

2019-02

Diversity in CO₂ concentrating mechanisms among chemolithoautotrophs from genera *Hydrogenovibrio*, *Thiomicrothrix*, and *Thiomicrospira*, ubiquitous in sulfidic habitats worldwide

Scott, KM

<http://hdl.handle.net/10026.1/12732>

10.1128/AEM.02096-18

Applied and Environmental Microbiology

American Society for Microbiology

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

1 **Diversity in CO₂ concentrating mechanisms among**
2 **chemolithoautotrophs from genera *Hydrogenovibrio*,**
3 ***Thiomicrothrix*, and *Thiomicrospira*, ubiquitous in sulfidic**
4 **habitats worldwide**

5
6 **Kathleen M. Scott^{1#}, Juliana M. Leonard¹, Rich Boden², Dale Chaput³, Clare**
7 **Dennison¹, Edward Haller¹, Tara L. Harmer⁴, Abigail Anderson¹, Tiffany Arnold¹,**
8 **Samantha Budenstein¹, Rikki Brown¹, Juan Brand¹, Jacob Byers¹, Jeanette**
9 **Calarco¹, Timothy Campbell¹, Erica Carter¹, Max Chase¹, Montana Cole¹, Deandra**
10 **Dwyer¹, Jonathon Grasham¹, Christopher Hanni¹, Ashlee Hazle¹, Cody Johnson¹,**
11 **Ryan Johnson¹, Brandi Kirby¹, Katherine Lewis¹, Brianna Neumann¹, Tracy**
12 **Nguyen¹, Jonathon Nino Charari¹, Ooreoluwa Morakinyo¹, Bengt Olsson¹,**
13 **Shanetta Roundtree¹, Emily Skjerve¹, Ashley Ubaldini¹, and Robert Whittaker¹.**
14

15 ¹Department of Integrative Biology, University of South Florida, Tampa, Florida, United
16 States of America

17 ²School of Biological & Marine Sciences, University of Plymouth, Drake Circus,
18 Plymouth, UK; Sustainable Earth Institute, University of Plymouth, Drake Circus,
19 Plymouth, UK,

20 ³Proteomics and Mass Spectrometry Facility, University of South Florida, Tampa,
21 Florida, United States of America,

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

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

24
25 Running title: Diverse CCMs in sulfur-oxidizing autotrophs
26
27

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

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

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

62 **KEYWORDS** CO₂ concentrating mechanism, chemolithoautotroph, autotroph, carbon
63 fixation

64

65

66

67 (INTRODUCTION)

68 Autotrophic members of domains *Bacteria* and *Archaea* are responsible for introducing
69 carbon into the biological portion of the global carbon cycle in virtually any habitat with
70 sufficient light or chemical energy to power the process of carbon fixation. They use
71 CO₂ from the air, or dissolved inorganic carbon (DIC = CO₂ + HCO₃⁻ + CO₃²⁻), if aquatic,
72 as their carbon source, and have a variety of mechanisms to compensate for variability
73 in the availabilities of these compounds.

74 CO₂-concentrating mechanisms (CCMs) are one such mechanism, and have
75 been particularly well-studied among members of the phylum "*Cyanobacteria*". In these
76 organisms, active transport of HCO₃⁻ elevates its intracellular concentration.

77 Cytoplasmic HCO₃⁻ enters carboxysomes. These protein-bound microcompartments
78 contain the enzymes carbonic anhydrase (EC 4.2.1.1), and ribulose 1,5-bisphosphate
79 carboxylase/oxygenase (RubisCO, EC 4.1.1.39). These enzymes act together to
80 dehydrate some of the HCO₃⁻ to form CO₂, and use the CO₂ to carboxylate ribulose
81 bisphosphate, leading to the formation of 3-phosphoglycerate for biosynthesis (1-3).

82 The HCO₃⁻ transporters characterized in members of phylum "*Cyanobacteria*" fall into
83 three evolutionarily independent lineages: BCT1, an ABC transporter (4); BicA, a
84 member of the SulP family of transporters (5); and SbtA (6). Loss of cytoplasmic DIC is
85 minimized by conversion of cytoplasmic CO₂ to HCO₃⁻ via membrane-associated
86 carbonic anhydrases, which couple CO₂ hydration to redox reactions (7, 8). This
87 arsenal of DIC transporters and traps is distributed among members of "*Cyanobacteria*"
88 based on their habitats; those inhabiting freshwater and sediments, in which DIC
89 concentration and composition can vary most greatly (e.g., due to pH differences) tend

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

96 CCMs are not well-characterized among autotrophs from the many other phyla of
97 *Bacteria* and *Archaea* with autotrophic members. Carboxysomes are present in many
98 autotrophic members of *Alpha-*, *Beta-*, and *Gammaproteobacteria*, and their structure
99 and function have been well-characterized for *Halothiobacillus neapolitanus* Parker X^T
100 from the *Chromatiales* of the *Gammaproteobacteria* (10-12). DIC uptake has only been
101 studied in detail for *Hydrogenovibrio crunogenus*, a sulfur-oxidizing chemolithoautotroph
102 from the *Thiotrichales* of the *Gammaproteobacteria*, isolated from deep-sea
103 hydrothermal vents (13). This organism generates elevated intracellular DIC
104 concentrations in an energy-dependent manner (14, 15), and has carboxysomes (16),
105 which likely facilitates its ability to grow rapidly under low-DIC conditions (14). Random
106 and site-directed mutagenesis of gene loci Tcr_0853 and Tcr_0854, which are located
107 downstream of the carboxysome locus in this organism, result in a high-CO₂ requiring
108 phenotype, and loss of an ability to generate high intracellular concentrations of DIC,
109 suggesting that these genes encode a two-component DIC transporter (17). Homologs
110 of these genes are common among autotrophic *Bacteria*, and one of them (Tcr_0854) is
111 from a PFAM without prior biochemical characterization (PFAM10070; (17)).

112 Several members of *Thiomicrospira*, *Thiomicrothrix*, and *Hydrogenovibrio*,
113 organisms taxonomically affiliated with *H. crunogenus* (18), have had their genomes
114 sequenced. Taxa were selected for sequencing to represent both the taxonomic
115 breadth of these genera, as well as the range of habitats from which these organisms
116 have been isolated, including shallow and deep-sea hydrothermal vents, coastal
117 sediments, and soda and salt lakes (19). Despite the rather narrow taxonomic range of
118 the organisms sequenced, a surprising diversity in mechanisms for DIC uptake and
119 fixation was suggested from the genome data. Genome sequences of some members
120 of *Thiomicrothrix* lack carboxysome loci altogether, suggesting the absence of a
121 CCM. For members of *Thiomicrospira*, genes encoding carboxysome components are
122 present, but those encoding carboxysomal carbonic anhydrase are lacking. Instead,
123 they each carry a gene in its place without apparent homologs beyond this genus (19),
124 raising the possibility that these genes might encode a novel form of carboxysomal
125 carbonic anhydrase. In all cases, when present, carboxysome loci are followed by
126 genes encoding transporters from four evolutionarily distinct families. Carboxysome
127 locus-associated genes encoding potential transporters include homologs to those
128 encoding the potential DIC transporter in *H. crunogenus* (Tcr_0853, Tcr_0854), and
129 members of the SulP and SbtA families distantly related to those known to transport
130 HCO₃⁻ in members of "*Cyanobacteria*". Also included are members of the Chr family,
131 which is widely distributed among prokaryotes. The two biochemically characterized
132 members of this family confer resistance to chromate by extruding this anion, perhaps
133 functioning as a chromate/sulfate antiporter (20).

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

142

143 RESULTS

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

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

167 **Growth under low-DIC conditions, and DIC concentrations *in situ*.** All taxa
168 tested here grew under high-DIC conditions; all but *Thiomicrothabodus* sp. Milos-T2 were
169 capable of growth under low-DIC conditions (2 mM DIC under ~400 ppm CO₂ ambient
170 headspace; Fig. 4). *Tmr. arctica* lacks genes encoding carboxysomes; therefore,
171 growth under low-DIC conditions was unexpected, motivating further consideration of its
172 habitat. CO₂ concentrations estimated for the marine Arctic sediments from which *Tmr.*
173 *arctica* was isolated are lower than those at the hydrothermal vent habitat from which *H.*
174 *crunogenus* was isolated (Table 1).

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

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

185 **Response of transporter transcript abundances to DIC concentration.**

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

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

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

201 heterologous expression based on collocation with the carboxysome locus and
202 upregulation under low-DIC conditions (Table 2): the two-component transporter from *H.*
203 *crunogenus* XCL-2 (Tcr_0853, Tcr_0854), members of Chr and SulP transporter
204 families from *H. thermophilus* JR2, and a member of the SbtA transporter family from
205 *Tmr. frisia* KP2. Mass spectrometric analysis of proteins from membranes from *E. coli*
206 constructs expressing these transporters verified their expression (Table 3). Signal
207 intensity was always low for the protein product of Tcr_0853, which may reflect that this
208 protein is likely to be particularly hydrophobic (11 predicted membrane-spanning alpha
209 helices), and therefore more difficult to solubilize, digest, elute from the C₁₈ column used
210 to resolve the peptides, and ionize for mass spectrometry.

211 *E. coli* expressing these genes were able to generate elevated intracellular DIC
212 concentrations (Fig. 6; Fig. S6). Intracellular DIC was similar to extracellular when
213 putative DIC transporters were oriented in reverse relative to the T7 promotor driving
214 their expression. When genes were correctly oriented relative to the T7 promoter,
215 intracellular DIC concentrations were higher than when in reverse orientation (Fig. 6; for
216 *sbtA*, *chr*, *sulP*: two-tailed *t*-test, F versus R orientations, $\alpha < 0.05$; for 8534 (F),
217 8534(R), 853(F), 854(F): ANOVA, with post-hoc multiple comparisons via Scheffe,
218 Bonferroni, and Tukey tests; $\alpha < 0.05$ for 8534 (F)). Intracellular DIC was particularly
219 high for cells expressing SbtA. The presence of both Tcr_0853 and Tcr_0854 were
220 necessary for the accumulation of intracellular DIC, suggesting that the gene products
221 of both are required for DIC uptake, and that they may form a two-subunit transporter.
222 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 CO₂
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*, *Thiomicrothrix*, 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.

245 *Tmr. arctica* and *Thiomicrospira* sp. Milos-T2 do not appear to have CCMs,
246 based on an absence of carboxysome loci and homologs to most of the transporter
247 genes associated with these loci. The inability of *Thiomicrospira* sp. Milos-T2 to
248 grow under low-DIC conditions is consistent with these genome traits. In contrast,
249 growth by *Tmr. arctica* under low-DIC conditions was surprising. This organism may
250 have a novel mechanism for growing under low-DIC conditions. Alternatively, its growth
251 may be facilitated by higher CO₂ concentrations that result from the lower temperatures
252 at which this organism grows, since lower temperatures increase the solubility of CO₂,
253 and also increase the pK_a of bicarbonate (24). However, these physical and chemical
254 factors do not appear likely to result in particularly high concentrations of CO₂ in the
255 habitat from which *Tmr. arctica* was isolated (Table 1). Perhaps the lower maximum
256 specific growth rates observed for this psychrophilic organism (25), may render
257 carboxysomes and CCMs unnecessary.

258 The branching order predicted from supertrees constructed for members of
259 *Thiomicrospira*, *Thiomicrospira*, and *Hydrogenovibrio* (18, 19); Fig. 1) suggests that
260 the loss of carboxysome loci may have occurred independently in the lineages leading
261 to *Tmr. arctica* and *Thiomicrospira* sp. Milos-T2. CCMs may not have provided a
262 selective advantage for these organisms. Given the size of carboxysomes, as well as
263 their abundance when expressed (21), carboxysome loss would provide an energetic
264 advantage for cells growing in habitats with consistently elevated concentrations of CO₂.
265 *Thiomicrospira* sp. Milos-T2 was cultured from the same hydrothermal vent system
266 as *Hydrogenovibrio* sp. Milos-T1 (26), which has a carboxysome locus. Perhaps they
267 inhabit different niches with different CO₂ abundances in this system.

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

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

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

293 DIC uptake by some transporters from "*Cyanobacteria*" cannot be successfully
294 assayed by silicone oil centrifugation and complementation of growth of carbonic
295 anhydrase-deficient *E. coli* (28). Perhaps shared membership with *E. coli* in the class
296 *Gammaproteobacteria*, or the stronger T7 promoter used for expression here, were
297 responsible for successful heterologous expression and activity. Successful
298 heterologous expression of the genes studied here bodes well for their potential use in
299 constructing organisms capable of synthesizing industrially relevant precursor
300 compounds from CO₂ and HCO₃⁻.

301 Multiple clades of each transporter family are present among the organisms
302 studied here. Among these organisms, homologs to Tcr_0853 and 0854, as well as
303 SbtA, fall into two clades each. In both cases, most of the genes fall together into a
304 single well-supported clade. Genes outside of this clade fall among those present in
305 distantly related members of "*Proteobacteria*", and organisms carrying these genes also
306 carry a copy falling within the clade. These 'extra copies' could have been relatively
307 recently acquired via horizontal gene transfer. Representatives from both clades from
308 both transporter families all have elevated transcript levels when cells are grown under
309 low-DIC conditions. Furthermore, genes encoding these two types of transporters are
310 usually present adjacent to genes encoding CO₂-metabolizing enzymes (Fig. 2; Fig. 3);
311 this collocation, as well as upregulation under low CO₂ conditions suggests that
312 members of these transporter families may predominantly transport DIC.

313 The only members of the Chr and SulP families to have elevated transcript levels
314 under low-DIC conditions fell within a single clade of each family (Table 2; Fig. 2; Fig.
315 3). The members of the Chr and SulP families tested here whose transcript levels were
316 not sensitive to DIC concentrations and are not collocated with carboxysome genes
317 may not play a role in CCMs, and may instead be either constitutively expressed DIC
318 transporters, or transport sulfate or other cations. Few members of these transporter
319 families are collocated with genes encoding CO₂-metabolizing enzymes (Fig. 2; Fig. 3),
320 suggesting that roles in DIC uptake may be less widespread in these transporter
321 families. However, it is important to note that a SulP-family transporter has been
322 implicated in DIC uptake in *E. coli* (5, 29).

323 The SulP-family transporters studied here have domains that distinguish them
324 from other members of this transporter family. Similar to the SulP transporters present
325 in "*Cyanobacteria*", they lack the β -carbonic anhydrase domain found in some members
326 of this family. When present in other organisms, this carbonic anhydrase domain is
327 located on the cytoplasmic side of the cell membrane. This absence of a carbonic
328 anhydrase domain is consistent with transporting HCO₃⁻ into the cytoplasm to generate
329 elevated DIC concentrations there; if it were present, a carbonic anhydrase domain
330 would convert the transported HCO₃⁻ into CO₂ at the cell membrane, where it would
331 diffuse out of the cell (30). Unlike those present in "*Cyanobacteria*", the SulP-family
332 transporters found in the organisms studied here lack the carboxy-terminal STAS
333 domain typically found in SulP proteins. The STAS domain is hypothesized to regulate
334 the activities of these transporters (31, 32). HCO₃⁻ transporters in "*Cyanobacteria*" are
335 post-translationally regulated, and inactive in the dark (reviewed in (33)). Though this

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

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

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

383 To test for ability to grow under low-DIC conditions, organisms were cultivated at
384 20°C in thiosulfate-supplemented artificial seawater (TASW; (14), pH 7.5, 15 µg/L
385 vitamin B-12). For *H. halophilus*, NaCl was raised to 1.5 M (43), and for *Tmr. arctica*, the
386 culture temperature was 10°C. High DIC cultures (50 mM NaHCO₃, 5% headspace
387 CO₂) in TASW were used to inoculate paired flasks containing high DIC and low-DIC
388 medium (2 mM NaHCO₃, 0.04% ambient air headspace CO₂; one low-DIC and high-DIC
389 culture per strain). Turbidity was monitored at 600 nm.

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

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

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

407 **qRT-PCR assay of transcript abundances from genes encoding**
408 **carboxysome components and potential DIC transporters.** To determine whether
409 carboxysome-associated transporter gene transcripts were more abundant when cells
410 were grown under low-DIC conditions, taxa that were amenable to cultivation in
411 chemostats were grown under two conditions: DIC limitation (low-DIC), and NH₃
412 limitation (high-DIC; (14). Cells were harvested via centrifugation (10,000 × *g*, 5 min,
413 4°C), flash-frozen with liquid nitrogen, and stored at -80°C for subsequent RNA
414 extraction using the Ambion RiboPure-Bacteria Kit (17). Primers were designed to
415 target genes encoding carboxysome components (*csoS2*; *csoS3*; positive control for
416 CCM induction), citrate synthase (calibrator for the $2^{-\Delta\Delta C_t}$ method; (47), and transporters
417 (Table 4), and qRT-PCR assays were implemented in an Applied Biosystems Step One
418 real-time PCR system, using QuantiTect SYBR Green RT-PCR (Qiagen, Inc.) as
419 described in (17).

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

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

443 *E. coli* carrying genes encoding potential DIC transporters were cultivated on a
444 gyrotary shaker (150 rpm, 37°C) in lysogeny broth supplemented with 100 mg/L
445 ampicillin and 30 mg/L chloramphenicol. Since rhamnose addition was not found in
446 pilot experiments to enhance expression of the target genes, it was not added to the
447 growth media. When OD₆₀₀ reached 0.5 – 1.2, IPTG was added (0.4 mM), and cells

448 were cultivated for another four hours at 30°C. Cells were harvested (10,000 × g 5 min,
449 4°C). For proteomic analysis, cells were stored at -20°C. For DIC uptake assays, cells
450 were resuspended in fresh medium (OD₆₀₀ ~50 0.4 mM IPTG) and stored on ice until
451 use, within 1 hr of harvest.

452 **Membrane preparations and proteomics.** To verify heterologous expression
453 of the potential DIC transporter genes, membranes were prepared from *E. coli* grown as
454 described above. Pellets from 20 ml cultures were thawed and resuspended in 10 ml
455 membrane buffer (50 mM TRIS pH 8, 10 mM EDTA) supplemented with chicken egg
456 lysozyme (0.1 mg/ml), and incubated for 30 min at 20°C. Lysate was sonicated on ice
457 for 15 sec to decrease viscosity, and centrifuged to remove debris and intact cells
458 (6,000 × g, 30 min, 4°C). Supernatant was centrifuged to pellet the membranes (75,000
459 × g, 30 min, 4°C). Pellets (membranes) were rinsed twice with membrane buffer.

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

466 Peptides were separated using a 50cm C₁₈ reversed-phase-HPLC column on an
467 Ultimate3000 UHPLC system (Thermo Fisher Scientific) with a 60 minute gradient (4-
468 40% acetonitrile with 0.1% formic acid) and analyzed on a hybrid quadrupole-Orbitrap
469 instrument (Q Exactive Plus, Thermo Fisher Scientific) using data-dependent
470 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 μ l portions of
481 suspended cells were added to 200 μ l lysogeny broth, 50 mM HEPES pH 8, 0.25 mM
482 DI^{14}C (2 mCi/ml $\text{NaH}^{14}\text{CO}_3$, 15 mCi/mmol; MP Biomedicals, Inc., Irvine, CA). These
483 200 μ l 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 $\times g$) 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., $^{14}\text{CO}_2$ 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 μ ci/ml, Amersham
493 Biosciences, Little Chalfont, UK). Cytoplasm volumes were calculated from cell

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

501 *E. coli* EDCM636 is only capable of growth under high DIC conditions due to
502 disruption of its β -carbonic anhydrase gene with a kanamycin resistance cartridge (22).
503 When genes encoding DIC transporters from "*Cyanobacteria*" are expressed by this
504 strain, it is capable of growing under ambient atmosphere (~400 ppm CO₂; (28). A
505 culture of this strain was obtained from the Coli Genetic Stock Center at Yale University
506 to screen transporters for DIC uptake activity. Unlike (28), in which DIC transporter
507 expression was driven by the *lac* promoter, target gene expression in this study was
508 driven by the T7 promoter. Since the transporter genes carried on pET101/D-TOPO
509 (described above) require T7 RNA polymerase for expression, it was necessary to
510 introduce a derivative of plasmid pAR1219 carrying an IPTG-inducible copy of the gene
511 encoding this enzyme (53) into *E. coli* EDCM636. Since both pET101/D-TOPO and
512 pAR1219 confer resistance to ampicillin, it was necessary to modify pAR1219
513 beforehand by interrupting its beta lactamase gene with a trimethoprim resistance
514 cartridge using the EZ-Tn5 <DHFR-1> insertion kit (Epicentre). Chemically competent
515 *E. coli* EDCM636 were transformed first with modified pAR1219 (conferring
516 trimethoprim resistance, but not ampicillin resistance). These cells were subsequently

517 transformed with pET101/D-TOPO plasmids carrying candidate DIC transporters, and
518 screened for an ability to grow under low DIC conditions. Lysogeny broth (100 mg/L
519 ampicillin, 50 mg/L kanamycin, 50 mg/L trimethoprim) was inoculated with each strain,
520 and 5-fold serial dilutions were prepared for each. These overnight cultures were grown
521 overnight under a headspace of 5% CO₂. Since pseudo-revertants capable of growing
522 under low-DIC conditions are frequent in *E. coli* EDCM636 (22, 28), the highest-titer
523 overnight culture was divided into two equal portions. One portion was incubated under
524 5% CO₂ with the other cultures, while the other was incubated under ambient air
525 overnight. The next morning, the absence of growth for the culture incubated under
526 ambient air was verified. The optical density of the serial dilutions incubated under 5%
527 CO₂ was measured, and cultures with OD₆₀₀ = 0.3 – 0.5 were selected to be used as
528 inocula. These cultures were split into two portions, one was brought to 0.4 mM IPTG,
529 and all were incubated another hour under 5% CO₂. Each was then added 1:100 v/v to
530 three portions of fresh lysogeny broth supplemented with the antibiotics described
531 above, plus 0.4 mM IPTG as appropriate for the experiment. These cultures were
532 incubated (30°C, 150 rpm) under ambient atmosphere and monitored for growth
533 spectrophotometrically at 600 nm.

534

535 SUPPLEMENTAL MATERIAL

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

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

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

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

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

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

548

549

550 **ACKNOWLEDGMENTS**

551 We are grateful for support from the National Science Foundation ((NSF-IOS-
552 1257532 to K.M.S.) and the University of South Florida Pilot Studies Awards program.

553 We are also grateful to Gordon Cannon for discussion and comments on an earlier
554 version of this manuscript, as well as the insightful comments of the anonymous
555 reviewers of this manuscript. Author contribution statement: KMS designed the study;
556 all authors conducted experiments, and contributed to data analysis and writing the
557 manuscript.

558

559 REFERENCES

- 560 1. **Price GD.** 2011. Inorganic carbon transporters of the cyanobacterial CO₂ concentrating
561 mechanism. *Photosynth Res* **109**:47-57.
- 562 2. **Whitehead L, Long BM, Price GD, Badger MR.** 2014. Comparing the in vivo function of alpha-
563 carboxysomes and beta-carboxysomes in two model cyanobacteria. *Plant Physiol* **165**:398-411.
- 564 3. **Kerfeld CA, Melnicki MR.** 2016. Assembly, function and evolution of cyanobacterial
565 carboxysomes. *Curr Opin Plant Biol* **31**:66-75.
- 566 4. **Omata T, Takahashi Y, Yamaguchi O, Nishimura T.** 2002. Structure, function, and regulation of
567 the cyanobacterial high-affinity bicarbonate transporter, BCT1. *Funct Plant Biol* **29**:151-159.
- 568 5. **Price G, Woodger F, Badger M, Howitt S, Tucker L.** 2004. Identification of a SulP-type
569 bicarbonate transporter in marine cyanobacteria. *Proc Natl Acad Sci USA* **101**:18228-18233.
- 570 6. **Shibata M, Katoh H, Sonoda M, Ohkawa H, Shimoyama M, Fukuzawa H, Kaplan A, Ogawa T.**
571 2002. Genes essential to sodium-dependent bicarbonate transport in cyanobacteria. *J Biol Chem*
572 **277**:18658-18664.
- 573 7. **Shibata M, Ohkawa H, Kaneko T, Fukuzawa H, tabata S, Kaplan A, Ogawa T.** 2001. Distinct
574 constitutive and low-CO₂-induced CO₂ uptake systems in cyanobacteria: Genes involved and
575 their phylogenetic relationship with homologous genes in other organisms. *Proc Natl Acad Sci*
576 **98**:11789-11794.
- 577 8. **Han X, Sun N, Xu M, Mi H.** 2017. Co-ordination of NDH and Cup proteins in CO₂ uptake in
578 cyanobacterium *Synechocystis* sp. PCC 6803. *J Exp Bot* **68**:3869-3877.
- 579 9. **Badger MR, Price GD, Long BM, Woodger FJ.** 2006. The environmental plasticity and ecological
580 genomics of the cyanobacterial CO₂ concentrating mechanism. *J Exp Bot* **57**:249-265.
- 581 10. **Heinhorst S, Williams E, Cai F, Murin D, Shively J, Cannon G.** 2006. Characterization of the
582 carboxysomal carbonic anhydrase CsoCSA from *Halothiobacillus neapolitanus*. *J Bacteriol*
583 **188**:8087-8094.
- 584 11. **Dou Z, Heinhorst S, Williams E, Murin E, Shively J, Cannon G.** 2008. CO₂ fixation kinetics of
585 *Halothiobacillus neapolitanus* mutant carboxysomes lacking carbonic anhydrase suggest the
586 shell acts as a diffusional barrier for CO₂. *J Biol Chem* **283**:10377-10384.
- 587 12. **Sutter M, Roberts EW, Gonzalez RC, Bates C, Dawoud S, Landry K, Cannon GC, Heinhorst S,**
588 **Kerfeld CA.** 2015. Structural characterization of a newly identified component of alpha-
589 carboxysomes: The AAA plus domain protein CsoCbbQ. *Sci Rep* **5**:16243.
- 590 13. **Jannasch H, Wirsén C, Nelson D, Robertson L.** 1985. *Thiomicrospira crunogena* sp. nov., a
591 colorless, sulfur-oxidizing bacterium from a deep-sea hydrothermal vent. *Int J Syst Bacteriol*
592 **35**:422-424.
- 593 14. **Dobrinski KP, Longo DL, Scott KM.** 2005. A hydrothermal vent chemolithoautotroph with a
594 carbon concentrating mechanism. *J Bacteriol* **187**:5761-5766.
- 595 15. **Menning KJ, Menon BB, Fox G, Scott UMLKM.** 2016. Dissolved inorganic carbon uptake in
596 *Thiomicrospira crunogena* XCL-2 is Δp- and ATP-sensitive and enhances RubisCO-mediated
597 carbon fixation. *Arch Microbiol* **198**:149-159.
- 598 16. **Dobrinski KP, Enkemann SA, Yoder SJ, Haller E, Scott KM.** 2012. Transcription response of the
599 sulfur chemolithoautotroph *Thiomicrospira crunogena* to dissolved inorganic carbon limitation. *J*
600 *Bacteriol* **194**:2074-2081.
- 601 17. **Mangiapia M, MCBL U, Brown T-RW, Chaput D, Haller E, Harmer TL, Hashemy Z, Keeley R,**
602 **Leonard J, Mancera P, Nicholson D, Stevens S, Wanjugi P, Zabinski T, Pan C, Scott KM.** 2017.
603 Proteomic and mutant analysis of the CO₂ concentrating mechanism of hydrothermal vent
604 chemolithoautotroph *Thiomicrospira crunogena*. *J Bacteriol* **199**:e00871-00816.

- 605 18. **Boden R, Scott KM, Williams J, Russel S, Antonen K, Rae AW, Hutt LP.** 2017. An evaluation of
606 *Thiomicrospira*, *Hydrogenovibrio* and *Thioalkalimicrobium*: Reclassification of four species of
607 *Thiomicrospira* to each *Thiomicrohabdus* gen. nov. and *Hydrogenovibrio*, and reclassification of
608 all four species of *Thioalkalimicrobium* to *Thiomicrospira*. *Int J Syst Evol Micr* **67**:1140-1151.
- 609 19. **Scott KM, Williams J, Porter CMB, Russel S, Harmer TL, Paul JH, Antonen KM, Bridges MK,**
610 **Camper GJ, Campa CK, Casella LG, Chase E, Conrad JW, Cruz MC, Dunlap DS, Duran L,**
611 **Fahsbender EM, Goldsmith DB, Keeley RF, Kondoff MR, Kussy BI, Lane MK, Lawler S, Leigh BA,**
612 **Lewis C, Lostal LM, Marking D, Mancera PA, McClenthan EC, McIntyre EA, Mine JA, Modi S,**
613 **Moore BD, Morgan WA, Nelson KM, Nguyen KN, Ogburn N, Parrino DG, Pedapudi AD, Pelham**
614 **RP, Preece AM, Rampersad EA, Richardson JC, Rodgers CM, Schaffer BL, Sheridan NE, Solone**
615 **MR, Staley ZR, Tabuchi M, Waide RJ, Wanjugi PW, Young S, Clum A, Daum C, Huntemann M,**
616 **Ivanova N, Kyrpides N, Mikhailova N, Palaniappan K, Pillay M, Reddy TBK, Shapiro N, Stamatis**
617 **D, Varghese N, Woyke T, Boden R, Freyermuth SK, Kerfeld CA.** 2018. Genomes of ubiquitous
618 marine and hypersaline *Hydrogenovibrio*, *Thiomicrohabdus*, and *Thiomicrospira* spp. encode a
619 diversity of mechanisms to sustain chemolithoautotrophy in heterogeneous environments.
620 *Environ Microbiol* **20**:2686 – 2708
- 621 20. **Nies DH, Koch S, Wachi S, Peitzsch N, Saier MH, Jr.** 1998. CHR, a novel family of prokaryotic
622 proton motive force-driven transporters probably containing chromate/sulfate antiporters. *J*
623 *Bacteriol* **180**:5799-5802.
- 624 21. **Kerfeld CA, Heinhorst S, Cannon GC.** 2010. Bacterial microcompartments. *Ann Rev Microbiol*
625 **64**:391-408.
- 626 22. **Merlin C, Masters M, McAteer S, Coulson A.** 2003. Why Is Carbonic Anhydrase Essential to
627 *Escherichia coli*? *J Bacteriol* **185**:6415-6424.
- 628 23. **Rosano GL, Ceccarelli EA.** 2014. Recombinant protein expression in *Escherichia coli*: advances
629 and challenges. *Front Microbiol* **5**:172.
- 630 24. **Zeebe RE, Wolf-Gladrow D.** 2003. CO₂ in seawater: Equilibrium, kinetics, isotopes. Elsevier, New
631 York.
- 632 25. **Knittel K, Kuever J, Meyerdierks A, Meinke R, Amann R, Brinkhoff T.** 2005. *Thiomicrospira*
633 *arctica* sp. nov. and *Thiomicrospira psychrophila* sp. nov., psychrophilic, obligately
634 chemolithoautotrophic, sulfur-oxidizing bacteria isolated from marine Arctic sediments. *Int J*
635 *Syst Evol Microbiol* **55**:781-786.
- 636 26. **Brinkhoff T, Sievert SM, Kuever J, Muyzer G.** 1999. Distribution and diversity of sulfur-oxidizing
637 *Thiomicrospira* spp. at a shallow-water hydrothermal vent in the Aegean Sea (Milos, Greece).
638 *Appl Environ Microb* **65**:3843-3849.
- 639 27. **Axen SD, Erbilgin O, Kerfeld CA.** 2014. A taxonomy of bacterial microcompartment loci
640 constructed by a novel scoring method. *PLoS Comput Biol* **10**: e1003898.
- 641 28. **Du J, Förster B, Rourke L, Howitt SM, Price GD.** 2014. Characterisation of cyanobacterial
642 bicarbonate transporters in *E. coli* shows that SbtA homologs are functional in this heterologous
643 expression system. *PLoS ONE* **9**:e115905.
- 644 29. **Babu M, Greenblatt JF, Emili A, Strynadka NCJ, Reithmeier RAF, Moraes TF.** 2010. Structure of
645 a SLC26 anion transporter STAS domain in complex with acyl carrier protein: Implications for *E.*
646 *coli* YchM in fatty acid metabolism. *Structure* **18**:1450-1462.
- 647 30. **Price GD, Howitt SM.** 2011. The cyanobacterial bicarbonate transporter BicA: its physiological
648 role and the implications of structural similarities with human SLC26 transporters. *Biochemistry*
649 and cell biology = *Biochimie et biologie cellulaire* **89**:178-188.
- 650 31. **Aravind L, Koonin EV.** 2000. The STAS domain - a link between anion transporters and
651 antisigma-factor antagonists. *Current biology : CB* **10**:R53-55.

- 652 32. **Shelden MC, Howitt SM, Price GD.** 2010. Membrane topology of the cyanobacterial bicarbonate
653 transporter, BicA, a member of the SulP (SLC26A) family. *Mol Membr Biol* **27**:12-22.
- 654 33. **Price GD, Badger MR, Woodger FJ, Long BM.** 2009. Advances in understanding the
655 cyanobacterial CO₂-concentrating-mechanism (CCM): functional components, Ci transporters,
656 diversity, genetic regulation and prospects for engineering into plants. *J Exp Bot* **59**:1441-1461.
- 657 34. **Le Bris N, Govenar B, Le Gall C, Fisher CR.** 2006. Variability of physico-chemical conditions in 9
658 degrees 50 ' N EPR diffuse flow vent habitats. *Mar. Chem.* **98**:167-182.
- 659 35. **Markowitz VM, Chen IA, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J, Woyke T,
660 Huntemann M, Anderson I, Billis K, Varghese N, Mavromatis K, Pati A, Ivanova NN, Kyrpides
661 NC.** 2014. IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucl
662 Acids Res* **42**:D560-D567.
- 663 36. **Edgar RC.** 2004. MUSCLE: Multiple sequence alignment with high accuracy and high
664 throughput. *Nucl Acids Res* **32**:1792-1797.
- 665 37. **Crooks GE, Hon G, Chandonia JM, Brenner SE.** 2004. WebLogo: A sequence logo generator.
666 *Genome Res* **14**:1188-1190.
- 667 38. **Talavera G, Castresana J.** 2007. Improvement of phylogenies after removing divergent and
668 ambiguously aligned blocks from protein sequence alignments. *Systematic biology* **56**:564-577.
- 669 39. **Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O.** 2010. New algorithms
670 and methods to estimate maximum-likelihood phylogenies: assessing the performance of
671 PhyML 3.0. *Systematic biology* **59**:307-321.
- 672 40. **Lefort V, Longueville JE, Gascuel O.** 2017. SMS: Smart model selection in PhyML. *Mol Bio Evol*
673 **34**:2422-2424.
- 674 41. **Le SQ, Gascuel O.** 2008. An improved general amino acid replacement matrix. *Mol Biol Evol*
675 **25**:1307-1320.
- 676 42. **Rambaut A.** 2016. FigTree: Tree Figure Drawing Tool Version 1.4.3, CoInstitute of Evolutionary
677 Biology, University of Edinburgh.
- 678 43. **Sorokin DY, Tourova TP, Kolganova TV, Spiridonova EM, Berg IA, Muyzer G.** 2006.
679 *Thiomicrospira halphila* sp. nov., a moderately halophilic, obligately chemolithoautotrophic,
680 sulfur-oxidizing bacterium from hypersaline lakes. *Int J Syst Evol Microbiol* **56**:2375-2380.
- 681 44. **Parkhurst DL, Appelo CAJ.** 1999. User's guide to PHREEQC (Version 2) : A computer program for
682 speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations.
683 Report 99-4259.
- 684 45. **Wolery T.** 1979. Calculation of chemical equilibrium between aqueous solution and minerals:
685 the EQ3/6 software package. [In FORTRAN extended 4. 6 for CDC6600 and 7600].
- 686 46. **Nordstrom DK, Plummer LN, Wigley TML, Wolery TJ, Ball JW, Jenne EA, Bassett RL, Crerar DA,
687 Florence TM, Fritz B, Hoffman M, Holdren GR, Lafon GM, Mattigod SV, McDuff RE, Morel F,
688 Reddy MM, Sposito G, Thraillkill J.** 1979. A comparison of computerized chemical models for
689 equilibrium calculations in aqueous systems, p. 857-892, *Chemical Modeling in Aqueous
690 Systems*, vol. 93. American Chemical Society.
- 691 47. **Livak KJ, Schmittgen TD.** 2001. Analysis of relative gene expression data using real-time
692 quantitative PCR and the 2^{-ΔΔC_T} method. *Methods* **25**:402-408.
- 693 48. **Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, Högbom M, van Wijk KJ,
694 Slotboom DJ, Persson JO, de Gier J-W.** 2008. Tuning *Escherichia coli* for membrane protein
695 overexpression. *Proc Nat Acad Sci USA* **105**:14371-14376.
- 696 49. **Schlegel S, Löfblom J, Lee C, Hjelm A, Klepsch M, Strous M, Drew D, Slotboom DJ, de Gier J-W.**
697 2012. Optimizing membrane protein overexpression in the *Escherichia coli* strain Lemo21(DE3). *J
698 Mol Biol* **423**:648-659.

- 699 50. **Sagné C, Isambert MF, Henry JP, Gasnier B.** 1996. SDS-resistant aggregation of membrane
700 proteins: application to the purification of the vesicular monoamine transporter. *Biochem J*
701 **316**:825-831.
- 702 51. **Sambrook J, Russell DW.** 2001. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor
703 Laboratory Press, Cold Spring Harbor.
- 704 52. **Oliver DB.** 1996. Periplasm. *In* Neidhardt FC (ed.), *Escherichia coli* and *Salmonella*: Cellular and
705 Molecular Biology, vol. 1. ASM Press, Washington, DC.
- 706 53. **Davanloo P, Rosenberg AH, Dunn JJ, Studier FW.** 1984. Cloning and expression of the gene for
707 bacteriophage T7 RNA polymerase. *Proc Natl Acad Sci U S A* **81**:2035-2039.
- 708 54. **Cervantes C, Ohtake H, Chu L, Misra TK, Silver S.** 1990. Cloning, nucleotide sequence, and
709 expression of the chromate resistance determinant of *Pseudomonas aeruginosa* plasmid
710 pUM505. *J Bacteriol* **172**:287-291.
- 711 55. **Zolotarev AS, Unnikrishnan M, Shmukler BE, Clark JS, Vandorpe DH, Grigorieff N, Rubin EJ,**
712 **Alper SL.** 2008. Increased sulfate uptake by *E. coli* overexpressing the SLC26-related SulP protein
713 Rv1739c from *Mycobacterium tuberculosis*. *Comparative biochemistry and physiology. Part A,*
714 *Molecular & integrative physiology* **149**:255-266.
- 715 56. **Weissgerber TL, Milic NM, Winham SJ, Garovic VD.** 2015. Beyond bar and line graphs: time for
716 a new data presentation paradigm. *PLoS Biol* **13**:e1002128.
- 717 57. **Goffredi SK, Childress JJ, Desaulniers NT, Lee RW, Lallier FH, Hammond D.** 1997. Inorganic
718 carbon acquisition by the hydrothermal vent tubeworm *Riftia pachyptila* depends upon high
719 external P-CO₂ and upon proton-equivalent ion transport by the worm. *J Exp Biol* **200**:883-896.
- 720 58. **Ahmad A, Barry JP, Nelson DC.** 1999. Phylogenetic affinity of a wide, vacuolate, nitrate-
721 accumulating *Beggiatoa* sp. from Monterey Canyon, California, with *Thioploca* spp. *Appl Environ*
722 *Microbiol* **65**:270-277.

723

724

725 **FIGURE LEGENDS**

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

734

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

751

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

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

767

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

772

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

776

777 **FIG 6** Intracellular DIC accumulation by *E coli* expressing potential DIC transporter
778 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 ($\alpha < 0.05$). 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 CO₂
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

801 support. On the left is a possible ancestral carboxysome locus. On the right are the
802 carboxysome loci in these organisms. Numbers in parentheses indicate the number of
803 scaffolds of the draft genome sequence; (C) indicates that the genome sequence is
804 complete. + *gene* = phylogenetic analysis and gene taxonomic distribution suggest
805 acquisition via horizontal gene transfer; - *gene* = gene is absent from sequenced
806 genome; CS = gene is collocated with carboxysome locus; to CS = gene moved from
807 elsewhere in the chromosome, to collocation with carboxysome locus; CCM = CO₂
808 concentrating mechanism.

809

810

811

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

| | Arctic marine sediments ^a (<i>Tmr. arctica</i>) | Hydrothermal vents ^a (<i>H. crunogenus</i>) |
|---------------------------------|---|---|
| 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 ^aTemperatures, 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 ^bCompounds 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*, *Thiomicrothrix*, 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) |
|---------------------------|--------------------|--|---|--|-------------------------------------|
| <i>Hydrogenovibrio</i> | | | | | |
| <i>crunogenus</i> XCL-2 | | | | | |
| | <i>csoS3</i> | | 9.7 ± 1.9 ^e | | 823 |
| | 853-I (CS) | | 8.0 ± 1.8 ^e | | 263 |
| | 854-I (CS) | | 8.4 ± 1.6 ^e | | 340 |
| | <i>chr-I</i> | - ^f | 0.4 ± 0.7 | | 1.3 |
| | <i>sulP-II</i> | <0.05 | 0.8 ± 0.2 | N/A ^g | 1.8 |
| <i>Hydrogenovibrio</i> | | | | | |
| <i>thermophilus</i> JR2 | | | | | |
| | <i>csoS3</i> | <0.005 | 11.0 ± 0.2 | <0.001 | 1984 |
| | <i>chr-I</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 |
| | <i>sulP-I</i> | <0.005 | 6.9 ± 0.1 | <0.001 | 123 |
| | <i>sulP-II</i> | <0.01 | -1.1 ± 0.1 | N/A | 0.5 |
| <i>Hydrogenovibrio</i> | | | | | |
| <i>thermophilus</i> MA2-6 | | | | | |
| | <i>csoS3</i> | <0.005 | 12.5 ± 0.4 | <0.001 | 5746 |
| | <i>chr-I</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 |
| | <i>sulP-II</i> | - | -0.1 ± 0.5 | N/A | 0.9 |
| <i>Hydrogenovibrio</i> | | | | | |
| <i>halophilus</i> | | | | | |
| | <i>csoS3</i> | <0.005 | 11.3 ± 0.3 | <0.001 | 2517 |
| | <i>sbtA-I</i> (CS) | <0.005 | 9.4 ± 0.2 | <0.001 | 680 |
| | <i>chr-I</i> | <0.05 | -0.5 ± 0.2 | N/A | 0.7 |
| | <i>sulP-III</i> | <0.05 | -0.8 ± 0.2 | N/A | 0.6 |

| | | | | | |
|-------------------------------------|-----------------------------|--------|-----------------|--------|------|
| <i>Hydrogenovibrio marinus</i> | <i>csoS3</i> | <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 |
| | <i>chr-II</i> | - | -0.9 ± 0.5 | N/A | 0.5 |
| <i>Hydrogenovibrio</i> sp. Milos-T1 | <i>csoS3</i> | <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 |
| | <i>hyp</i> (853) | <0.005 | 4.1 ± 0.2 | <0.005 | 17 |
| | 854-II | <0.005 | 4.1 ± 0.3 | <0.005 | 17 |
| | <i>chr-II</i> | - | -0.3 ± 0.2 | N/A | 0.8 |
| <i>Thiomicrothabodus frisia</i> Kp2 | <i>csoS3</i> | <0.005 | 12.4 ± 0.3 | <0.001 | 5327 |
| | <i>sbtA-I</i> (CS) | <0.05 | 9.2 ± 0.6 | <0.025 | 597 |
| | <i>sulP-II</i> | <0.05 | 1.4 ± 0.4 | - | 2.7 |
| <i>Thiomicrospira pelophila</i> | <i>csoS2</i> | <0.025 | 1.8 ± 0.3 | <0.05 | 3 |
| | <i>hyp</i> (<i>csoS2</i>) | N/A | ND ^h | N/A | |
| | <i>sulP-I</i> (CS) | <0.025 | 1.6 ± 0.4 | - | 3 |
| | <i>sbtA-I</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 |
| | <i>chr-I</i> | - | -0.2 ± 0.3 | N/A | 1 |
| | <i>sbtA-II</i> | <0.05 | 10.4 ± 2.7 | <0.025 | 1391 |
| | <i>sulP-III</i> | - | -0.1 ± 0.5 | N/A | 0.9 |

822

823 ^aGene abbreviations: *chr* = chromate ion transporter family; *sulP* = sulfate transporter family; 853, 854 = homologs to824 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 ^bFor all species, citrate synthase was used as the calibrator gene.

828 ^ctwo-tailed *t*-test, n=3 for all except *Tmr. frisia* Kp2 *sbtA* (CS), for which n=2

829 ^done-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 ^gN/A: Not applicable

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

835

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

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

838
 839 ^aSamples consisted of membranes prepared from *E. coli* cells expressing potential DIC
 840 transporters. Sample 8534 is from *E. coli* expressing both Tcr_0853 and Tcr_0854
 841 (8534), while samples 853 and 854 were from *E. coli* expressing either Tcr_0853 or
 842 Tcr_0854.

843

844

845

846

847

848

849

850

851

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

| Taxon | IMG gene object ID^a | Predicted gene product^b | Forward primer | Reverse primer |
|---------------------------------|---------------------------------------|---|-------------------------|---------------------------|
| <i>H. crunogenus</i> XCL-2 | 637785059 | Citrate synthase | CTTTGATGCGGGCTTGTTTAC | CCCCTGTGTAGATTTGAGTCG |
| | 637785561 | CsoSCA | CTCCGCTTACCTTATGCCTTAG | AGTAACGTGTTGGTTCATCCG |
| | 637786436 | Chr-I | GGTTTCGGCCTGGACTATTT | GCGCTTCATCAAACCAAGAC |
| | 637786269 | SulP-II | CGGATTGATTACCGCCATCT | TGCCATGCTCCATCACTAAA |
| <i>H. thermophilus</i> JR2 | 2507072380 | Citrate synthase | CGAATCCGTGCTCGGTTATT | GAACCGATTTTCATCCAGCATTTTC |
| | 2507073759 | CsoSCA | GCGTTCCAGGCTCTAAAGATAG | GGATGCCGACAATTCCTGATA |
| | 2507073746 | Chr-I (CS) | GCTGGAGCTTGATCGTGTTA | CCATCTCCGATCGACCAAATAC |
| | 2507074342 | 853-I | CCTGTTTATGGCCGGTTACA | GTCACCCATTTCGTCCAGATAAA |
| | 2507074343 | 854-I | GCTTCGCCTCAGTGTCTATATC | GAGTCCCAACGGAAACAGAA |
| | 2507074344 | SulP-I | TGTGTGGCTGTGGCTTTAT | TTGGAGTTACAGGGTCGTTTC |
| | 2507073582 | SulP-II | ACACCTTGTCGGGCATTAC | GAGGTGATGAAACCGACGATAA |
| <i>H. thermophilus</i> MA2-6 | 2572250326 | Citrate synthase | CGAATCCGTGCTCGGTTATT | GAACCGATTTTCATCCAGCATTTTC |
| | 2572249265 | CsoSCA | GAAATCGGAAGACAGGACAGAG | CCATTTTCATAACGACGCAACAA |
| | 2572249252 | Chr-I (CS) | TGGCGCTGAATCTGGTATTG | CGCCACCCAACTCCAATAAA |
| | 2572249866 | 853-I | CCTACATGGCCGGAATAAG | ATCCAAGCGGTCATCATCAG |
| | 2572249867 | 854-I | ATGTGCGCTCGGAAATCA | GGGCGGTATTCTATCGGTAATC |
| | 2572249063 | SulP-II | ACACCTTGTCGGGCATAAC | GACGATAATCGCGGCATACA |
| <i>H. halophilus</i> | 2518266203 | Citrate synthase | CCAGACGGGTCAAATACAATCT | ATGTATTCCGGTGCTGGGATAAA |
| | 2518265324 | CsoSCA | AGGGTCTGTACCCGGATATT | CGTGTCCAGAAACCCGTAAT |
| | 2518265315 | SbtA-I (CS) | ATTGGCCACGTCGGATTT | GAATCGCAACAGCGCATAAC |
| | 2518266741 | Chr-I | CTGGCTGGGACAAACCTATT | CCCAGAGCCGAACCTCATTT |
| | 2518265727 | SulP-III | TGGCCGGGTATCTGAATTTG | CGGTAGCTGAGCCATGAATATC |
| <i>H. marinus</i> | 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 |
| <i>Hydrogenovibrio</i> sp. Milos-T1 | 2579718736 | Citrate synthase | GACTGGTGAAGAGCCAGATAAG | CCCAGTCACCATAGTGGTAAAG |
| | 2579719002 | CsoSCA | GATGTCAGCGAGAGTGTTAGAG | TCTGCTTTTCGAGAAGTGGTAAA |
| | 2579719013 | 853-I (CS) | GGGATTTGTGGATGGCATTTC | ACACTCTTCTGGTCTTCATCATC |
| | 2579719014 | 854-I (CS) | GGGATGTATAGCGAGTGGTTAG | GGTATCTGGCAAGCTGAGAA |
| | 2579719194 | 853-II | GTTTGGCGAGCAGCTTTATG | CCCGTAGCCAGCCAATAATAA |
| | 2579719193 | Hyp(853) | TCCTAACTGGGTGATGATTTGG | CATGCAGACGTGCAATAAAAG |
| | 2579719192 | 854-II | ACTACCAACCCAAGCCTAAAG | GTAGTCTCCCATGCCTTCTAAAT |
| | 2579720167 | Chr-II | CATCAGGAGCTGGTGGATAAG | ATATAGGTGGCGAGCTGTTG |
| <i>Tmr. frisia</i> Kp2 | 2517375157 | Citrate synthase | ACCCTTGTTCCGTTATCTCTTC | CAGGCGAACCAATCTCATCT |
| | 2517375722 | CsoSCA | GCGTACGTAAGTGGTCTTTAT | TTGGTGGTGTGGATCTGATTT |
| | 2517375731 | SbtA-I (CS) | GCACACGAAAGCTACCCTATTA | CTGAGAACCATCACCTGAAGTC |
| | 2517376429 | SulP-II | ATGACACCCTATCAGGCATTAC | CAGGACGACCACCAATACA |
| <i>Tms. pelophila</i> | 2568511528 | Citrate synthase | GATCCAATTGAACCGCGTAAAG | CATCAAGTAAACGTGCCCAATC |
| | 2568509998 | CsoS2 | GCTAATGCTTACTCTGCACCTA | CGGCTCCATCTCCTGTTATTC |
| | 2568509999 | Hyp(csoS2) | TAGGTTTCGCCAAGCAGTAAG | GCATCAGTCAAGGCATACAAAC |
| | 2568510006 | SulP-I (CS) | GAAGCGCCTAAACAAGACAAAG | TAGCCATTGCGGCTGTAATAA |
| | 2568510007 | SbtA-I (CS) | GCGGCTAGTGCATCCTATATT | ACGGAAAGGTAAGTCCAAGTG |
| | 2568511025 | 853-I | CGGGTAATGTTGAGGAAGAGAA | AACCCACGCCAGCATAAA |
| | 2568511024 | 854-I | GGAGTCCAAGTTCAGCCGATTAC | AGCCCGTCTTGCCAATTTA |
| | 2568510063 | Chr-I | CGCCGCCTTAATAAATCCAATC | AACAGGTTTCAGACCCAGTAATC |
| | 2568511393 | SbtA-II | GCGGCTACCTATGGTTCAATTA | ACCATTGCCACCGTCATATAG |
| | 2568511001 | SulP-III | GTCAGGGCGTGGCAAATA | CCTGCCGTAAAGGTGGATAAA |

853

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

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

861

862

863 **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 |
|-------------------------------|---------------------------------------|---|--|--|
| <i>H. crunogenus</i> XCL-2 | 637785573 | 853 (CS) | CACCATGAATATGCAATGGGTAGG | TCATTGAAATAACTCCTCTTTAGGAACTT |
| | 637785574 | 854 (CS) 853 & 854 | CACCATGATGTTGCACAACGC CACCATGAATATGCAATGGGTAGG | TCAGGCAGATTCCAACCACT TCAGGCAGATTCCAACCACT |
| <i>H. thermophilus</i> JR2 | 2507073746 | chr (CS) | CACCATGTCATTGCCTGTCTTTTTG | TTAGCCCGAAACAAACGACACC |
| | | chr (reverse) | ATGTCATTGCTTGTCTTTTTGGC | CACCTTAGCCCGAAACAAACG |
| | 2507074344 | SulP SulP (reverse) | CACCATGACACAGGAAAACATAAAC ATGACACAGGAAAACATAAACACAG | TCAATTTAATTCTTTATCGTCTTCTTTAAATTTTTTG CACCTCAATTTAATTCTTTATCGTCTTCTTTAA |
| <i>Tmr. frisia</i> Kp2 | 2517375731 | SbtA (CS) | CACCATGTTGGGATTGGATAGC | TTATACCGCTGAATACCACATAGC |

864

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

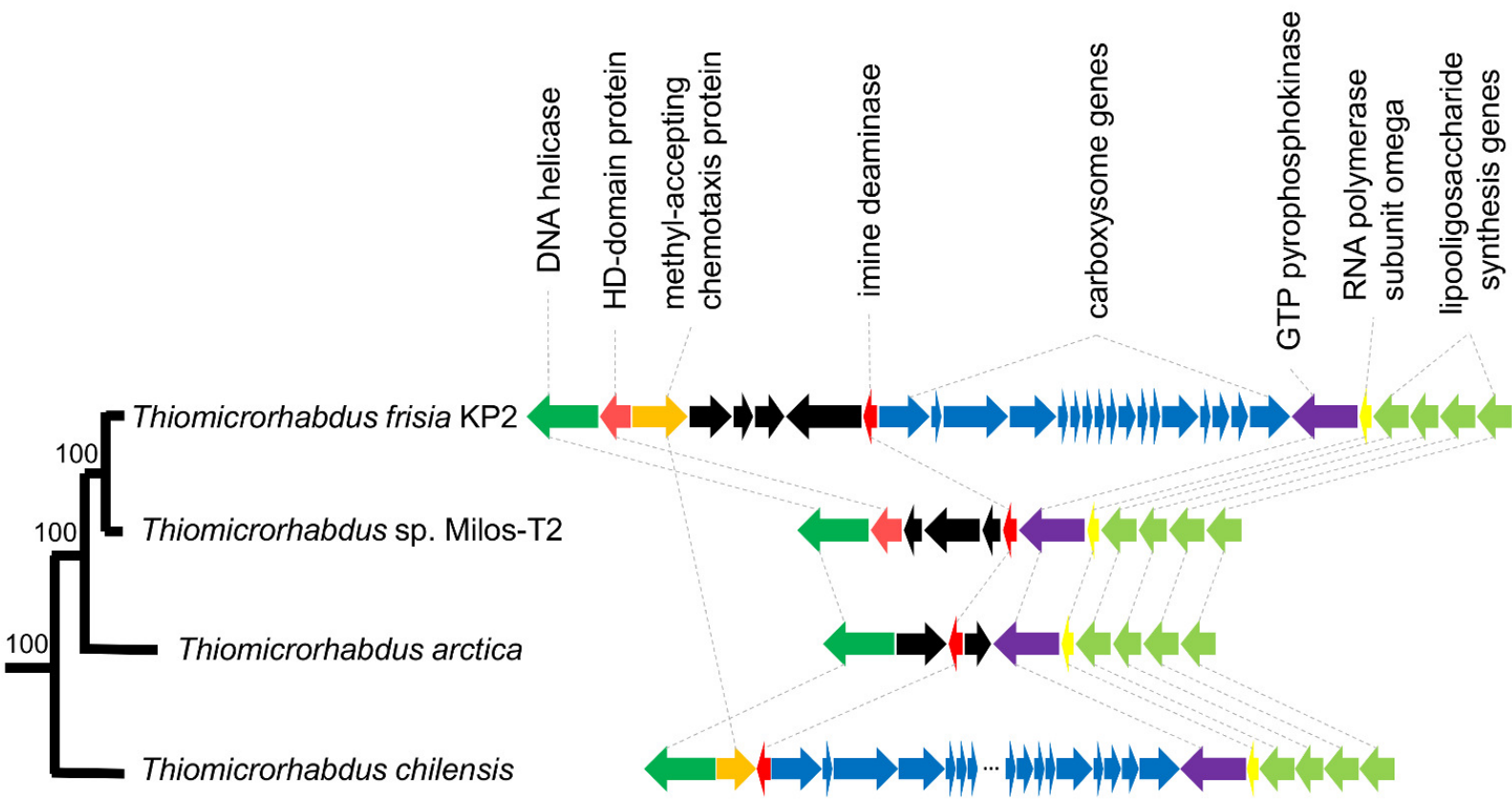
867 ^bGene product abbreviations: Chr = chromate ion transporter family; SulP = sulfate permease family; 853, 854 = proteins
868 encoded by Tcr_0853, 0854; SbtA = sodium-dependent bicarbonate transporter family. Gene product names followed by
869 (CS) indicate that these genes are adjacent to those encoding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

870

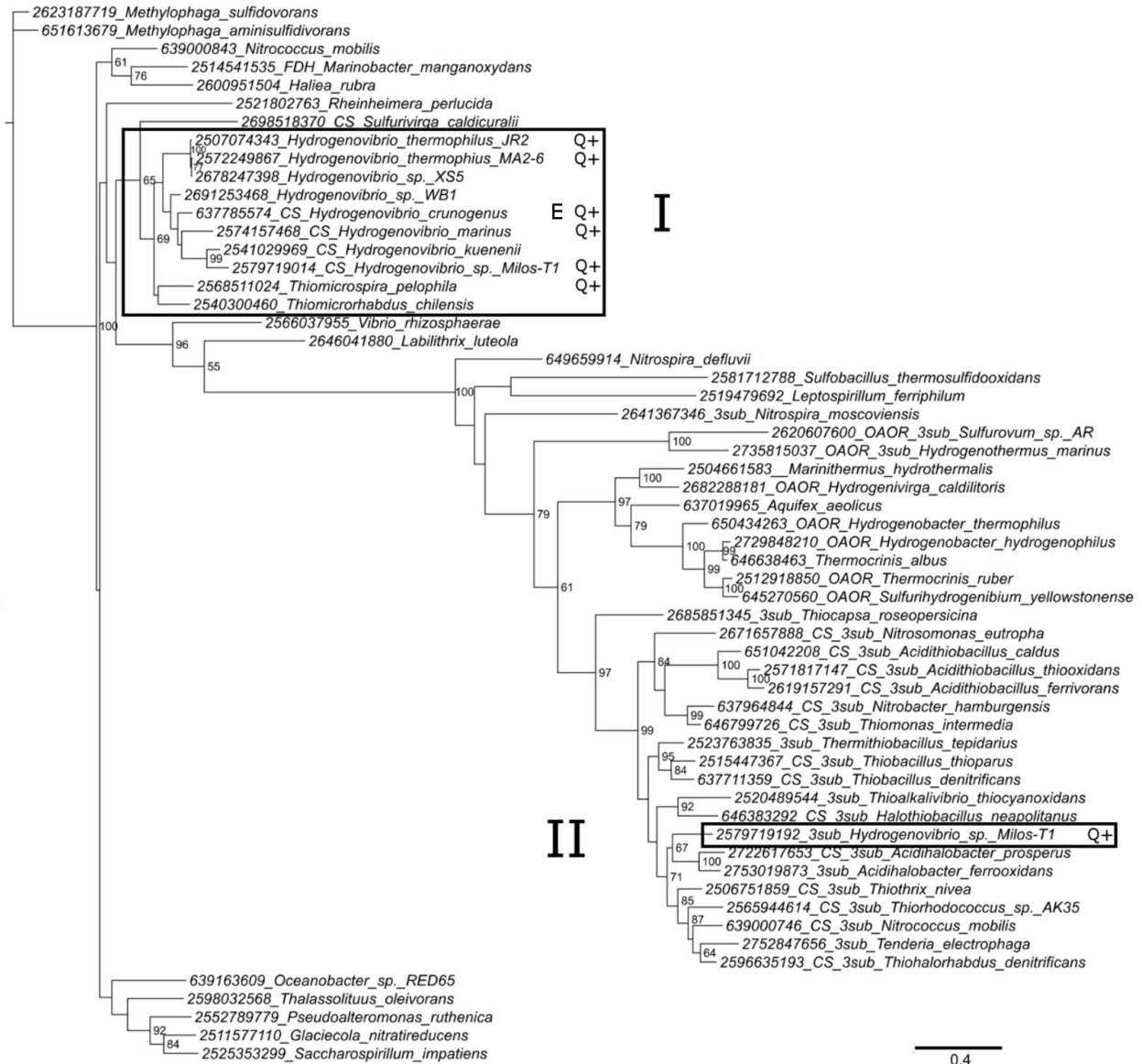
871

872

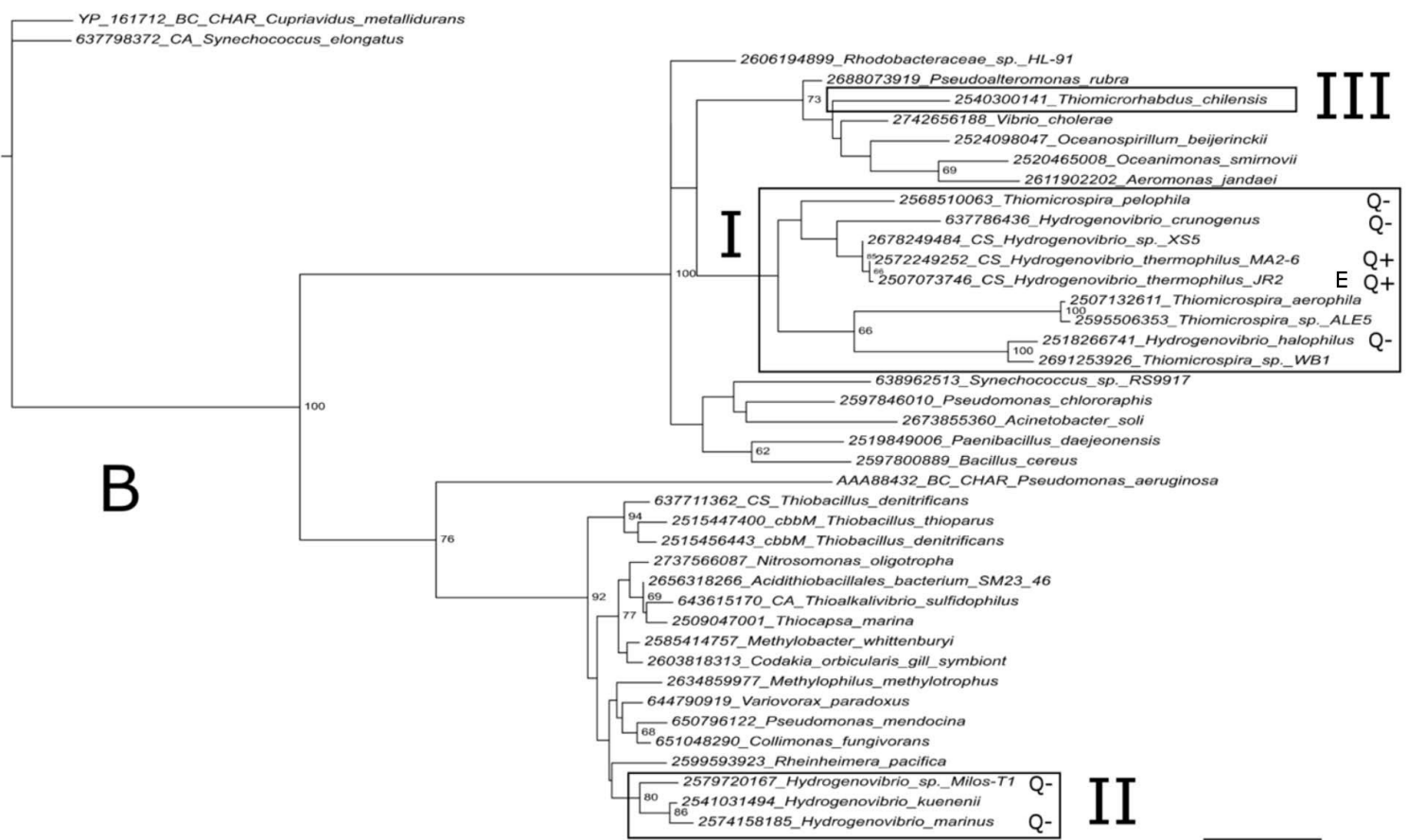
873

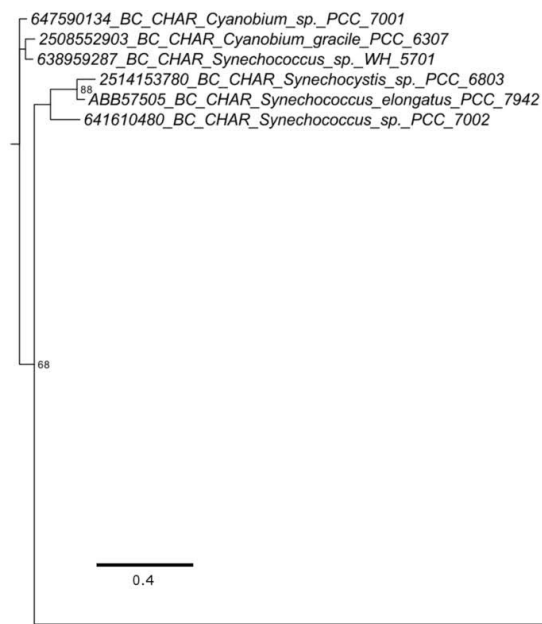


A

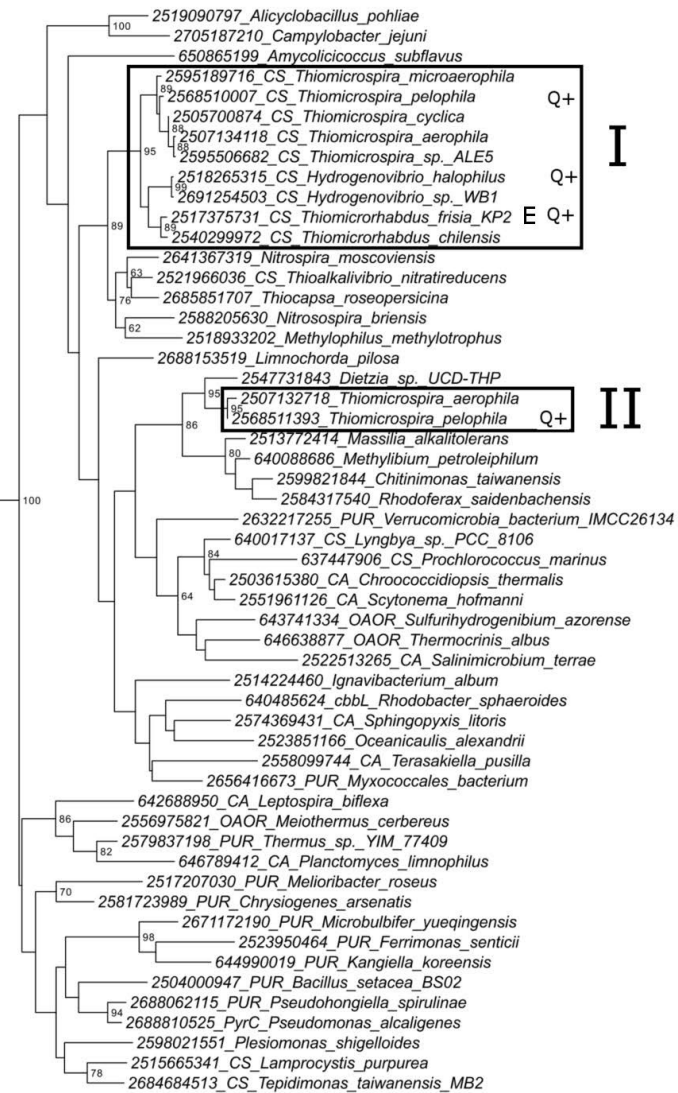


B



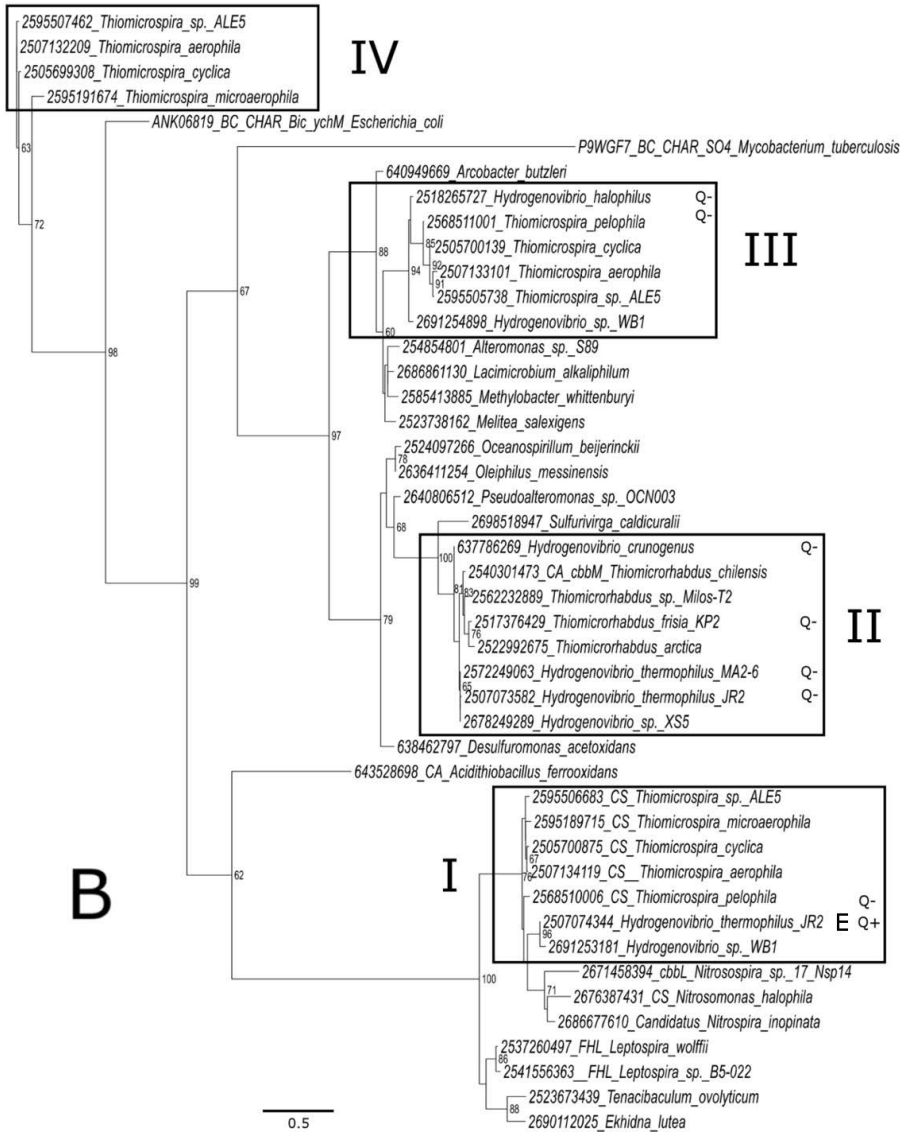


A



I

II



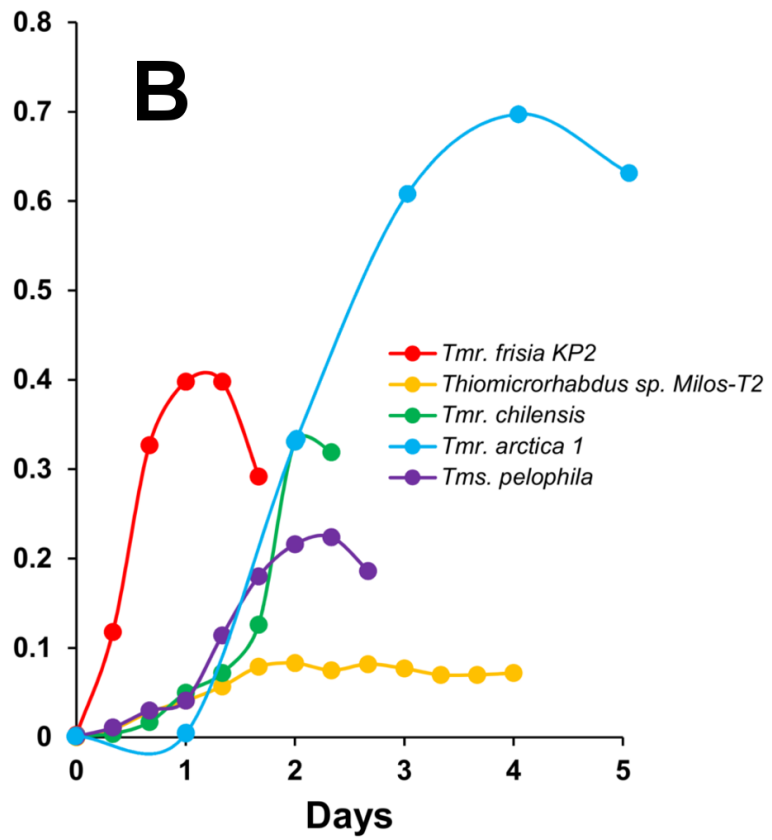
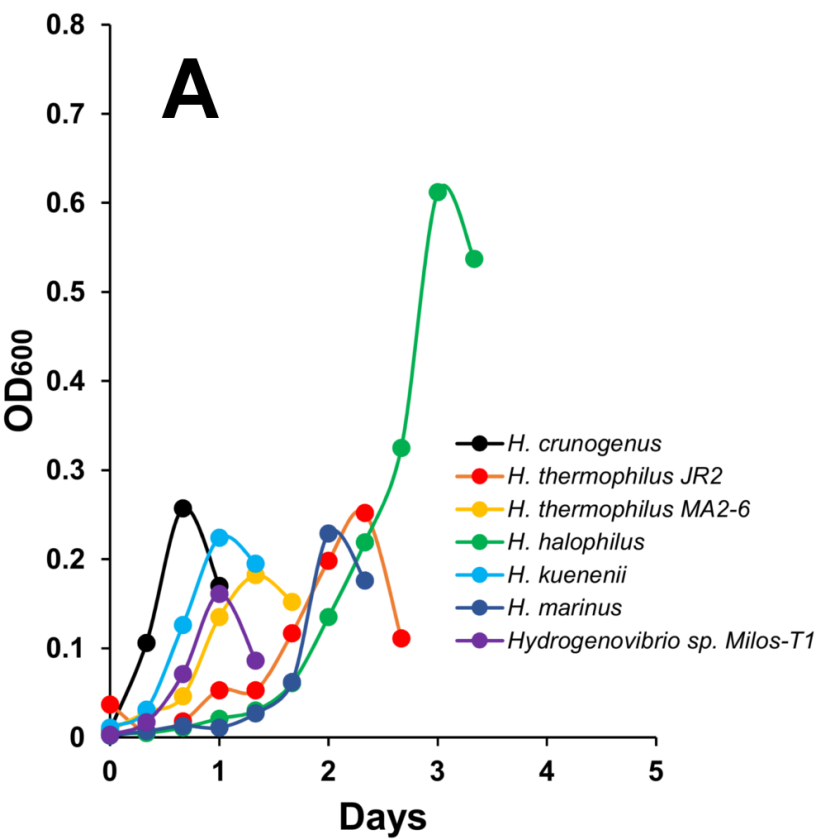
IV

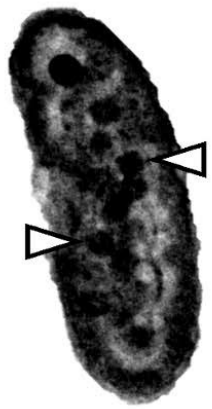
III

II

I

B

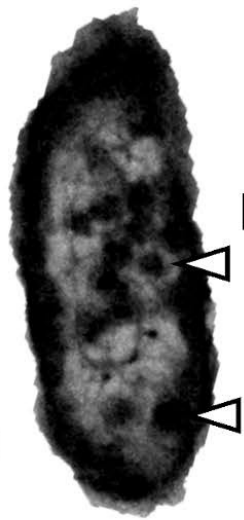




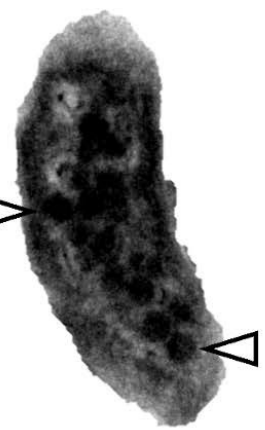
H. crunogenus



H. thermophilus
JR2



H. thermophilus
MA2-6



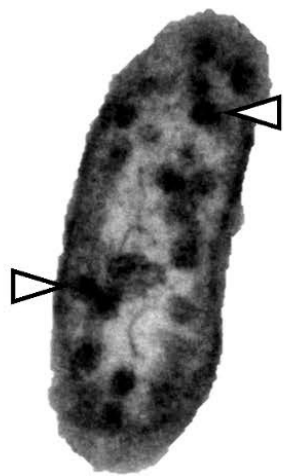
H. kuenenii



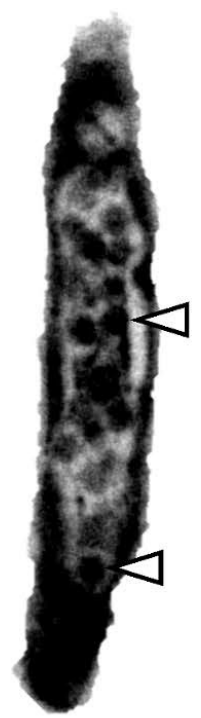
H. halophilus



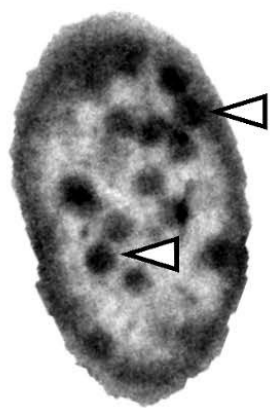
Thiomicroorhabdus sp.
Milos T2



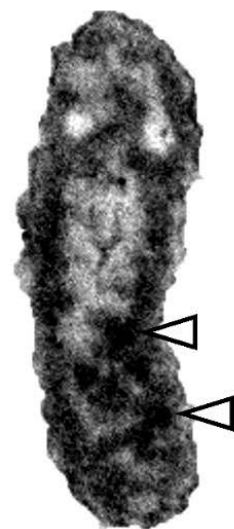
H. marinus



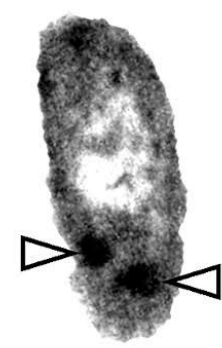
Hydrogenovibrio sp.
Milos T1



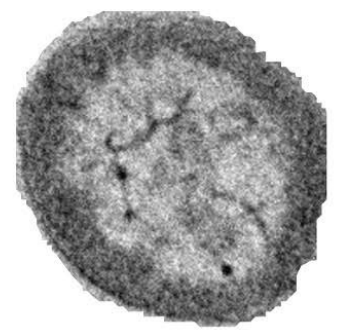
Tmr. frisia Kp2



Tmr. chilensis



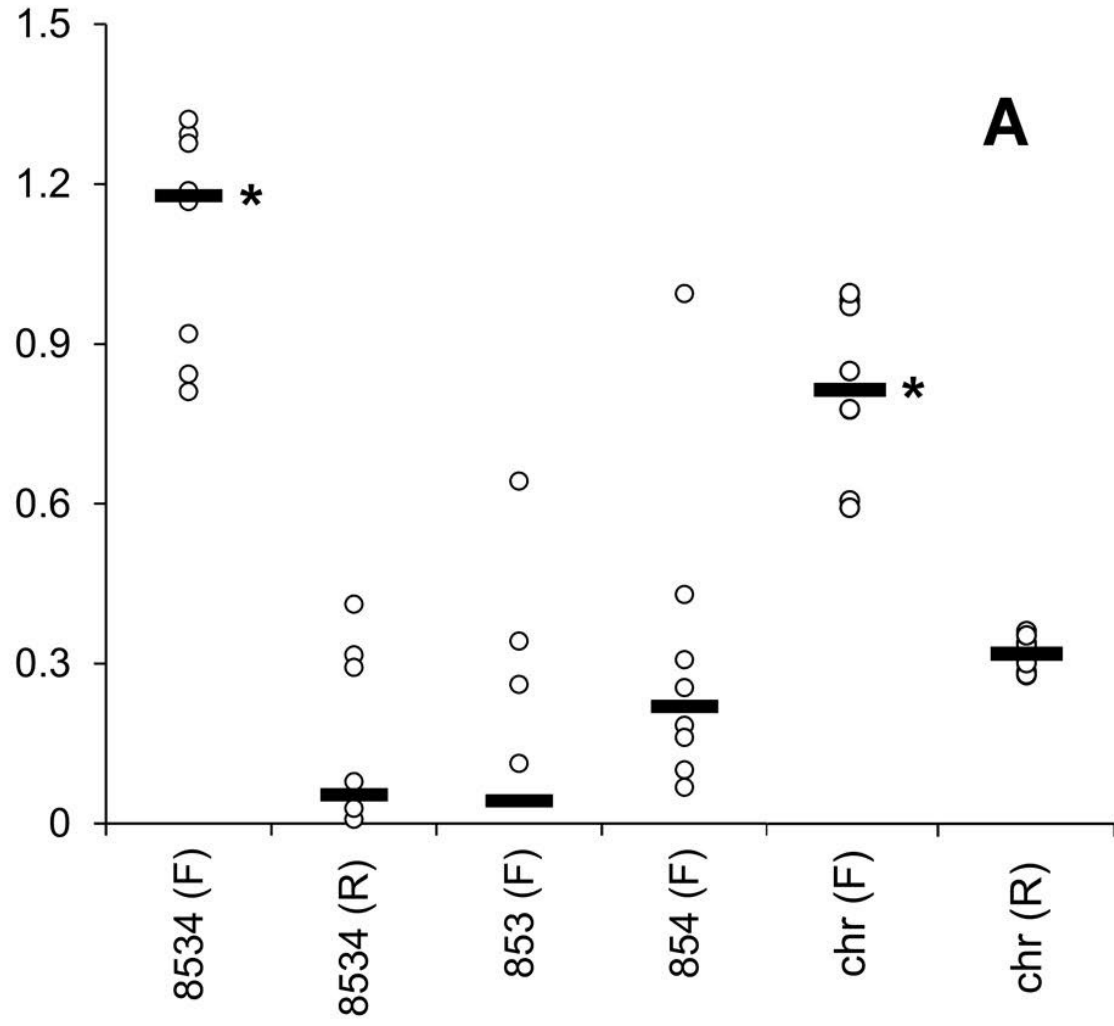
Tms. pelophila



Tmr. arctica



Intracellular DIC (mM)



Intracellular DIC (mM)

