of membranes prepared from *E. coli* cells expressing potential DIC transporters. Sample 8534 is from *E. coli* expressing both Tcr_0853 and Tcr_0854 (8534), while samples 853 and 854 were from *E. coli* expressing either Tcr_0853 or Tcr_0854.

852 **TABLE 4** Primers used for qRT-PCR

Taxon	IMG gene object ID[a]	Predicted gene product[b]	Forward primer	Reverse primer
H. crunogenus XCL-2	637785059	Citrate synthase	CTTTGATGCGGGCTTGTTTAC	CCCCTGTGTAGATTTGAGTCG
	637785561	CsoSCA	CTCCGCTTACCTTATGCCTTAG	AGTAACGTGTTGGTTCATCCG
	637786436	Chr-I	GGTTTCGGCCTGGACTATTT	GCGCTTCATCAAACCAAGAC
	637786269	SulP-II	CGGATTGATTACCGCCATCT	TGCCATGCTCCATCACTAAA
H. thermophilus. JR2	2507072380	Citrate synthase	CGAATCCGTGCTCGGTTATT	GAACCGATTTCATCCAGCATTTC
	2507073759	CsoSCA	GCGTTCCAGGCTCTAAAGATAG	GGATGCCGACAATTCCTGATA
	2507073746	Chr-I (CS)	GCTGGAGCTTGATCGTGTTA	CCATCTCCGATCGACCAAATAC
	2507074342	853-I	CCTGTTTATGGCCGGTTACA	GTCACCCATTCGTCCAGATAAA
	2507074343	854-I	GCTTCGCCTCAGTGTCTATATC	GAGTCCCAACGGAAACAGAA
	2507074344	SulP-I	TGTGTGGCTGTGGCTTTAT	TTGGAGTTACAGGGTCGTTTC
	2507073582	SulP-II	ACACCTTGTCGGGCATTAC	GAGGTGATGAAACCGACGATAA
H. thermophilus MA2-6	2572250326	Citrate synthase	CGAATCCGTGCTCGGTTATT	GAACCGATTTCATCCAGCATTTC
	2572249265	CsoSCA	GAAATCGGAAGACAGGACAGAG	CCATTTCATAACGACGCAACAA
	2572249252	Chr-I (CS)	TGGCGCTGAATCTGGTATTG	CGCCACCCAACTCCAATAAA
	2572249866	853-I	CCTACATGGCCGGGAATAAG	ATCCAAGCGGTCATCATCAG
	2572249867	854-I	ATGTGCGCTCGGAAATCA	GGGCGGTATTCTATCGGTAATC
	2572249063	SulP-II	ACACCTTGTCGGGCATAAC	GACGATAATCGCGGCATACA
H. halophilus	2518266203	Citrate synthase	CCAGACGGGTCAAATACAATCT	ATGTATTCGGTGCTGGGATAAA
	2518265324	CsoSCA	AGGGTCTGTACCCGGATATT	CGTGTCCAGAAACCCGTAAT
	2518265315	SbtA-I (CS)	ATTGGCCACGTCGGATTT	GAATCGCAACAGCGCATAAC
	2518266741	Chr-I	CTGGCTGGGACAAACCTATT	CCCAGAGCCGAACTCATTT
	2518265727	SulP-III	TGGCCGGGTATCTGAATTTG	CGGTAGCTGAGCCATGAATATC
H. marinus	2574157295	Citrate synthase	AGAAGAGTTGGGAGCGTTTG	GGATGTGTTGTTGACGGTAGAT
	2574157483	CsoSCA	CGTTATCAGGAGTTGTCGGTATC	CTGCCAGGTTGGGTAGTTT

	2574157469	853-I (CS)	GCGGCACTGCTATTTGTTTAC	ACTGATCCACAGGTCTCCTATC
	2574157468	854-I (CS)	TTATGACTGGCAGCAGGATAAG	CGACGCGTAGTACTGAAGATTG
	2574158185	Chr-II	TGCTGCTGGTCTGCTATTT	CGGGCTTAATGCCGTAGAA
Hydrogenovibrio sp. Milos-T1	2579718736	Citrate synthase	GACTGGTGAAGAGCCAGATAAG	CCCAGTCACCATAGTGGTAAAG
	2579719002	CsoSCA	GATGTCAGCGAGAGTGTTAGAG	TCTGCTTTCGAGAAGTGGTAAA
	2579719013	853-I (CS)	GGGATTTGTGGATGGCATTTC	ACACTCTTCTGGTCTTCATCATC
	2579719014	854-I (CS)	GGGATGTATAGCGAGTGGTTAG	GGTATCTGGCAAGCTGAGAA
	2579719194	853-II	GTTTGGCGAGCAGCTTTATG	CCCGTAGCCAGCCAATAATAA
	2579719193	Hyp(853)	TCCTAACTGGGTGATGATTTGG	CATGCAGACGTCGCAATAAAG
	2579719192	854-II	ACTACCAACCCAAGCCTAAAG	GTAGTCTCCCATGCCTTCTAAAT
	2579720167	Chr-II	CATCAGGAGCTGGTGGATAAG	ATATAGGTGGCGAGCTGTTG
Tmr. frisia Kp2	2517375157	Citrate synthase	ACCCTTGTTCGGTTATCTCTTC	CAGGCGAACCAATCTCATCT
	2517375722	CsoSCA	GCGTACGTAACTGGGTCTTTAT	TTGGTGGTGTGGATCTGATTT
	2517375731	SbtA-I (CS)	GCACACGAAAGCTACCCTATTA	CTGAGAACCATCACCTGAAGTC
	2517376429	SulP-II	ATGACACCCTATCAGGCATTAC	CAGGACGACCACCAAATACA
Tms. pelophila	2568511528	Citrate synthase	GATCCAATTGAACCGCGTAAAG	CATCAAGTAAACGTGCCCAATC
	2568509998	CsoS2	GCTAATGCTTACTCTGCACCTA	CGGCTCCATCTCCTGTTATTC
	2568509999	Hyp(csoS2)	TAGGTTCGCCAAGCAGTAAG	GCATCAGTCAAGGCATACAAAC
	2568510006	SulP-I (CS)	GAAGCGCCTAAACAAGACAAAG	TAGCCATTGCGGCTGTAATAA
	2568510007	SbtA-I (CS)	GCGGCTAGTGCATCCTATATT	ACGGAAAGGTAACTCCAAGTG
	2568511025	853-I	CGGGTAATGTTGAGGAAGAGAA	AACCCACGCCAGCATAAA
	2568511024	854-I	GGAGTCCAACTTAGCCGATTAC	AGCCCGTCTTGCCAATTTA
	2568510063	Chr-I	CGCCGCCTTAATAAATCCAATC	AACAGGTTCAGACCCAGTAATC
	2568511393	SbtA-II	GCGGCTACCTATGGTTCAATTA	ACCATTGCCACCGTCATATAG
	2568511001	SulP-III	GTCAGGGCGTGGCAAATA	CCTGCCGTAAAGGTGGATAAA

[a]Gene object identification numbers from the Integrated Microbial Genomes system (https://img.jgi.doe.gov/cgi-bin/m/main.cgi)

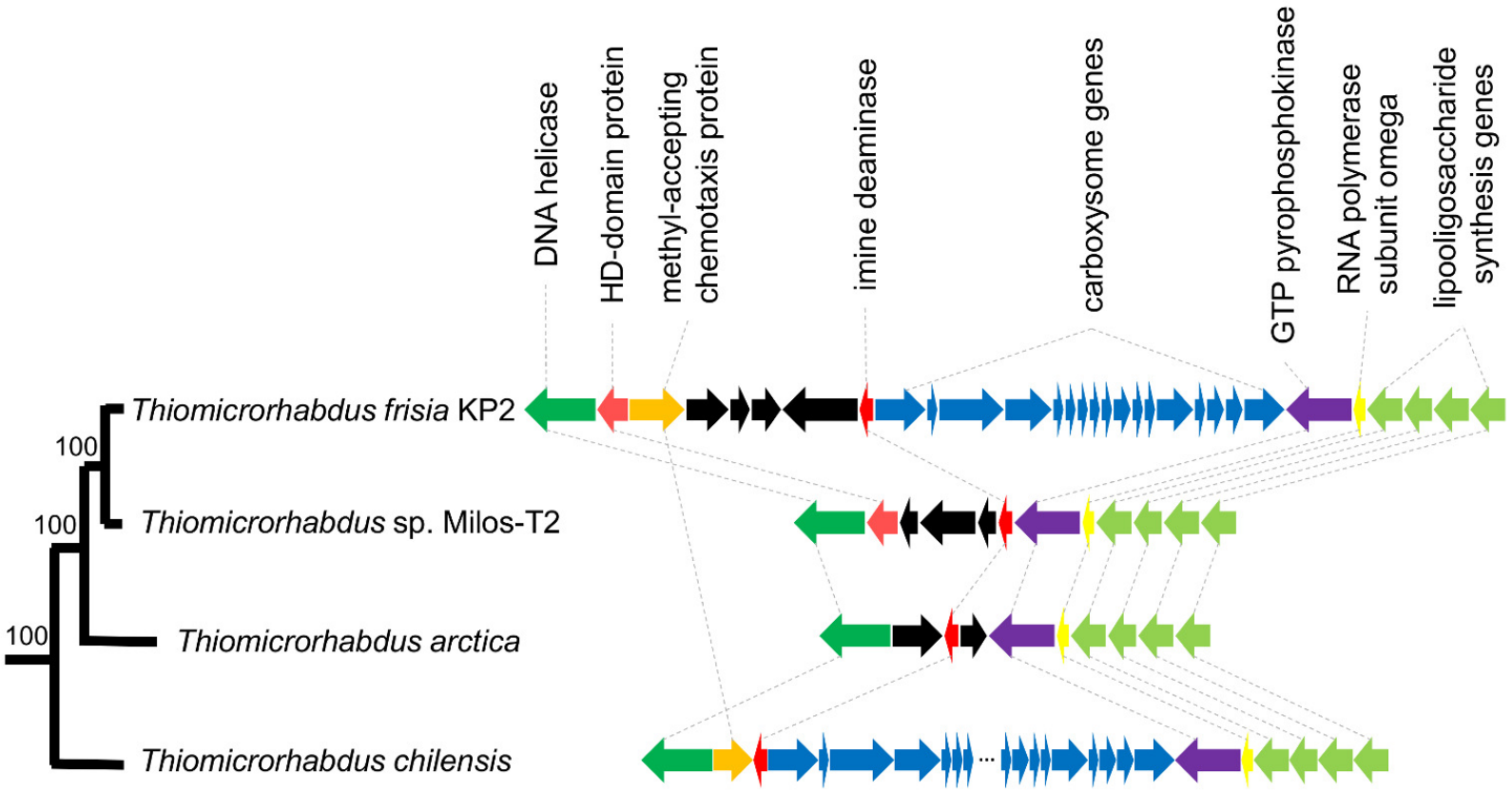
[b]Gene product abbreviations: CsoSCA = carboxysomal carbonic anhydrase; Chr = chromate ion transporter family; SulP = sulfate permease family; 853, 854 = homologs to proteins encoded by *Tcr_0853, 0854;* Hyp(x) = hypothetical protein adjacent to gene x; SbtA = sodium-dependent bicarbonate transporter family. Roman numerals after predicted gene product names indicate the clade in which the gene is found (Fig. 2, Fig. 3). Gene product names followed by (CS) indicate that these genes are adjacent to those encoding the components of carboxysomes.

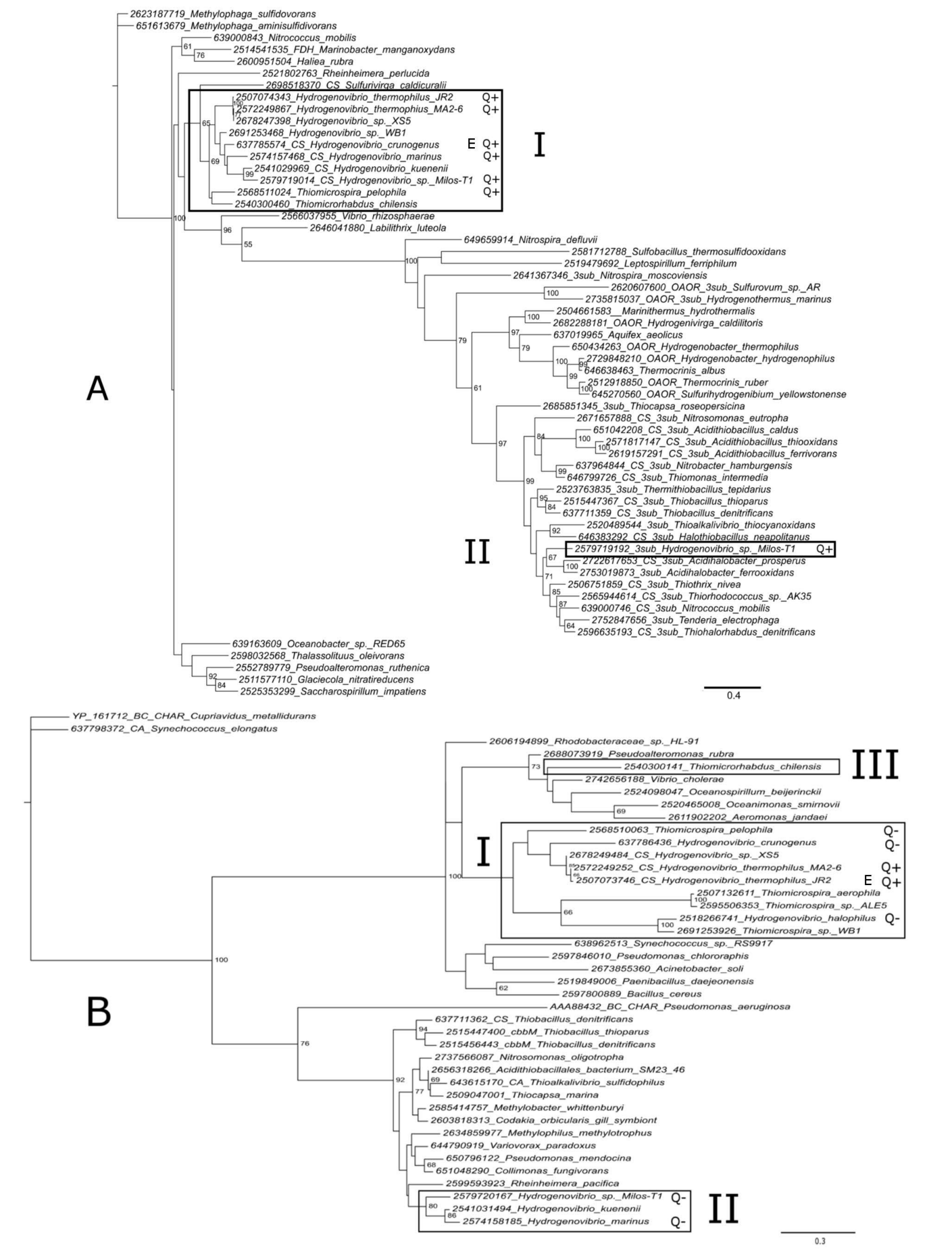
TABLE 5 Primers used for heterologous expression of potential DIC transporters

Taxon	IMG gene object ID[a]	Predicted gene product[b]	Forward primer	Reverse primer
H. crunogenus XCL-2	637785573	853 (CS)	CACCATGAATATGCAATGGGTAGG	TCATTGAAATAACTCCTCTTTAGGAACTT
	637785574	854 (CS)	CACCATGATGTTGCACAACGC	TCAGGCAGATTCCAACCACT
		853 & 854	CACCATGAATATGCAATGGGTAGG	TCAGGCAGATTCCAACCACT
H. thermophilus JR2	2507073746	chr (CS)	CACCATGTCATTGCCTGTCTTTTG	TTAGCCCGAAACAAACGACACC
		chr (reverse)	ATGTCATTGCTTGTCTTTTTGGC	CACCTTAGCCCGAAACAAACG
	2507074344	SulP	CACCATGACACAGGAAAACATAAAC	TCAATTTAATTCTTTATCGTCTTCTTTAAATTTTTTG
		SulP (reverse)	ATGACACAGGAAAACATAAACACAG	CACCTCAATTTAATTCTTTATCGTCTTCTTTAA
Tmr. frisia Kp2	2517375731	SbtA (CS)	CACCATGTTGGGATTGGATAGC	TTATACCGCTGAATACCACATAGC

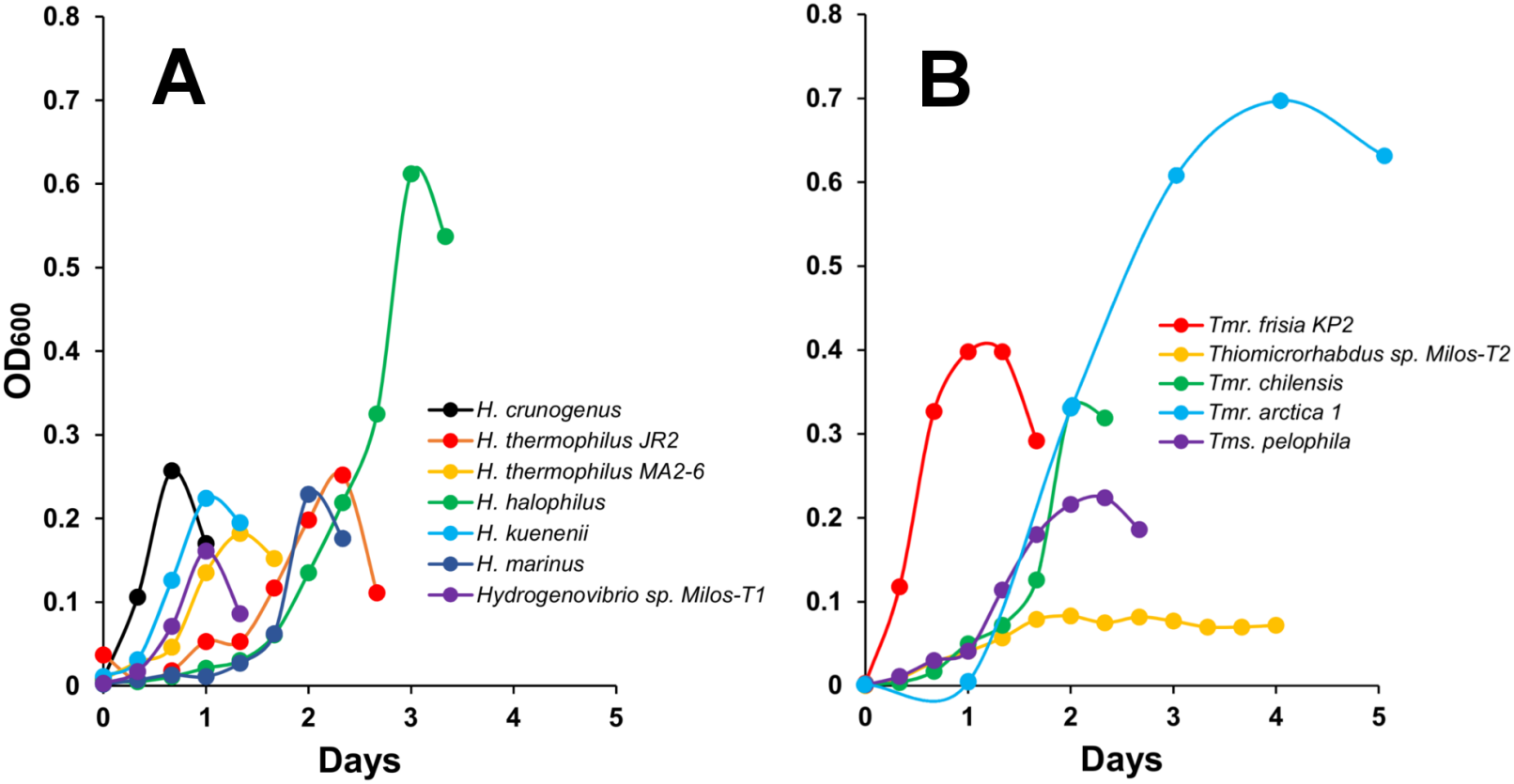
[a]Gene object identification numbers from the Integrated Microbial Genomes system (https://img.jgi.doe.gov/cgi-bin/m/main.cgi)

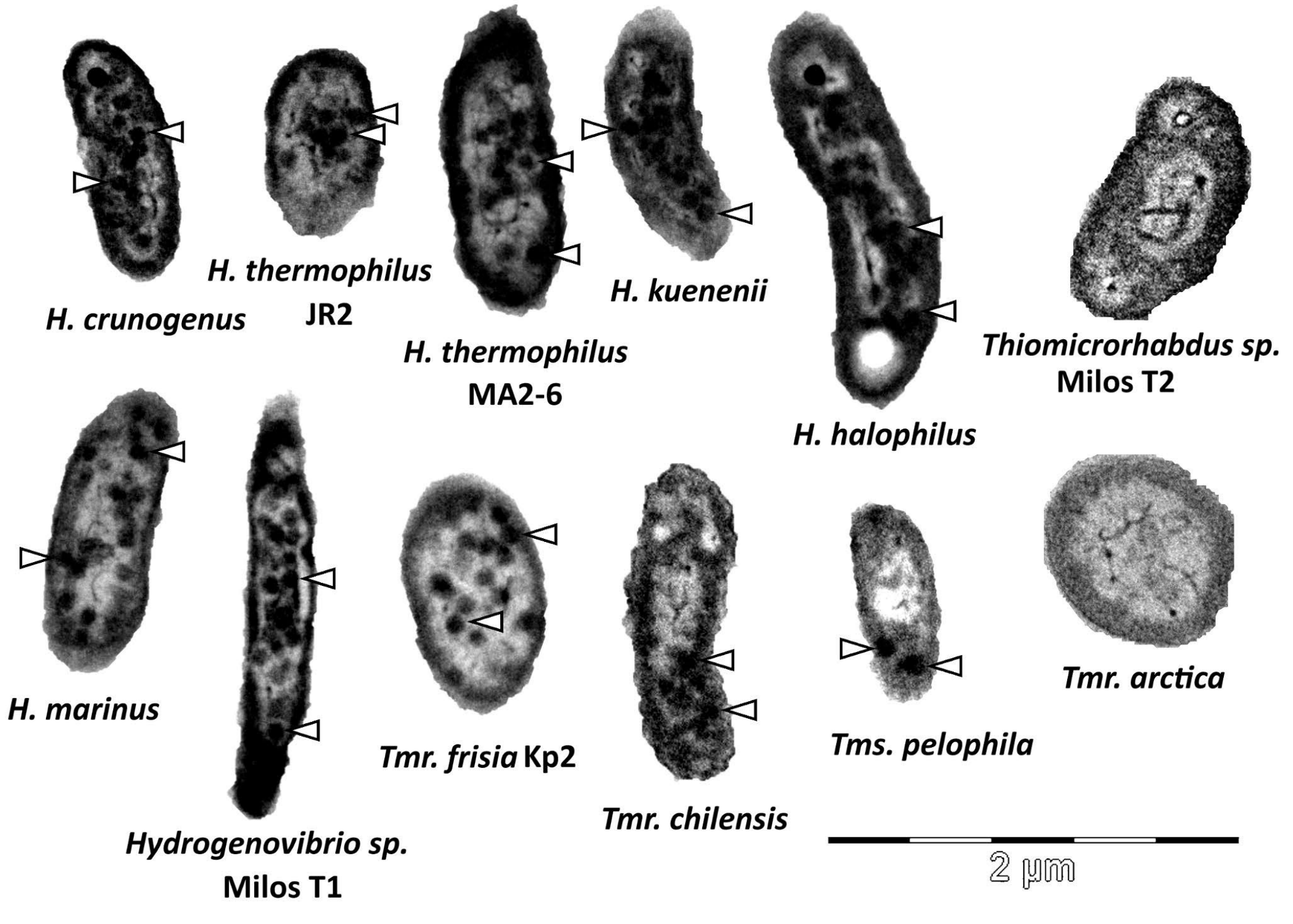
[b]Gene product abbreviations: Chr = chromate ion transporter family; SulP = sulfate permease family; 853, 854 = proteins encoded by Tcr_0853, 0854; SbtA = sodium-dependent bicarbonate transporter family. Gene product names followed by (CS) indicate that these genes are adjacent to those encoding the components of carboxysomes.

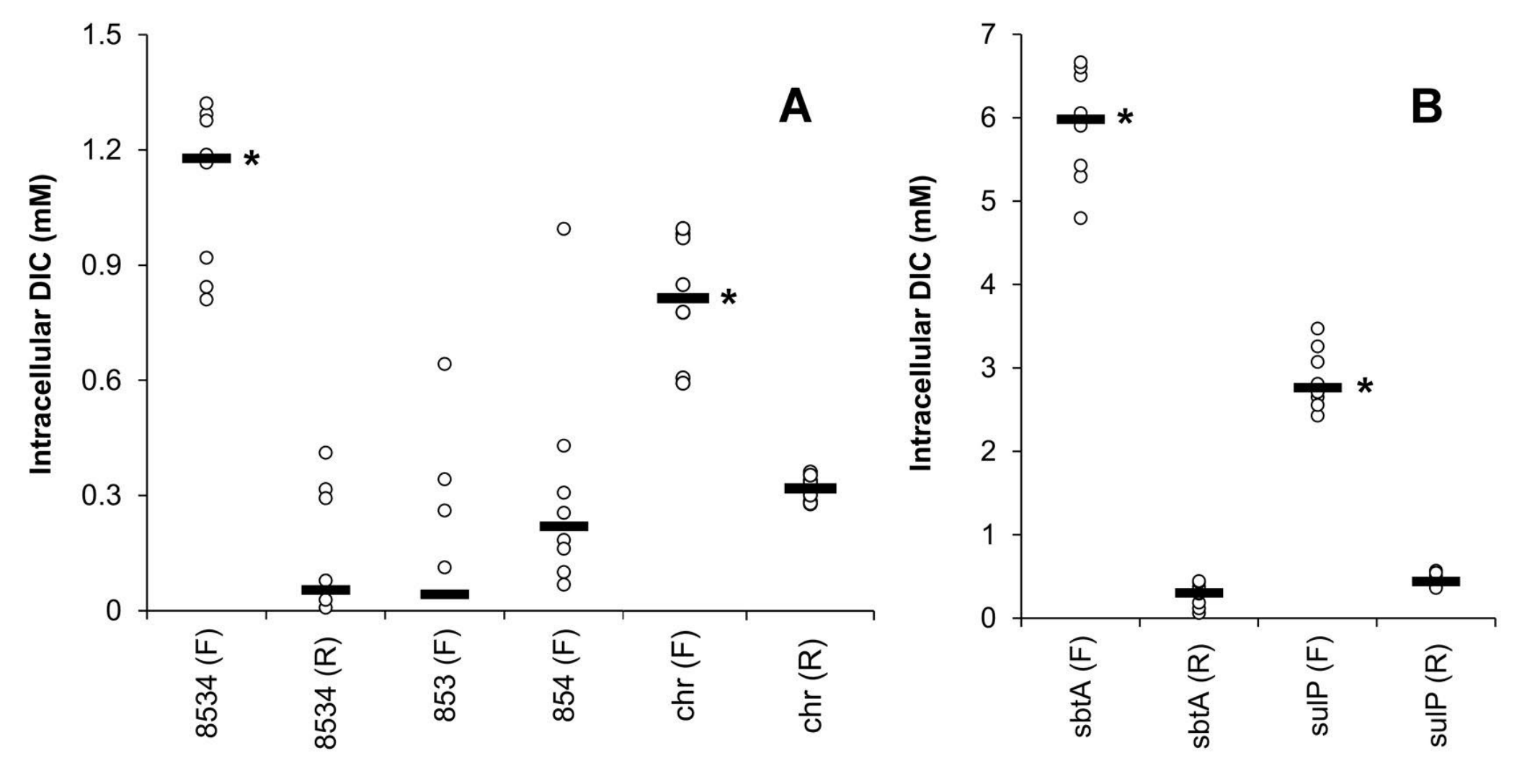

DNA helicase
HD-domain protein
methyl-accepting chemotaxis protein
imine deaminase
carboxysome genes
GTP pyrophosphokinase
RNA polymerase subunit omega
lipooligosaccharide synthesis genes
Thiomicrorhabdus frisia KP2
Thiomicrorhabdus sp. Milos-T2
Thiomicrorhabdus arctica
Thiomicrorhabdus chilensis
100
100
100

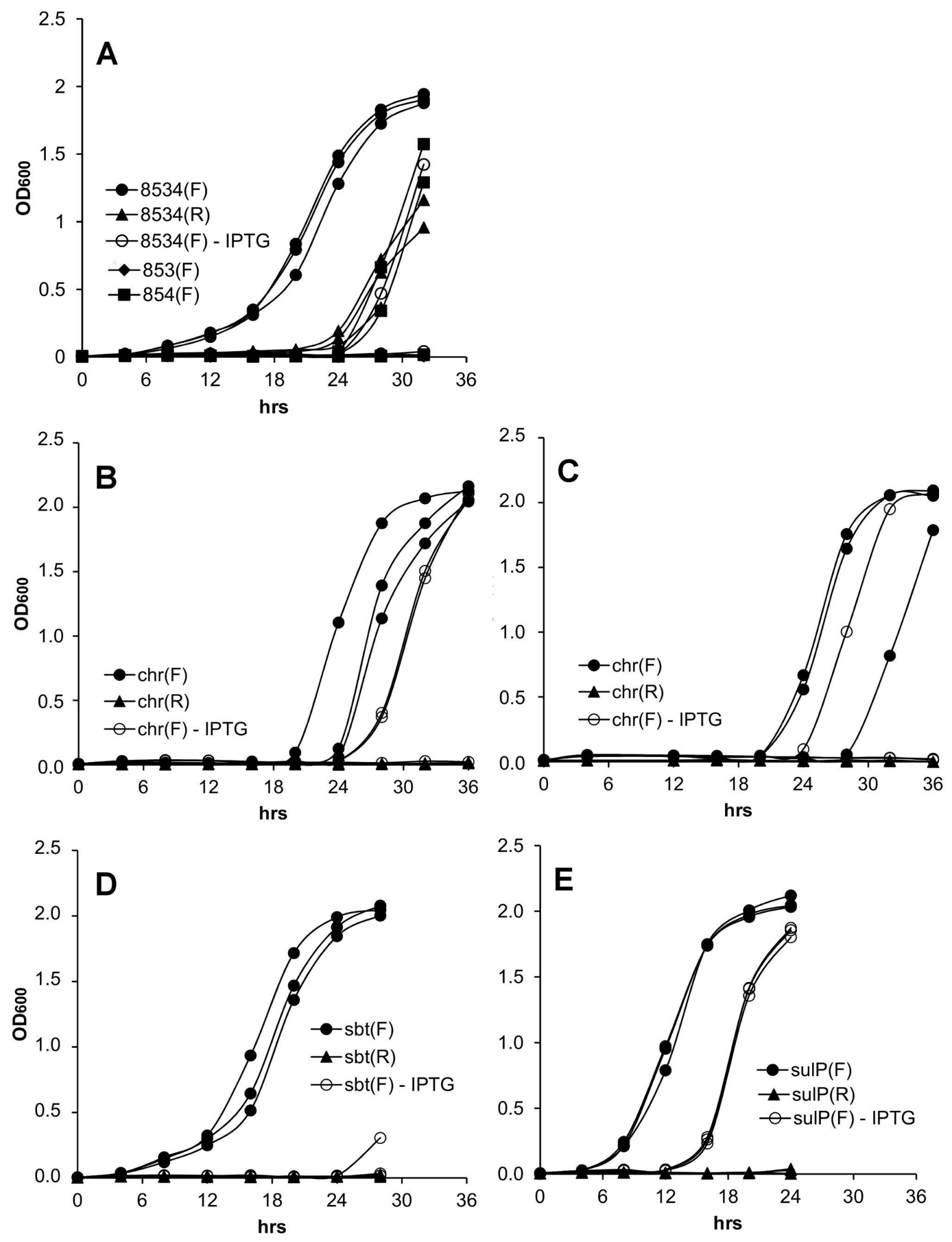

2623187719_Methylophaga_sulfidovorans
651613679_Methylophaga_aminisulfidivorans
639000843_Nitrococcus_mobilis
2514541535_FDH_Marinobacter_manganoxydans
2600951504_Haliea_rubra
2521802763_Rheinheimera_perlucida
2698518370_CS_Sulfurivirga_caldicuralii
2507074343_Hydrogenovibrio_thermophilus_JR2 Q+
2572249867_Hydrogenovibrio_thermophius_MA2-6 Q+
2678247398_Hydrogenovibrio_sp._XS5
2691253468_Hydrogenovibrio_sp._WB1
637785574_CS_Hydrogenovibrio_crunogenus E Q+
2574157468_CS_Hydrogenovibrio_marinus Q+
2541029969_CS_Hydrogenovibrio_kuenenii
2579719014_CS_Hydrogenovibrio_sp._Milos-T1 Q+
2568511024_Thiomicrospira_pelophila Q+
2540300460_Thiomicrorhabdus_chilensis
I
2566037955_Vibrio_rhizosphaerae
2646041880_Labilithrix_luteola
649659914_Nitrospira_defluvii
2581712788_Sulfobacillus_thermosulfidooxidans
2519479692_Leptospirillum_ferriphilum
2641367346_3sub_Nitrospira_moscoviensis
2620607600_OAOR_3sub_Sulfurovum_sp._AR
2735815037_OAOR_3sub_Hydrogenothermus_marinus
2504661583__Marinithermus_hydrothermalis
2682288181_OAOR_Hydrogenivirga_caldilitoris
637019965_Aquifex_aeolicus
650434263_OAOR_Hydrogenobacter_thermophilus
2729848210_OAOR_Hydrogenobacter_hydrogenophilus
646638463_Thermocrinis_albus
2512918850_OAOR_Thermocrinis_ruber
645270560_OAOR_Sulfurihydrogenibium_yellowstonense
2685851345_3sub_Thiocapsa_roseopersicina
2671657888_CS_3sub_Nitrosomonas_eutropha
651042208_CS_3sub_Acidithiobacillus_caldus
2571817147_CS_3sub_Acidithiobacillus_thiooxidans
2619157291_CS_3sub_Acidithiobacillus_ferrivorans
637964844_CS_3sub_Nitrobacter_hamburgensis
646799726_CS_3sub_Thiomonas_intermedia
2523763835_3sub_Thermithiobacillus_tepidarius
2515447367_CS_3sub_Thiobacillus_thioparus
637711359_CS_3sub_Thiobacillus_denitrificans
2520489544_3sub_Thioalkalivibrio_thiocyanoxidans
646383292_CS_3sub_Halothiobacillus_neapolitanus
2579719192_3sub_Hydrogenovibrio_sp._Milos-T1 Q+
2722617653_CS_3sub_Acidihalobacter_prosperus
2753019873_3sub_Acidihalobacter_ferrooxidans
2506751859_CS_3sub_Thiothrix_nivea
2565944614_CS_3sub_Thiorhodococcus_sp._AK35
639000746_CS_3sub_Nitrococcus_mobilis
2752847656_3sub_Tenderia_electrophaga
2596635193_CS_3sub_Thiohalorhabdus_denitrificans
II
A
639163609_Oceanobacter_sp._RED65
2598032568_Thalassolituus_oleivorans
2552789779_Pseudoalteromonas_ruthenica
2511577110_Glaciecola_nitratireducens
2525353299_Saccharospirillum_impatiens
0.4
YP_161712_BC_CHAR_Cupriavidus_metallidurans
637798372_CA_Synechococcus_elongatus
2606194899_Rhodobacteraceae_sp._HL-91
2688073919_Pseudoalteromonas_rubra
2540300141_Thiomicrorhabdus_chilensis
III
2742656188_Vibrio_cholerae
2524098047_Oceanospirillum_beijerinckii
2520465008_Oceanimonas_smirnovii
2611902202_Aeromonas_jandaei
2568510063_Thiomicrospira_pelophila Q-
637786436_Hydrogenovibrio_crunogenus Q-
2678249484_CS_Hydrogenovibrio_sp._XS5
2572249252_CS_Hydrogenovibrio_thermophilus_MA2-6 Q+
2507073746_CS_Hydrogenovibrio_thermophilus_JR2 E Q+
2507132611_Thiomicrospira_aerophila
2595506353_Thiomicrospira_sp._ALE5
2518266741_Hydrogenovibrio_halophilus Q-
2691253926_Thiomicrospira_sp._WB1
I
638962513_Synechococcus_sp._RS9917
2597846010_Pseudomonas_chlororaphis
2673855360_Acinetobacter_soli
2519849006_Paenibacillus_daejeonensis
2597800889_Bacillus_cereus
AAA88432_BC_CHAR_Pseudomonas_aeruginosa
637711362_CS_Thiobacillus_denitrificans
2515447400_cbbM_Thiobacillus_thioparus
2515456443_cbbM_Thiobacillus_denitrificans
2737566087_Nitrosomonas_oligotropha
2656318266_Acidithiobacillales_bacterium_SM23_46
643615170_CA_Thioalkalivibrio_sulfidophilus
2509047001_Thiocapsa_marina
2585414757_Methylobacter_whittenburyi
2603818313_Codakia_orbicularis_gill_symbiont
2634859977_Methylophilus_methylotrophus
644790919_Variovorax_paradoxus
650796122_Pseudomonas_mendocina
651048290_Collimonas_fungivorans
2599593923_Rheinheimera_pacifica
2579720167_Hydrogenovibrio_sp._Milos-T1 Q-
2541031494_Hydrogenovibrio_kuenenii
2574158185_Hydrogenovibrio_marinus Q-
II
B
0.3

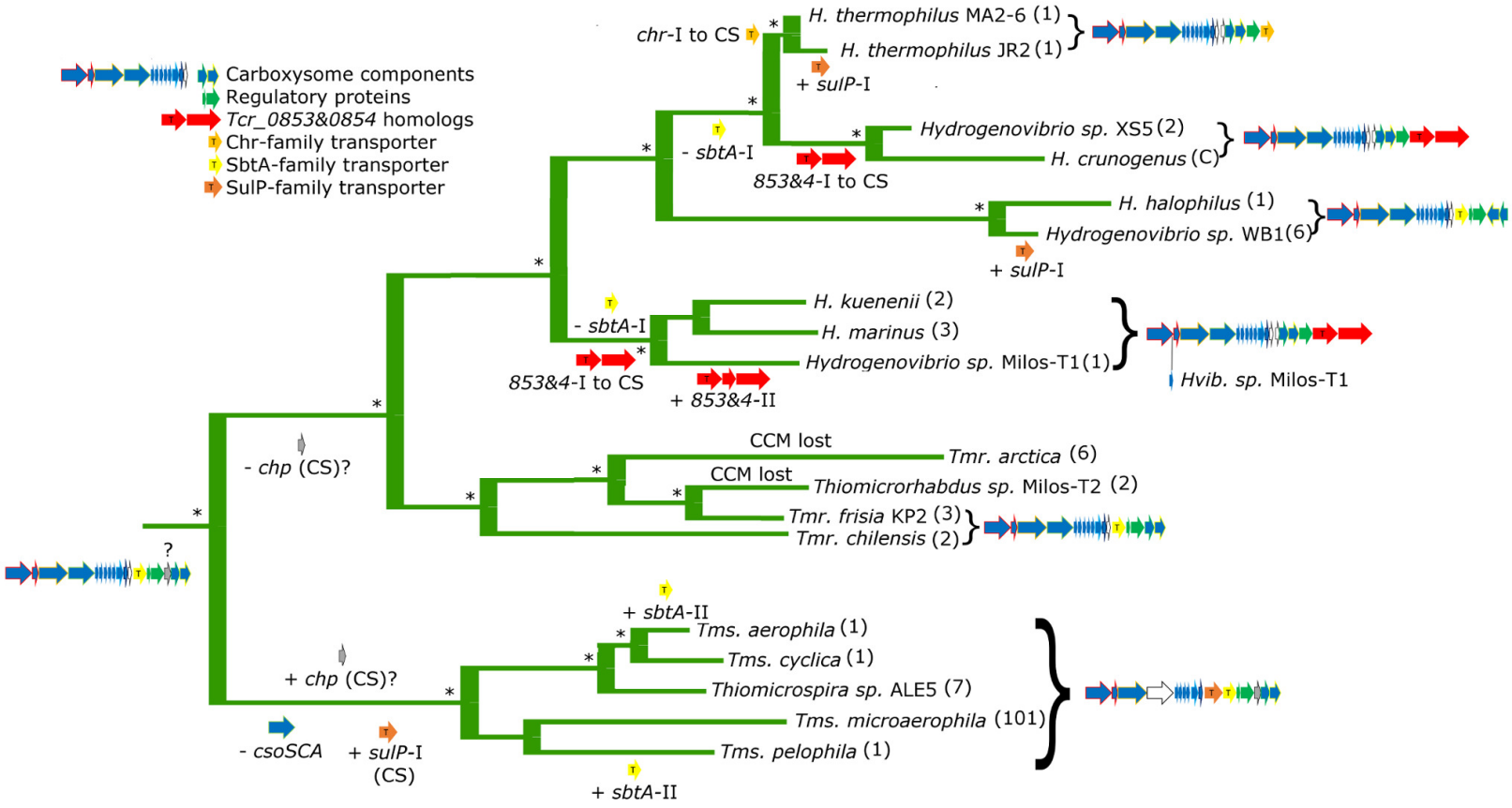

647590134_BC_CHAR_Cyanobium_sp._PCC_7001
2508552903_BC_CHAR_Cyanobium_gracile_PCC_6307
638959287_BC_CHAR_Synechococcus_sp._WH_5701
2514153780_BC_CHAR_Synechocystis_sp._PCC_6803
ABB57505_BC_CHAR_Synechococcus_elongatus_PCC_7942
641610480_BC_CHAR_Synechococcus_sp._PCC_7002
A
0.4
2519090797_Alicyclobacillus_pohliae
2705187210_Campylobacter_jejuni
650865199_Amycolicicoccus_subflavus
2595189716_CS_Thiomicrospira_microaerophila
2568510007_CS_Thiomicrospira_pelophila
2505700874_CS_Thiomicrospira_cyclica
2507134118_CS_Thiomicrospira_aerophila
2595506682_CS_Thiomicrospira_sp._ALE5
2518265315_CS_Hydrogenovibrio_halophilus
2691254503_CS_Hydrogenovibrio_sp._WB1
2517375731_CS_Thiomicrorhabdus_frisia_KP2
2540299972_CS_Thiomicrorhabdus_chilensis
Q+
E
I
2641367319_Nitrospira_moscoviensis
2521966036_CS_Thioalkalivibrio_nitratireducens
2685851707_Thiocapsa_roseopersicina
2588205630_Nitrosospira_briensis
2518933202_Methylophilus_methylotrophus
2688153519_Limnochorda_pilosa
2547731843_Dietzia_sp._UCD-THP
2507132718_Thiomicrospira_aerophila
2568511393_Thiomicrospira_pelophila
II
2513772414_Massilia_alkalitolerans
640088686_Methylibium_petroleiphilum
2599821844_Chitinimonas_taiwanensis
2584317540_Rhodoferax_saidenbachensis
2632217255_PUR_Verrucomicrobia_bacterium_IMCC26134
640017137_CS_Lyngbya_sp._PCC_8106
637447906_CS_Prochlorococcus_marinus
2503615380_CA_Chroococcidiopsis_thermalis
2551961126_CA_Scytonema_hofmanni
643741334_OAOR_Sulfurihydrogenibium_azorense
646638877_OAOR_Thermocrinis_albus
2522513265_CA_Salinimicrobium_terrae
2514224460_Ignavibacterium_album
640485624_cbbL_Rhodobacter_sphaeroides
2574369431_CA_Sphingopyxis_litoris
2523851166_Oceanicaulis_alexandrii
2558099744_CA_Terasakiella_pusilla
2656416673_PUR_Myxococcales_bacterium
642688950_CA_Leptospira_biflexa
2556975821_OAOR_Meiothermus_cerbereus
2579837198_PUR_Thermus_sp._YIM_77409
646789412_CA_Planctomyces_limnophilus
2517207030_PUR_Melioribacter_roseus
2581723989_PUR_Chrysiogenes_arsenatis
2671172190_PUR_Microbulbifer_yueqingensis
2523950464_PUR_Ferrimonas_senticii
644990019_PUR_Kangiella_koreensis
2504000947_PUR_Bacillus_setacea_BS02
2688062115_PUR_Pseudohongiella_spirulinae
2688810525_PyrC_Pseudomonas_alcaligenes
2598021551_Plesiomonas_shigelloides
2515665341_CS_Lamprocystis_purpurea
2684684513_CS_Tepidimonas_taiwanensis_MB2
2595507462_Thiomicrospira_sp._ALE5
2507132209_Thiomicrospira_aerophila
2505699308_Thiomicrospira_cyclica
2595191674_Thiomicrospira_microaerophila
IV
ANK06819_BC_CHAR_Bic_ychM_Escherichia_coli
P9WGF7_BC_CHAR_SO4_Mycobacterium_tuberculosis
640949669_Arcobacter_butzleri
2518265727_Hydrogenovibrio_halophilus
2568511001_Thiomicrospira_pelophila
2505700139_Thiomicrospira_cyclica
2507133101_Thiomicrospira_aerophila
2595505738_Thiomicrospira_sp._ALE5
2691254898_Hydrogenovibrio_sp._WB1
Q-
Q-
III
254854801_Alteromonas_sp._S89
2686861130_Lacimicrobium_alkaliphilum
2585413885_Methylobacter_whittenburyi
2523738162_Melitea_salexigens
2524097266_Oceanospirillum_beijerinckii
2636411254_Oleiphilus_messinensis
2640806512_Pseudoalteromonas_sp._OCN003
2698518947_Sulfurivirga_caldicuralii
637786269_Hydrogenovibrio_crunogenus
2540301473_CA_cbbM_Thiomicrorhabdus_chilensis
2562232889_Thiomicrorhabdus_sp._Milos-T2
2517376429_Thiomicrorhabdus_frisia_KP2
2522992675_Thiomicrorhabdus_arctica
2572249063_Hydrogenovibrio_thermophilus_MA2-6
2507073582_Hydrogenovibrio_thermophilus_JR2
2678249289_Hydrogenovibrio_sp._XS5
Q-
Q-
Q-
II
638462797_Desulfuromonas_acetoxidans
643528698_CA_Acidithiobacillus_ferrooxidans
2595506683_CS_Thiomicrospira_sp._ALE5
2595189715_CS_Thiomicrospira_microaerophila
2505700875_CS_Thiomicrospira_cyclica
2507134119_CS__Thiomicrospira_aerophila
2568510006_CS_Thiomicrospira_pelophila
2507074344_Hydrogenovibrio_thermophilus_JR2
2691253181_Hydrogenovibrio_sp._WB1
Q-
E
Q+
I
B
2671458394_cbbL_Nitrosospira_sp._17_Nsp14
2676387431_CS_Nitrosomonas_halophila
2686677610_Candidatus_Nitrospira_inopinata
2537260497_FHL_Leptospira_wolffii
2541556363__FHL_Leptospira_sp._B5-022
2523673439_Tenacibaculum_ovolyticum
2690112025_Ekhidna_lutea
0.5


A
B
OD600
Days
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
H. crunogenus
H. thermophilus JR2
H. thermophilus MA2-6
H. halophilus
H. kuenenii
H. marinus
Hydrogenovibrio sp. Milos-T1
Tmr. frisia KP2
Thiomicrorhabdus sp. Milos-T2
Tmr. chilensis
Tmr. arctica 1
Tms. pelophila


H. crunogenus
H. thermophilus JR2
H. thermophilus MA2-6
H. kuenenii
H. halophilus
Thiomicrorhabdus sp. Milos T2
H. marinus
Hydrogenovibrio sp. Milos T1
Tmr. frisia Kp2
Tmr. chilensis
Tms. pelophila
Tmr. arctica
2 µm


A
B
Intracellular DIC (mM)
Intracellular DIC (mM)
1.5
1.2
0.9
0.6
0.3
0
7
6
5
4
3
2
1
0
*
*
*
*
8534 (F)
8534 (R)
853 (F)
854 (F)
chr (F)
chr (R)
sbtA (F)
sbtA (R)
sulP (F)
sulP (R)


A
OD600
hrs
8534(F)
8534(R)
8534(F) - IPTG
853(F)
854(F)
B
chr(F)
chr(R)
chr(F) - IPTG
C
chr(F)
chr(R)
chr(F) - IPTG
D
sbt(F)
sbt(R)
sbt(F) - IPTG
E
sulP(F)
sulP(R)
sulP(F) - IPTG


Carboxysome components
Regulatory proteins
Tcr_0853&0854 homologs
Chr-family transporter
SbtA-family transporter
SulP-family transporter
?
- chp (CS)?
+ chp (CS)?
- csoSCA
+ sulP-I (CS)
+ sbtA-II
Tms. aerophila (1)
Tms. cyclica (1)
Thiomicrospira sp. ALE5 (7)
Tms. microaerophila (101)
Tms. pelophila (1)
+ sbtA-II
CCM lost
Tmr. arctica (6)
CCM lost
Thiomicrorhabdus sp. Milos-T2 (2)
Tmr. frisia KP2 (3)
Tmr. chilensis (2)
- sbtA-I
853&4-I to CS
+ 853&4-II
H. kuenenii (2)
H. marinus (3)
Hydrogenovibrio sp. Milos-T1 (1)
Hvib. sp. Milos-T1
- sbtA-I
853&4-I to CS
Hydrogenovibrio sp. XS5 (2)
H. crunogenus (C)
H. halophilus (1)
Hydrogenovibrio sp. WB1 (6)
+ sulP-I
chr-I to CS
H. thermophilus MA2-6 (1)
H. thermophilus JR2 (1)
+ sulP-I