

1 **Genome-wide screening reveals a novel class of carbonic anhydrase-like inorganic carbon**  
2 **pumps in chemoautotrophic bacteria**

3  
4 John J. Desmarais<sup>1</sup>, Avi I. Flamholz<sup>1</sup>, Cecilia Blikstad<sup>1</sup>, Eli J. Dugan<sup>1</sup>, Thomas G. Laughlin<sup>1</sup>, Luke M.  
5 Oltrogge<sup>1</sup>, Allen W. Chen<sup>3</sup>, Kelly Wetmore<sup>2</sup>, Spencer Diamond<sup>3</sup>, Joy Y. Wang<sup>4</sup>, David F. Savage<sup>1</sup>

6  
7 <sup>1</sup>Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720

8 <sup>2</sup>Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley,  
9 CA, USA

10 <sup>3</sup>Department of Earth and Planetary Science, University of California, Berkeley, CA 94720

11 <sup>4</sup>Department of Chemistry, University of California, Berkeley, CA 94720

12  
13  
14 Correspondence should be addressed to: savage@berkeley.edu.

15    **Abstract**

16    Many bacterial autotrophs rely on CO<sub>2</sub> concentrating mechanisms (CCMs) to assimilate carbon. Although  
17    many CCM proteins have been identified, including a 200+ MDa protein organelle called the carboxysome,  
18    a systematic screen of CCM components has not been carried out. Here, we performed a genome-wide  
19    barcoded transposon screen to identify essential and CCM-related genes in the γ-proteobacterium *H.*  
20    *neapolitanus*. Our screen revealed an operon critical for CCM function which encodes a domain of  
21    unknown function (PFAM:PF10070) and putative cation transporter subunit (PFAM:PF00361). These  
22    two proteins, which we name DabA and DabB for “DABs accumulate bicarbonate,” function as a  
23    heterodimeric, energy-coupled inorganic carbon pump in *E. coli*. Furthermore, DabA binds zinc and has  
24    a an active site homologous to a β-carbonic anhydrase. Based on these results, we propose that DABs  
25    function as vectorial CAs coupled to cation gradients and serve as inorganic carbon pumps throughout  
26    prokaryotic phyla.

27 **Introduction**

28 Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (Rubisco) is the primary carboxylase of the  
29 Calvin-Benson-Bassham (CBB) cycle and the major entry point of inorganic carbon ( $C_i$ ) into the biosphere.  
30 Rubisco activity is thus critical to agriculture and a major flux removing anthropogenic  $CO_2$  from the  
31 atmosphere. Despite its centrality and abundance, Rubisco is not a very fast enzyme (Bar-Even et al., 2011;  
32 Bathellier et al., 2018; Flamholz et al., 2018). Nor is Rubisco very specific - all known Rubiscos can use  
33 molecular oxygen ( $O_2$ ) as a substrate in place of  $CO_2$  (Tcherkez, 2016). The resulting oxygenation reaction  
34 is often described as “wasteful” as it fails to incorporate inorganic carbon and produces a product, 2-  
35 phosphoglycolate, that is not part of the CBB cycle and must be recycled through metabolically-expensive  
36 photorespiratory pathways (Bauwe et al., 2010; Buchanan et al., 2015). Many studies support the  
37 hypothesis that improvements to Rubisco could improve crop yields, but Rubisco has proven quite  
38 recalcitrant to improvement by engineering. Indeed, it remains unclear whether or how Rubisco can be  
39 improved (Flamholz et al., 2018; Savir et al., 2010; Tcherkez et al., 2006).

40 Organisms that depend on Rubisco for growth often employ supplemental physiological  
41 mechanisms to improve its rate and specificity. These mechanisms are collectively termed  $CO_2$   
42 concentrating mechanisms (CCMs) because they serve to concentrate  $CO_2$  at the site of Rubisco, ensuring  
43 Rubisco is saturated with  $CO_2$ , so that carboxylation proceeds at its maximum rate and oxygenation is  
44 competitively inhibited (Buchanan et al., 2015; Mangan et al., 2016; Raven et al., 2017). All cyanobacteria  
45 and many chemotrophic proteobacteria have a CCM (Badger and Price, 2003; Cannon et al., 2001). The  
46 bacterial CCM has garnered particular interest among bioengineers because it is well-understood,  
47 composed of only ~20 genes and operates inside single cells (Long et al., 2016). Detailed modeling  
48 suggests that transplantation of the bacterial CCM into crops might improve yields (McGrath and Long,  
49 2014; Price et al., 2011) and efforts towards transplantation are already underway (Lin et al., 2014; Long et  
50 al., 2018; Occhialini et al., 2016).

51 Based on diverse experimental studies, a general model of the bacterial CCM function has emerged.  
52 This model requires two major components: active transport of  $C_i$  leading to the accumulation of  $HCO_3^-$  in  
53 the cytosol and organization of RuBisCO with carbonic anhydrase (CA) in the lumen of a 200+ MDa protein  
54 organelle known as the carboxysome (Hopkinson et al., 2014; Mangan et al., 2016; Price and Badger,  
55 1989a, 1989b; Reinhold et al., 1991). Energy-coupled  $C_i$  pumps ensure that the cytosolic  $HCO_3^-$   
56 concentration is high (> 10 mM) and, crucially, out-of-equilibrium with  $CO_2$  (Holthuijzen et al., 1987;  
57 Hopkinson et al., 2014; Kaplan et al., 1980; Price and Badger, 1989a, 1989b; Whitehead et al., 2014).  
58 Inside the carboxysome, the luminal CA converts the high cytosolic  $HCO_3^-$  concentration into a high  
59 carboxysomal  $CO_2$  concentration, which promotes faster carboxylation by Rubisco and also competitively  
60 inhibits oxygenation (Mangan et al., 2016). Genetic lesions to either component - uptake systems or  
61 carboxysomes - disrupt the CCM and mutants require elevated  $CO_2$  for growth (Cai et al., 2009; Dou et al.,  
62 2008; Ogawa et al., 1987) This high- $CO_2$  requiring (HCR) mutant phenotype is commonly used to identify  
63 CCM components in screens (Mackinder et al., 2016; Marcus et al., 1986; Ogawa et al., 1987; Price and  
64 Badger, 1989b).

65 Despite these early screens, a comprehensive list of bacterial CCM components remains unknown,  
66 leaving the possibility that additional activities are required for CCM function. Although well-assembled  
67 carboxysome structures can be heterologously expressed in bacteria and plants (Bonacci et al., 2012;  
68 Fang et al., 2018; Long et al., 2018), functionality of these carboxysomes in a heterologous CCM has not  
69 been demonstrated. Moreover, genetic and bioinformatic studies show that several additional genes are  
70 associated with carboxysome function (Axen et al., 2014; Jorda et al., 2013). For example, it was recently  
71 demonstrated that carboxysome-associated genes may function as Rubisco chaperones and assembly  
72 factors (Aigner et al., 2017; Wheatley et al., 2014). Moreover, many experimental (e.g. (Price and Badger,  
73 1989b; Shibata et al., 2002b)) and modeling studies (e.g. (Hopkinson et al., 2014; Mangan et al., 2016;  
74 Reinhold et al., 1991)) make it clear that energy-coupled  $C_i$  uptake systems are required for the CCM to

75 function. Several different C<sub>i</sub> pump families, including transporters and facilitated uptakes systems are now  
76 known (Long et al., 2016; Price, 2011). However, since model carbon-fixing bacteria often express multiple  
77 C<sub>i</sub> uptake systems and these integral membrane protein systems are difficult to assay biochemically, our  
78 mechanistic biochemical understanding of C<sub>i</sub> uptake is limited (Artier et al., 2018; Battchikova et al., 2011;  
79 Price, 2011).

80 Here we use a genome-wide barcoded transposon mutagenesis screen (RB-TnSeq) to interrogate  
81 the CCM of *Halothiobacillus neapolitanus* (henceforth *Hnea*). *Hnea* is a sulfur oxidizing γ-proteobacterial  
82 chemoautotroph and a model system for studying α-carboxysomes (Heinhorst et al., 2006; Robertson and  
83 Kuenen, 2006). In addition to producing the first catalog of essential genes for a bacterial chemotroph, we  
84 leverage our pooled mutant library to comprehensively screen for knockouts that produce an HCR  
85 phenotype. This screen identified all known CCM components and confirmed that a two-gene operon  
86 containing a large, conserved, poorly-characterized protein (PFAM:PF10070, hereafter DabA) and a  
87 member of a large family of cation transporters (PFAM:PF00361, hereafter DabB) is required for CCM  
88 function. Recent proteomic analyses and physiological experiments have shown that this operon is involved  
89 in C<sub>i</sub> transport in proteobacteria (Mangiapia et al., 2017; Scott et al., 2018). For reasons outlined below, we  
90 term this operon and its gene products DAB for “**DABs Accumulate Bicarbonate**.”

91 As confirmed here, the genes of the DAB operon form a protein complex that is capable of  
92 energetically-coupled C<sub>i</sub> uptake when heterologously expressed in *E. coli*. Both proteins are necessary for  
93 activity in our experiments and treatment with a generic cation ionophore (CCCP) abrogates DAB-mediated  
94 C<sub>i</sub> uptake. Structural homology modeling suggests that DabA contains a domain homologous to a type II □-  
95 carbonic anhydrase. We demonstrate that, like all known type II □-CAs, DabA binds a zinc ion and depends  
96 on two cysteines, one histidine and one aspartic acid residue for activity (Krishnamurthy et al., 2008;  
97 Rowlett, 2010). Taken together, these results suggest that the C<sub>i</sub> uptake systems of proteobacterial  
98 chemotrophs rely on a vectorial CA mechanism that is coupled to a cation gradient (e.g., H<sup>+</sup> or Na<sup>+</sup>). Similar  
99 mechanisms have been proposed for the cyanobacterial C<sub>i</sub> uptake proteins (CUPs) (Han et al., 2017; Price,  
100 2011; Shibata et al., 2002b). Because the cyanobacterial systems appear to associate with a modified  
101 complex I in electron micrographs, they are thought to facilitate C<sub>i</sub> uptake by coupling vectorial CO<sub>2</sub>  
102 hydration with favorable electron flow. However, DAB complexes do not appear to possess any redox active  
103 subunits, nor do they associate with any redox-active proteins (e.g. complex I) in our *E. coli* reconstitution.  
104 We therefore propose a novel model of vectorial CA activity in which DABs couple dissipation of a cation  
105 gradient (e.g. of H<sup>+</sup> or Na<sup>+</sup>) to active hydration of CO<sub>2</sub> to HCO<sub>3</sub><sup>-</sup> in the cytosol. The net effect of this  
106 proposed activity would be energetically-coupled C<sub>i</sub> uptake compatible with CCM function.

107 **Results**

108 *Transposon mutagenesis and gene essentiality*

109 We generated a randomly-barcoded genome-wide pooled knockout library of *Hnea* by conjugation  
110 (Wetmore et al., 2015). This process is diagrammed in Figure 1A. The donor strain (*E. coli* APA 766)  
111 contains a vector with a barcoded Tn5-derived transposon encoding a kanamycin resistance marker.  
112 Conjugation was performed under 5% CO<sub>2</sub> so that CCM genes could be knocked out and the resulting  
113 *Hnea* conjugants were selected for growth in the presence of kanamycin at 5% CO<sub>2</sub> to ensure transposon  
114 insertion.

115

116 The presence of a unique barcode in each transposon simplifies the use of the library for pooled screens  
117 (Wetmore et al., 2015). However, transposon insertion sites and associated barcodes must be mapped to  
118 the *Hnea* genome in order to perform these screens. We mapped transposon insertions using standard  
119 TnSeq methods (Wetmore et al., 2015) and found that our library contains ~10<sup>5</sup> insertions, or roughly one  
120 insertion for every ~25 base pairs in the *Hnea* genome. Since the average gene contains ~35 insertions,  
121 genes with no insertions are almost certainly essential for growth (Rubin et al., 2015). Following this logic,  
122 we used a simple statistical model to identify 551 essential genes and 1787 nonessential genes out of 2408  
123 genes in the *Hnea* genome (Methods, Figure 1A-B, Table S2). The remaining 70 genes were classified as  
124 “ambiguous” due either to their short length or because replicate mapping experiments were discordant  
125 (Methods). Genes associated with known essential functions including central carbon metabolism, ribosome  
126 production, and DNA replication were categorized as essential (Figure 1C). As the library was generated  
127 under 5% CO<sub>2</sub> (Figure 1A) it contains multiple knockouts of known CCM genes, including carboxysome  
128 components (Figure 2C).

129

130 *Comprehensive screen for Hnea CCM components*

131 Based on the current model of the bacterial CCM (diagrammed in Figure 2A) knockouts of CCM genes are  
132 expected to require high CO<sub>2</sub> for growth (Mackinder et al., 2016; Marcus et al., 1986; Price and Badger,  
133 1989b). Strains in our library that harbor CCM gene knockouts should therefore have low fitness in ambient  
134 CO<sub>2</sub> concentrations. As our pooled library contains ~70,000 barcodes that map to exactly one position in the  
135 *Hnea* genome, we were able to use the barseq method to quantify the fitness defects associated with single  
136 gene knockouts for all nonessential *Hnea* genes (Figure 2B). In barseq, a preculture of the library is grown  
137 in permissive conditions (5% CO<sub>2</sub>) and then back-diluted into two conditions: a reference condition (5% CO<sub>2</sub>  
138 again) and a condition of interest (e.g. ambient CO<sub>2</sub>). Genomic DNA is extracted from the preculture (called  
139 t<sub>0</sub>) and both culture outgrowths and barcodes are PCR-amplified and sequenced. In this pooled competition  
140 assay the proportional change in barcode abundance is taken to reflect the fitness effect of gene knockouts  
141 (Wetmore et al., 2015). A CCM gene knockout should have no fitness defect in 5% CO<sub>2</sub> but a large defect in  
142 ambient CO<sub>2</sub>. Since the library contains >20 knockouts with unique barcodes per gene (on average), these  
143 screens contain multiple internal biological replicates testing the effect of single gene knockouts.

144

145 As expected, knockouts to nearly all carboxysome-associated genes produced large fitness defects in  
146 ambient CO<sub>2</sub> (Figures 2B-C). These genes include *cbbLS* - the large and small subunits of the α-  
147 carboxysomal Rubisco; *csos2* - an intrinsically disordered protein required for α-carboxysome assembly;  
148 *csosCA* - the carboxysomal carbonic anhydrase; *csos4AB* - the pentameric proteins thought to form vertices  
149 of the α-carboxysome; and *csos1CAB* - the hexamers that form the faces of the α-carboxysome shell  
150 (Cannon et al., 2001; Heinhorst et al., 2006). Knockouts of *csos1D*, a shell hexamer with a large central  
151 pore (Bonacci et al., 2012; Roberts et al., 2012), confer a very weak HCR phenotype in this screen and so  
152 *csos1D* did not cross the threshold for being called HCR (Figures 2B-C). The *Hnea* genome also contains a  
153 secondary, non-carboxysomal Form II Rubisco that is likely not involved in CCM activity as its disruption  
154 confers no fitness defect in ambient CO<sub>2</sub>. A number of genes that are not structurally associated with the

carboxysome also exhibited HCR phenotypes. These include two LysR transcriptional regulators, a crp/fnr type transcriptional regulator, a protein called acRAF that is involved in Rubisco assembly (Aigner et al., 2017; Wheatley et al., 2014), and two paralogous loci encoding DAB genes (hereafter DAB1 and DAB2, Figure 2B-F).

#### *dabA2 and dabB2 are necessary and sufficient for energy-coupled C<sub>i</sub> accumulation in E. coli*

DAB1 is a cluster of 3 genes found in an operon directly downstream of the carboxysome operon (Figure 2C). Though DAB1 is part of a larger 11-gene operon containing several genes associated with Rubisco proteostasis, including acRAF (Aigner et al., 2017; Wheatley et al., 2014) and a cbbOQ-type Rubisco activase (Mueller-Cajar, 2017), we refer to DAB1 as an “operon” for simplicity. DAB2 is a true operon and is not proximal to the carboxysome operon in the *Hnea* genome. These “operons” are unified in that they both display HCR phenotypes and possess similar genes (Figures 2B-D).

Both operons contain a conserved helical protein of unknown function (PFAM:PF10070, DabA). Since DabA proteins have no predicted transmembrane helices or signal peptides they appear to be large (DabA1: 1046 AA, DabA2: 827 AA), soluble, cytoplasmic proteins (Methods, Figure 3A). DAB1-2 operons also contain a member of the cation transporter family (PFAM:PF00361) that includes H<sup>+</sup>-pumping subunits of respiratory complex I and Mrp Na<sup>+</sup>:H<sup>+</sup> antiporters. This protein, which we call DabB, is smaller than DabA (DabB1: 559 AA, DabB2: 551 AA) and is predicted to have 12-13 transmembrane helices (Figure 3A). The complex I subunits in PF00361 are H<sup>+</sup>-pumping proteins that contain no iron-sulfur clusters, flavin binding sites, or quinone binding sites. Moreover, DabB proteins form a distinct clade in a phylogenetic tree of PF00361. This clade appears to be most closely-related to cyanobacterial proteins involved in C<sub>i</sub> uptake and not complex I subunits (Figure 3 S1). Therefore, homology between DabB and canonical complex I subunits (e.g. NuoL) suggests that DabB is a cation transporter but does not necessarily imply redox activity. Operons of this type were recently demonstrated to be involved C<sub>i</sub> transport in proteobacterial chemotrophs (Mangiapia et al., 2017; Scott et al., 2018).

Since DAB2 disruption is associated with a larger fitness defect than DAB1 (Figure 2B), we used an *E. coli*-based system to test DAB2 for C<sub>i</sub> uptake activity. Knocking out carbonic anhydrases produces an HCR phenotype in *E. coli* (Merlin and Masters, 2003) that is complemented by expression of cyanobacterial bicarbonate transporters (Du et al., 2014). We generated an *E. coli* strain, CAfree, that contains no CA genes (Methods) and found that DAB2 expression enables growth of CAfree in ambient CO<sub>2</sub> (Figure 3B). CAfree complementation requires both DabA2 and DabB2 (Figure 3B) and leads to uptake of radiolabeled C<sub>i</sub> that is substantially above background (grey bars in Figure 3C). Moreover, DAB2-associated C<sub>i</sub> uptake is strongly inhibited by the ionophore CCCP (white bars in Figure 3C), indicating that DAB2 is energetically-coupled, either directly or indirectly, to a cation gradient (e.g. H<sup>+</sup> or Na<sup>+</sup>).

#### *DabA2 and DabB2 interact to form a complex*

In order to determine if the genetic interaction between *dabA2* and *dabB2* is due to a physical interaction, we attempted to purify the two proteins as a complex. DabA2 was genetically fused to a C-terminal Strep-tag, DabB2 was fused to a C-terminal GFP with 6xHis-tag, and the genes were assayed for co-expression in *E. coli* (Methods). Tandem-affinity purification revealed that DabA2 and DabB2 interact physically to form a complex in *E. coli* (Figure 4A). The complex runs as a single major peak on size exclusion chromatography and has a retention volume consistent with a heterodimer of DabA2 and DabB2 (Figure 4B). Notably, we did not observe co-purification of *E. coli* complex I subunits or any other proteins with the DabA-DabB complex (Figure 4A), suggesting that the DAB2 operates as an independent complex within the membrane.

203 *DabA contains a CA-like active site required for zinc binding and activity*

204 Structural homology modeling software predicted that the middle of DabA2 has sequence elements related  
205 to a  $\square$ -CA (Figure 3A). Specifically, Phyre2 predictions identified C539 and H524 as part of a potential Zn<sup>2+</sup>  
206 binding site distantly homologous to a bacterial type II  $\square$ -CA (10% coverage of DabA, 90.8% confidence). I-  
207 TASSER predicted a Zn<sup>2+</sup> binding site including the same residues along with an additional cysteine (C351),  
208 and aspartic acid (D353). As shown in Figure 4C, these residues could make up the active site of a type II  
209  $\square$ -CA (Cronk et al., 2006, 2001; Supuran, 2016). We generated individual alanine mutants for each of these  
210 putative active site residues (C351A, D353A, and H534A) and tested their ability to rescue CAfree. All  
211 mutants failed to produce growth of CAfree in ambient CO<sub>2</sub> (Figure 4D). We proceeded to assay zinc  
212 binding of purified dabAB complex using X-ray fluorescence spectroscopy and found that wild-type dabAB2  
213 and all single mutants bind zinc (Figure 4E). These single mutants retain three of four zinc-coordinating  
214 residues (Rowlett, 2010), which could explain why single mutation was insufficient to abrogate zinc binding.  
215 This is consistent with mutational studies of the human CA II, where mutation of Zn<sup>2+</sup>-binding residues  
216 reduces but does not abrogate zinc binding (Ippolito et al., 1995; Krishnamurthy et al., 2008).

217  
218 *Purified DAB2 does not have conspicuous CA activity.*

219 We tested whether purified DabAB2 had CA activity (Figure 4F) but no obvious CA activity was observed.  
220 We tested for activity in CO<sub>2</sub> concentrations that are typically saturating for CAs and at high concentrations  
221 of purified DabAB2 (> 650-fold more protein than the positive control) but did not detect any activity (Figure  
222 4F). We estimate that activities 20x lower than the positive control would have been detected. The absence  
223 of activity *in vitro* implies either that DabAB2 has extremely low activity or that DabAB2 must reside in a cell  
224 membrane holding a cation gradient to function as an activated carbonic anhydrase.

225 **Discussion**

226 Since oxygenic photosynthesis is responsible for our contemporary O<sub>2</sub>-rich and relatively CO<sub>2</sub>-poor  
227 atmosphere, it is likely that Rubisco evolved in an ancient CO<sub>2</sub>-rich environment where its modest rate and  
228 limited specificity posed no problem (Shih et al., 2016; Tabita et al., 2008). However, over the subsequent  
229 2.5 billion years of Earth's history, atmospheric O<sub>2</sub> increased and CO<sub>2</sub> declined to the point where, today,  
230 autotrophic bacteria that grow in atmosphere appear to uniformly have CCMs (Raven et al., 2017). Bacterial  
231 CCMs come in two convergently-evolved forms - α-carboxysomes are found in proteobacteria and marine  
232 cyanobacteria while β-carboxysomes are found in freshwater cyanobacteria (Rae et al., 2013). Because the  
233 bacterial CCM is well-studied and known to function in single cells it is an attractive target for synthetic  
234 biology and efforts to transplant it into crops are already underway (Lin et al., 2014; Long et al., 2018;  
235 Occhialini et al., 2016).

236 In principle, the bacterial CCM requires two major components: i. energy-coupled uptake of inorganic  
237 carbon to concentrate HCO<sub>3</sub><sup>-</sup> in the cytosol and ii. carboxysome structures that co-localize Rubisco with CA  
238 enzymes that convert concentrated HCO<sub>3</sub><sup>-</sup> into a high concentration of the Rubisco substrate CO<sub>2</sub> (Mangan  
239 et al., 2016). While the carboxysome components are well-documented for both α- and β-families, C<sub>i</sub> uptake  
240 systems of the proteobacterial CCM have only been identified very recently (Mangiapia et al., 2017; Scott et  
241 al., 2018). Moreover, though numerous laboratories have spent decades studying the bacterial CCM, it  
242 remains unclear whether our current "parts list" for α- and β-CCMs is complete.

243 Here we undertook an effort to complete the genetic "parts list" of the α-family CCM of the proteobacterial  
244 chemotroph *H. neapolitanus*. We generated a genome-wide knockout library containing ~35 individual  
245 knockouts for every gene in the *Hnea* genome and compiled the first list of essential genes for a  
246 chemotroph (Figure 1). Because we generated the library at elevated CO<sub>2</sub> (5%, Figure 1A) we were able to  
247 knockout all known CCM components, including all genes known to form the α-carboxysome (Figure 2C).  
248 We subsequently used this library to screen for genes associated with CCM activity by screening for  
249 knockouts with fitness defects specific to ambient CO<sub>2</sub> growth conditions (Figure 2B). As expected, this  
250 screen identified most carboxysome components and highlighted several genes whose relationship to the  
251 CCM is not fully understood (Figures 2B-F). These genes include several transcriptional regulators, a  
252 putative Rubisco chaperone and two small operons (DAB1 and DAB2) that are involved in CCM-associated  
253 C<sub>i</sub> uptake in chemotrophic proteobacteria (Mangiapia et al., 2017; Scott et al., 2018).

254 Freshwater cyanobacteria express several well-studied C<sub>i</sub> transporters (Price, 2011) that take up HCO<sub>3</sub><sup>-</sup> and  
255 are coupled to energy in the form of ATP or an Na<sup>+</sup> gradient. The substrate, energy coupling, and chemical  
256 mechanism are unclear for the recently-identified proteobacterial transporters (Mangiapia et al., 2017; Scott  
257 et al., 2018). We note, however, that the preferred substrate for C<sub>i</sub> uptake will depend on the extracellular  
258 pH because pH determines the relative abundance of CO<sub>2</sub>, H<sub>2</sub>CO<sub>3</sub>, HCO<sub>3</sub><sup>-</sup> and CO<sub>3</sub><sup>2-</sup> (Mangan et al., 2016).  
259 Since *Hnea* and many other sulfur-oxidizing proteobacteria are acidophilic and CO<sub>2</sub> is more abundant than  
260 HCO<sub>3</sub><sup>-</sup> at acidic pH (Figure 5 S2), it stands to reason that they might have evolved a mechanism to take up  
261 CO<sub>2</sub> instead of HCO<sub>3</sub><sup>-</sup>.

262 We showed that the DAB2 operon encodes a two-component protein complex that has C<sub>i</sub> uptake activity in  
263 *E. coli* (Figure 3B-C). This complex may be a heterodimer, as suggested by size-exclusion chromatography  
264 (Figure 4B). As this activity is strongly inhibited by the ionophore CCCP (Figure 3C), we suspect that DAB2-  
265 mediated C<sub>i</sub> uptake is energetically-coupled to a cation gradient (Figure 5A). Moreover, the DabA unit of this  
266 complex has limited homology to a type II β-carbonic anhydrase and binds a zinc (Figures 3-4). Mutations to  
267 the putative zinc-binding residues (C351A, D353A, and H534A) ablate function in-vivo, but do not abolish  
268 zinc binding (Figure 4D-E). For all these reasons, we propose a model of DAB activity wherein CO<sub>2</sub> is

passively taken into the cell and then vectorially (unidirectionally) hydrated to  $\text{HCO}_3^-$  by DabA. Model carbonic anhydrases are not directly coupled to any energy source (e.g., ATP) and so they only accelerate the equilibration of  $\text{CO}_2$  and  $\text{HCO}_3^-$  (Krishnamurthy et al., 2008; Supuran, 2016). Coupling conversion of  $\text{CO}_2$  into  $\text{HCO}_3^-$  to dissipation of an existing cation gradient, though, would result in unidirectional hydration and enable the DAB system to actively accumulate  $\text{HCO}_3^-$  in the cytosol and power the CCM (as diagrammed in Figure 2A). We draw this activity as being coupled to the  $\text{H}^+$  gradient in Figure 5A for simplicity, but our results are equally consistent with other cation gradients, e.g.  $\text{Na}^+$ . This mechanism requires tight coupling of cation flow to  $\text{CO}_2$  hydration by the CA-like DabA protein, which is consistent with our observation that purified DabAB2 displays no measurable CA activity. Notably, type II  $\beta$ -CAs are the only CAs that display allosteric regulation (Cronk et al., 2006; Rowlett, 2010). Allosteric control is postulated to be mediated by  $\text{Zn}^{2+}$  binding and unbinding by the active site aspartic acid (D353 in DabA2). A similar mechanism might couple ion movement through DabB to the active site of DabA (schematized in Figure 5A).

Cyanobacteria possess two distinct uptake systems (CupA/B) that perform vectorial conversion of  $\text{CO}_2$  to bicarbonate (Maeda et al., 2002; Price, 2011; Rae et al., 2013; Shibata et al., 2002b, 2001). Unlike in *Hnea*, however, these are typically secondary transporters that are not required for growth in standard lab conditions. CupA/B have proven challenging to study for this reason. Because CupA/B appear to associate with the cyanobacterial complex I in transmission electron micrographs (Battchikova et al., 2011; Birungi et al., 2010), their  $\text{CO}_2$  hydration activity is thought to be coupled to energetically-favorable electron flow (Figure 5 S1). Though DabB is part of the MrpA protein family (PF00361) that also contains the  $\text{H}^+$ -pumping subunits of complex I, this is a broad and diverse protein family (Figure 3 S1) that contains many cation transporters (e.g.  $\text{H}^+:\text{Na}^+$  antiporters) that do not associate with complex I or any other redox-coupled membrane complex (Krulwich et al., 2009; Mangiapia et al., 2017; Marreiros et al., 2013). Moreover, as shown in Figures 3-4, the DAB complex functions in *E. coli* but does not appear to engage the *E. coli* complex I. Rather, the two subunits of the DAB complex co-purify alone (Figure 4A), suggesting that they function as a single unit in the *E. coli* membrane. Moreover, treatment with the ionophore CCCP strongly inhibits DAB2 activity, implying that a gradient is important for activity. We therefore propose that DAB activity is coupled to a cation gradient and not electron flow (Figure 5A).

We observed that DabAB2 functions substantially better in CAfree *E. coli* than SbtA (Figures 3C and 3S4), the primary inorganic carbon transporter of model freshwater cyanobacteria (Du et al., 2014; Rae et al., 2013). As *E. coli* and *Hnea* are both proteobacteria, this observation is likely due to greater “compatibility” of proteobacterial proteins with *E. coli* expression as compared to proteins derived from cyanobacteria. It may also be the case that the  $\alpha$ -CCM of proteobacteria is more “portable” than the  $\beta$ -CCM of freshwater cyanobacteria. Indeed,  $\alpha$ -CCM genes are typically found in a single gene cluster in chemoautotrophs throughout  $\alpha$ -  $\beta$ - and  $\gamma$ -proteobacteria and the  $\alpha$ -CCM was clearly horizontally transferred at least once from proteobacteria to marine cyanobacteria (Rae et al., 2013). We examined the phylogeny of DabA1-2 homologs in prokaryotes and found that they are both widespread and likely to have undergone multiple horizontal transfer events (Figure 5B). Since DabAB2 appears to be so much more active in *E. coli* than SbtA and the  $\alpha$ -CCM appears to have undergone widespread horizontal transfer, DAB-family transporters are an attractive target for protein engineering and heterologous expression in plants and industrial microbes, where elevated intracellular  $\text{C}_i$  could be technologically useful (Antonovsky et al., 2016).

Finally, we were surprised to find evidence of DABs outside of known carbon-fixing bacteria. For example, high-confidence DabA homologs are found in notable heterotrophic pathogens including *V. cholerae*, *B. anthracis*, *L. pneumophila* (Figure 5B). Carbonic anhydrase activity is essential for heterotrophic growth of *E. coli* and *S. cerevisiae* in ambient  $\text{CO}_2$  (Aguilera et al., 2005; Merlin and Masters, 2003) and is required for

321 growth or virulence of several pathogens including *M. tuberculosis* and *H. pylori* (Supuran, 2008). In the  
322 heterotrophic context, CA activity is thought to supply bicarbonate for the biotin-dependent carboxylases of  
323 central metabolism, for which  $\text{HCO}_3^-$  is the true substrate (Aguilera et al., 2005; Merlin and Masters, 2003).  
324 Prokaryotic CAs may also be involved in pH regulation (Supuran, 2008). Perhaps DAB-family C<sub>i</sub> uptake  
325 systems play similar roles in these important pathogens? We hope that future research will delineate the  
326 role of energetically-activated C<sub>i</sub> uptake in clades that do not perform net carbon fixation.

<b>Reagent type (species) or resource</b>	<b>Designation</b>	<b>Source or reference</b>	<b>Identifiers</b>	<b>Additional information</b>
strain, strain background ( <i>Escherichia coli</i> )	APA 766	(Wetmore et al., 2015)		APA 766 obtained from the Arkin laboratory at University of California, Berkeley
strain, strain background ( <i>Escherichia coli</i> )	CAfree	This paper		BW25113 ΔcanA ΔcynT
strain, strain background ( <i>Escherichia coli</i> )	BL21-AI	Thermo fischer catalog C607003		F-ompT hsdS <sub>B</sub> (r <sub>B</sub> <sup>-</sup> m <sub>B</sub> <sup>-</sup> ) gal dcm araB::T7RNAP-tetA
strain, strain background ( <i>Halothiobacillus neapolitanus</i> c2)	Hnea	ATCC 23641		
Gene ( <i>Halothiobacillus neapolitanus</i> c2)	dabA2	This paper	Uniprot: D0KWS7	
Gene ( <i>Halothiobacillus neapolitanus</i> c2)	dabB2	This paper	Uniprot: D0KWS8	
Gene ( <i>Escherichia coli</i> )	eCA	(Cronk et al., 2001)	Uniprot: P61517	
Gene ( <i>Homo sapiens</i> )	hCA	(Murakami et al., 1987)	Uniprot: P00918	
Gene ( <i>Synechococcus elongatus</i> )	sbtA	(Du et al., 2014; Shibata et al., 2002a)	Uniprot: Q6UP27	
Gene ( <i>Aequorea victoria</i> )	sfGFP	(Pédelacq et al., 2006)		
recombinant DNA reagent	pFE	This paper		A plasmid backbone adapted from pZE21 by the addition of a tet repressor gene (Lutz and Bujard, 1997). aTc inducible and

				carries a kanamycin resistance marker.
recombinant DNA reagent	pFE-sfGFP	This paper		pFE carrying a sfGFP gene
recombinant DNA reagent	pFE-hCA	This paper		pFE carrying a human carbonic anhydrase II gene (Uniprot: P00918)
recombinant DNA reagent	pFE-dabB2	This paper		pFE carrying a dabB2 gene (Uniprot: D0KWS8)
recombinant DNA reagent	pFE-dabA2	This paper		pFE carrying a dabA2 gene (Uniprot: D0KWS7)
recombinant DNA reagent	pFE-dabAB2	This paper		pFE carrying both the dabA2 (Uniprot: D0KWS7) and dabB2 genes (Uniprot: D0KWS8)
recombinant DNA reagent	pFE-dabAB2 C351A	This paper		pFE carrying a dabA2 gene with a C351A mutation and a dabB2 gene (Uniprot: D0KWS8)
recombinant DNA reagent	pFE-dabAB2 D353A	This paper		pFE carrying a dabA2 gene (Uniprot: D0KWS7) with a D353A mutation and a dabB2 gene (Uniprot: D0KWS8)
recombinant DNA reagent	pFE-dabAB2 H524A	This paper		pFE carrying a dabA2 gene (Uniprot: D0KWS7) with a H524A mutation and a dabB2 gene (Uniprot: D0KWS8)
recombinant DNA reagent	pFE-dabAB2 C539A	This paper		pFE carrying a dabA2 gene (Uniprot: D0KWS7) with a C539A mutation and a dabB2 gene (Uniprot: D0KWS8)
recombinant DNA	pFA	This paper		A plasmid

reagent				backbone adapted from pZA31 by the addition of a tet repressor gene (Lutz and Bujard, 1997). aTc inducible and carries a chloramphenicol resistance marker.
recombinant DNA reagent	pFA-GFP	This paper		pFA carrying a sfGFP gene
recombinant DNA reagent	pFA-eCA	This paper		pFA carrying a <i>E. coli</i> canA gene (Uniprot: P61517)
recombinant DNA reagent	pFA-sbtA	This paper		pFA carrying a <i>Synechococcus elongatus</i> PCC 7942 sbtA gene (Uniprot: Q6UP27)
recombinant DNA reagent	pFA-dabAB2	This paper		pFA carrying the dabA2 (Uniprot: D0KWS7) and dabB2 genes (Uniprot: D0KWS8)
recombinant DNA reagent	pET14b-dabAB2	This paper		pET14b carrying the dabA2 gene (Uniprot: D0KWS7) with a c-terminal strep tag fusion and the dabB2 gene (Uniprot: D0KWS8) with a c-terminal sfGFP V206K fusion and 6xHis tag
recombinant DNA reagent	pET14b-dabAB2 C351A	This paper		pET14b carrying the dabA2 gene (Uniprot: D0KWS7) with a D353A mutation fused to a c-terminal strep tag and the dabB2 gene (Uniprot: D0KWS8) with a c-terminal sfGFP V206K fusion and 6xHis tag

recombinant DNA reagent	pET14b-dabAB2 D353A	This paper		pET14b carrying the dabA2 gene (Uniprot: D0KWS7) with a H524A mutation fused to a c-terminal strep tag and the dabB2 gene (Uniprot: D0KWS8) with a c-terminal sfGFP V206K fusion and 6xHis tag
-------------------------	---------------------	------------	--	--

329

330 *Bacterial strains and growth conditions*

331 *E. coli* strain APA 766 was used as the conjugation donor to transfer the Tn5 transposon to *Halothiobacillus*  
 332 *neapolitanus* C2 (*Hnea*) via conjugation (Wetmore et al., 2015). The *E. coli* double CA deletion strain  
 333 “CAfree” (BW25113  $\Delta$ canA  $\Delta$ cynT) was generated by curing the KEIO collection *cynT* knockout (BW25113  
 334  $\Delta$ cynT, KEIO strain JW0330) of kanamycin resistance via pCP20-mediated FLP recombination and  
 335 subsequent P1 transduction (and curing) of kanamycin resistance from the *canA* knockout strain EDCM636  
 336 (MG1655  $\Delta$ canA, Yale Coli Genomic Stock Center, (Baba et al., 2006; Merlin and Masters, 2003)).  
 337 Lysogeny broth (LB) and LB agar were used as *E. coli* growth media unless otherwise specified. *E. coli*  
 338 strains were grown at 37 °C in the presence of 0.1 mg/ml Carbenicillin, 0.06 mg/ml Kanamycin, or 0.025  
 339 mg/ml Chloramphenicol as appropriate. *Hnea* was grown in DSMZ-68 media at 30 °C and in the presence of  
 340 0.03 mg/ml Kanamycin when appropriate.

341

342 *Transposon mutagenesis and RB-TnSeq library production*

343 A barcoded library of *Hnea* transposon mutants was generated by adapting the methods of (Wetmore et al.,  
 344 2015). Conjugations were performed as follows. *Hnea* and APA 766 were cultured and harvested by  
 345 centrifugation. Both cultures were washed once in 10 mL antibiotic-free growth media per conjugation  
 346 reaction and resuspended in 100  $\mu$ L. 5 OD600 units of *Hnea* were mixed with 20 OD600 units of APA 766 on  
 347 a 0.45  $\mu$ M millipore MCE membrane filter and cultured overnight at 30 °C in 5% CO<sub>2</sub> on an antibiotic-free LB  
 348 agar plate containing 0.06 mg/ml diaminopimelic acid. Cells were scraped from the filter into 2 mL DSMZ-68  
 349 and collected in a 2 mL microcentrifuge tube. Recovered cells were pelleted by centrifugation at 16000  $\times$  g  
 350 for 1 minute, washed in 2 mL DSMZ-68, pelleted again at 9000  $\times$  g for 1 minute, and resuspended in 2 mL  
 351 DSMZ-68 before 200  $\mu$ L was plated onto 10 separate DSMZ-68 kanamycin plates (per conjugation). Plates  
 352 were incubated at 30 °C under 5% CO<sub>2</sub> until colonies formed (~ 7 days). Colonies were counted and  
 353 scraped into 55 mL DSMZ-68. Two 1.4 OD600 unit samples were taken and used to prepare genomic DNA  
 354 (Qiagen DNeasy blood and tissue kit). Transposon insertions were amplified from gDNA following protocols  
 355 in (Wetmore et al., 2015). Transposons were mapped after Illumina sequencing using software developed in  
 356 (Wetmore et al., 2015) 1.6 OD600 unit aliquots were then flash frozen in 50% glycerol for subsequent Bar-  
 357 seq experiments.

358

359 *Essential gene assignment*

360 Following the logic of (Rubin et al., 2015; Wetmore et al., 2015), we categorized genes as essential if we  
 361 observed significantly fewer transposon insertions than would be expected by chance. If insertion occurred  
 362 uniformly at random, the number of insertions per gene would be expected to follow a binomial distribution.  
 363 The probability of observing at most  $k$  insertions into a gene of length  $n$  is therefore expressed as:

$$P(k; n, p) = \sum_{i=0}^{i=k} \frac{n!}{k!(n-k)!} p^i (1-p)^{n-i}$$

364 Here,  $p$  is the average rate of transposon insertion per base pair genome-wide. Genes were determined to  
 365 be essential if they received a lower-than-expected number of insertions in both replicates of the library  
 366 mapping, i.e. if the probability of observing  $k$  or fewer insertions was beneath 0.05 after Bonferroni  
 367 correction. Genes were called “ambiguously essential” in two cases: (i) the replicates were discordant or (ii)  
 368 zero insertions were observed but the gene was short enough that the formula could not yield a Bonferroni-  
 369 corrected probability below 0.05 threshold even in the case of zero insertions.  
 370

### 371 *Gene fitness experiments*

372 Fitness experiments were performed according to a modification of the protocol in (Wetmore et al., 2015). A  
 373 library aliquot was thawed and used to inoculate three 33 mL cultures. Cultures were grown to OD600 ~0.08  
 374 in 5% CO<sub>2</sub>. At this point, 20 mL were removed and harvested by centrifugation as two t<sub>0</sub> (input) samples.  
 375 Cultures were back-diluted 1:64 into 128 mL and incubated for 6.5-7.5 doublings under 5% CO<sub>2</sub> or ambient  
 376 conditions. 50 mL of culture was harvested by centrifugation. gDNA was prepared and barcodes were  
 377 amplified for fitness determination via Illumina sequencing as described in (Wetmore et al., 2015).  
 378

### 379 *CAfree rescue experiments*

380 Electrocompetent CAfree cells were prepared using standard protocols and transformed with pFE plasmids  
 381 containing genes of interest by electroporation. CAfree pre-cultures were grown overnight in 10% CO<sub>2</sub> and  
 382 diluted into 96 well plates (3 µl cells in 250 µl media). Growth curves were measured by culturing cells in a  
 383 Tecan M1000 microplate reader under ambient conditions with continuous shaking, and measuring OD600  
 384 every 15 minutes. When samples are marked “induced,” 200 nM anhydrotetracycline (aTc) was added to  
 385 the media. Growth yields are calculated as the maximum OD600 achieved after 24 hours of growth and  
 386 normalized to the yield of a wild type control.  
 387

### 388 *Silicone oil centrifugation measurement of inorganic carbon uptake*

389 The silicone oil filtration method was modified from (Dobrinski et al., 2005) and used to measure uptake of  
 390 labeled inorganic carbon. Assay tubes were generated using 0.6 ml microcentrifuge tubes containing 20 µl  
 391 of dense kill solution (66.7% v/v 1 M glycine pH 10, 33.3% v/v triton X-100) covered by 260 µl of silicone oil  
 392 (4 parts AR20:3.5 parts AR200). Electrocompetent CAfree cells were prepared using standard protocols  
 393 and transformed with pFA plasmids containing genes of interest by electroporation. CAfree cultures were  
 394 grown overnight in 10% CO<sub>2</sub>, back diluted to an OD600 of 0.1 and allowed to grow to mid-log phase in 10%  
 395 CO<sub>2</sub> in the presence of 200 nM aTc for induction. Cells were then harvested by centrifugation, washed once  
 396 in PBS (pH 7.0) and resuspended to OD600 0.6 in PBS + 0.4% glucose. <sup>14</sup>C-labeled sodium bicarbonate  
 397 (PerkinElmer) was added to a final concentration of 4.1 nM and an activity of 0.23 µCi. Cells were incubated  
 398 with <sup>14</sup>C for 4 minutes before centrifugation at 17,000 x g for 4 minutes to separate cells from buffer. Pellets  
 399 were clipped into scintillation vials containing 5 ml Ultima Gold scintillation fluid and 300 µl 3M NaOH using  
 400 microcentrifuge tube clippers or medium dog toenail clippers. Counts were measured on a PerkinElmer  
 401 scintillation counter. <sup>14</sup>C counts are normalized to 1 OD600 unit of cells added. During inhibition assays,  
 402 cells were incubated in PBS pH 7 with 0.4% glucose + 0.4% DMSO and the inhibitor (100 µM CCCP) for 10  
 403 minutes before assay.  
 404

### 405 *Generation of DabA Phylogenetic Tree*

406 We searched the Uniprot reference proteome database using the Pfam Hidden Markov Model PF10070.9  
 407 with a cutoff e-value of 1e<sup>-4</sup>. Our search recovered 941 candidate DabA proteins. These sequences were  
 408 aligned using MAFFT and manually pruned to remove fragments and poorly aligning sequences. The  
 409 remaining 878 candidate DabA sequences were re-aligned MAFFT and an approximate maximum

410 likelihood phylogenetic tree was constructed using FastTree. Taxonomy was assigned to nodes in the tree  
411 based on NCBI taxonomy information for the genomes harboring each sequence.  
412

#### 413 *Generation of DabB Phylogenetic Tree*

414 DabB homologs were collected manually by searching MicrobesOnline for close homologs of four PF00361  
415 members in the *Hnea* genome (dabB1, dabB2, Hneap\_1953, Hneap\_1130) and other characterized  
416 PF00361 members including *Syneccococcus elongatus* ndhF1, *Syneccococcus elongatus* ndhF3, and  
417 *Syneccococcus elongatus* ndhF4. Genes were clustered to 95% similarity and genes with divergent operon  
418 structure were removed manually using MicrobesOnline treeview (Dehal et al., 2010). NuoL from  
419 *Escherichia coli*, Nqo12 from *Thermus thermophilus*, and NdhF1/3/4 from *Thermosynechococcus elongatus*  
420 BP-1 were added as markers. ClustalOmega was used to construct a multiple sequence alignment and the  
421 resulting nearest-neighbor tree was visualized using the Interactive Tree of Life (Letunic and Bork, 2016;  
422 Sievers and Higgins, 2018).

#### 423 *Protein Annotation and Structural Homology Modeling*

424 Secondary structural annotations for DabAB2 were generated using XtalPred (Slabinski et al., 2007).  
425 Structural Homology modeling of DabA was performed using Phyre2 and I-TASSER web servers with  
426 default parameters (Kelley et al., 2015; Roy et al., 2010). A list of close DabB homologs was assembled by  
427 searching MicrobesOnline for PF00361 members with similar operon structure. A ClustalOmega alignment  
428 was used to calculate residue-level conservation of DabB proteins while the MAFFT alignment generated  
429 during the creation of the DabA tree was used to calculate residue level conservation of DabA proteins  
430 (Figure 3 S1).

#### 431 *Purification of DAB2*

432 Chemically competent BL21-AI *E. coli* were transformed with pET14b vectors containing dabAB constructs.  
433 1 liter of 2xYT media was inoculated with 20 ml of an overnight culture of BL21-AI *E. coli* in LB+CARB and  
434 allowed to grow to mid log at 37 °C. When midlog was reached, cells were induced with 20 ml of 50 mg/ml  
435 arabinose and transitioned to 20 °C for overnight growth. Cultures were pelleted and resuspended in 10 ml  
436 TBS (50 mM Tris, 150 mM NaCl, pH 7.5) supplemented with 1.2 mM phenylmethylsulfonyl fluoride, 0.075  
437 mg/ml lysozyme and 0.8 ug/ml DNase I per liter of starting culture and then incubated at room temperature  
438 on a rocker for 20 minutes. Cells were lysed with four passes through a homogenizer (Avestin). Lysate was  
439 clarified at 15,000 x g for 30 minutes. Membranes were pelleted at 140,000 x g for 90 minutes. Membrane  
440 pellets were resuspended overnight in 25 ml TBS supplemented with 1 mM phenylmethylsulfonyl fluoride  
441 and 1% β-dodecyl-maltoside (DDM, Anatrace) per liter of culture following (Newby et al., 2009). Membranes  
442 were then repelleted at 140,000 - 200,000 x g for 60 minutes and the supernatant was incubated with Ni-  
443 NTA beads (Thermo Fisher) for 90 min at 4 °C. The resin was washed with “Ni buffer” (20 mM Tris + 300  
444 mM NaCl + 0.03% DDM, pH 7.5) supplemented with 30 mM imidazole and eluted with Ni buffer  
445 supplemented with 300 mM imidazole. Eluent was then incubated with Strep-Tactin (Millipore) resin for 90  
446 min at 4 °C. Resin was washed with “strep buffer” (TBS + 0.03% DDM) and eluted with strep buffer  
447 supplemented with 2.5 mM desthiobiotin. Eluent was concentrated using Vivaspin 6 100 kDa spin  
448 concentrators and buffer exchanged into strep buffer by either spin concentration or using Econo-Pac 10DG  
449 (Biorad) desalting columns. For analytical purposes, 300 µg of strep-purified protein was injected onto a  
450 Superdex 200 Increase 3.2/300 size-exclusion column pre-equilibrated in strep buffer and eluted  
451 isocratically in the same buffer.

#### 452 *CA Assays*

453 CA catalyzed CO<sub>2</sub> hydration of purified DAB2 complex and human carbonic anhydrase (hCA) was  
454 measured using the buffer/indicator assay of Khalifah (Khalifah, 1971) on a KinTek AutoSF-120 stopped-

458 flow spectrophotometer at 25 °C. The buffer/indicator pair used was TAPS/*m*-cresol purple measured at a  
459 wavelength of 578 nm using a pathlength of 0.5 cm. Final buffer concentration after mixing was 50 mM  
460 TAPS, pH 8.0 with the ionic strength adjusted to 50 mM with Na<sub>2</sub>SO<sub>4</sub>, and 50 µM of pH-indicator. Final  
461 protein concentration used was: 9.8 µM DAB2 (His-elution) and 0.015 µM hCA (positive control; Sigma  
462 Aldrich C6624). Saturated solution of CO<sub>2</sub> (32.9 mM) was prepared by bubbling CO<sub>2</sub> gas into milli-Q water  
463 at 25 °C. The saturated solution was injected into the stopped-flow using a gas-tight Hamilton syringe, and  
464 measurements were performed in a final CO<sub>2</sub> concentration of 16.5 mM. Progression curves were  
465 measured in 7 replicates.

466

467 *X-ray fluorescence spectroscopy for metal analysis*

468 50-100 µg of protein dissolved in 20-200 µl of TBS + 0.03% DDM was precipitated by addition of 4 volumes  
469 of acetone and incubation at -20 °C for 1 hour. Samples were centrifuged at 21,130 x g for 15 minutes in a  
470 benchtop centrifuge and the supernatant was removed. Pellets were stored at 4 °C until analysis.  
471 Fluorescence analysis was performed by breaking up the pellet into 5 µl of TBS + 0.03% DDM with a  
472 pipette tip. Small pieces of the pellet were looped with a nylon loop and flash frozen in place on a  
473 goniometer under a nitrogen stream. The sample was excited with a 14 keV X-ray beam and a fluorescence  
474 spectrum was collected. Sample emission spectra were then used to identify metals. Metal analysis was  
475 performed on wild-type DAB2, Zn-binding mutants C351A and D353A, bovine CA (positive control; Sigma  
476 Aldrich C7025) and a buffer blank was used as a negative control. A Rubisco crystal with a containing  
477 cobalt salts was also used as a zinc free control. Displayed traces are averages of at least two experiments.  
478 Experiments were performed at the Lawrence Berkeley National Laboratory Advanced Light Source  
479 Beamline 8.3.

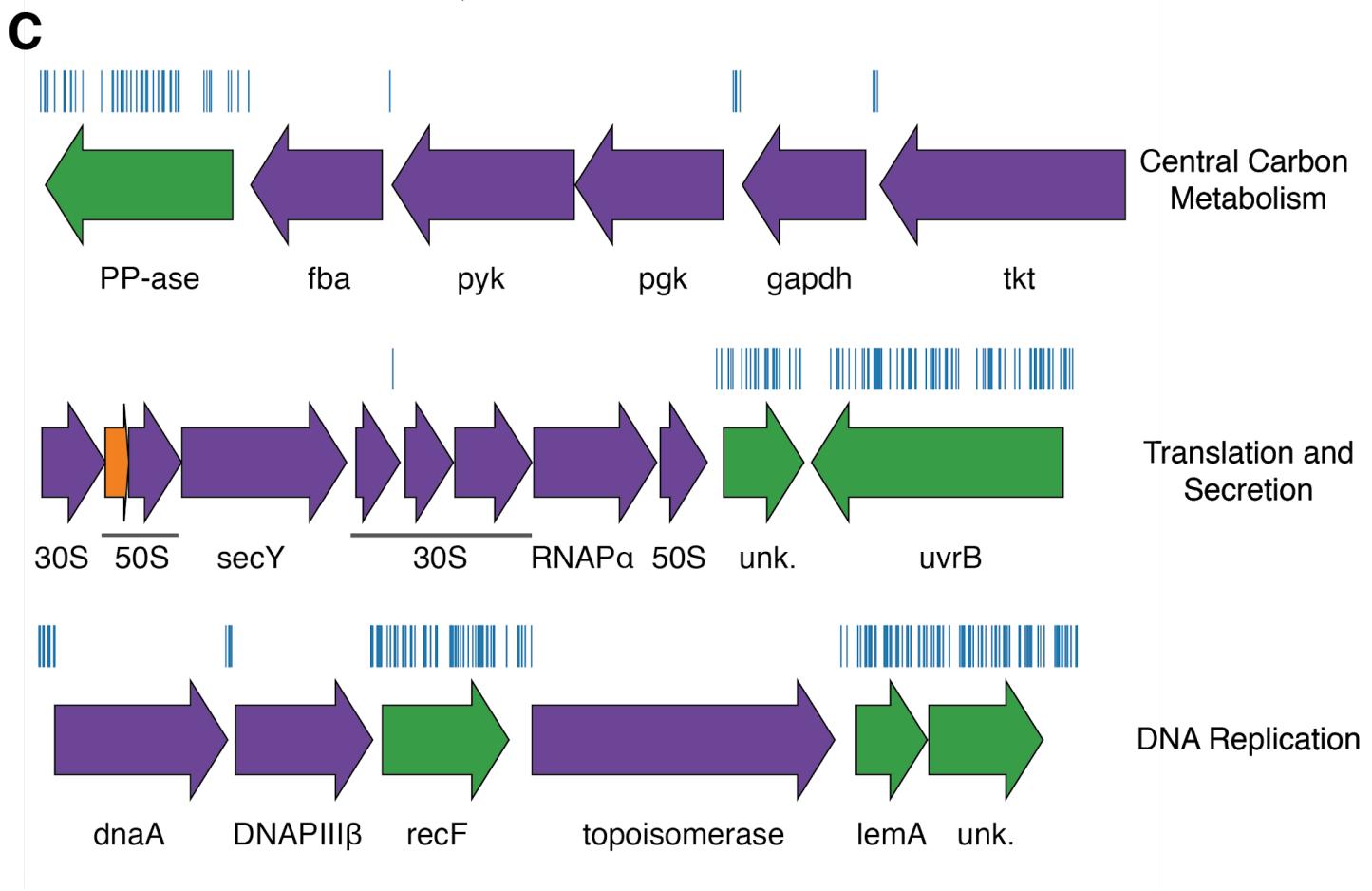
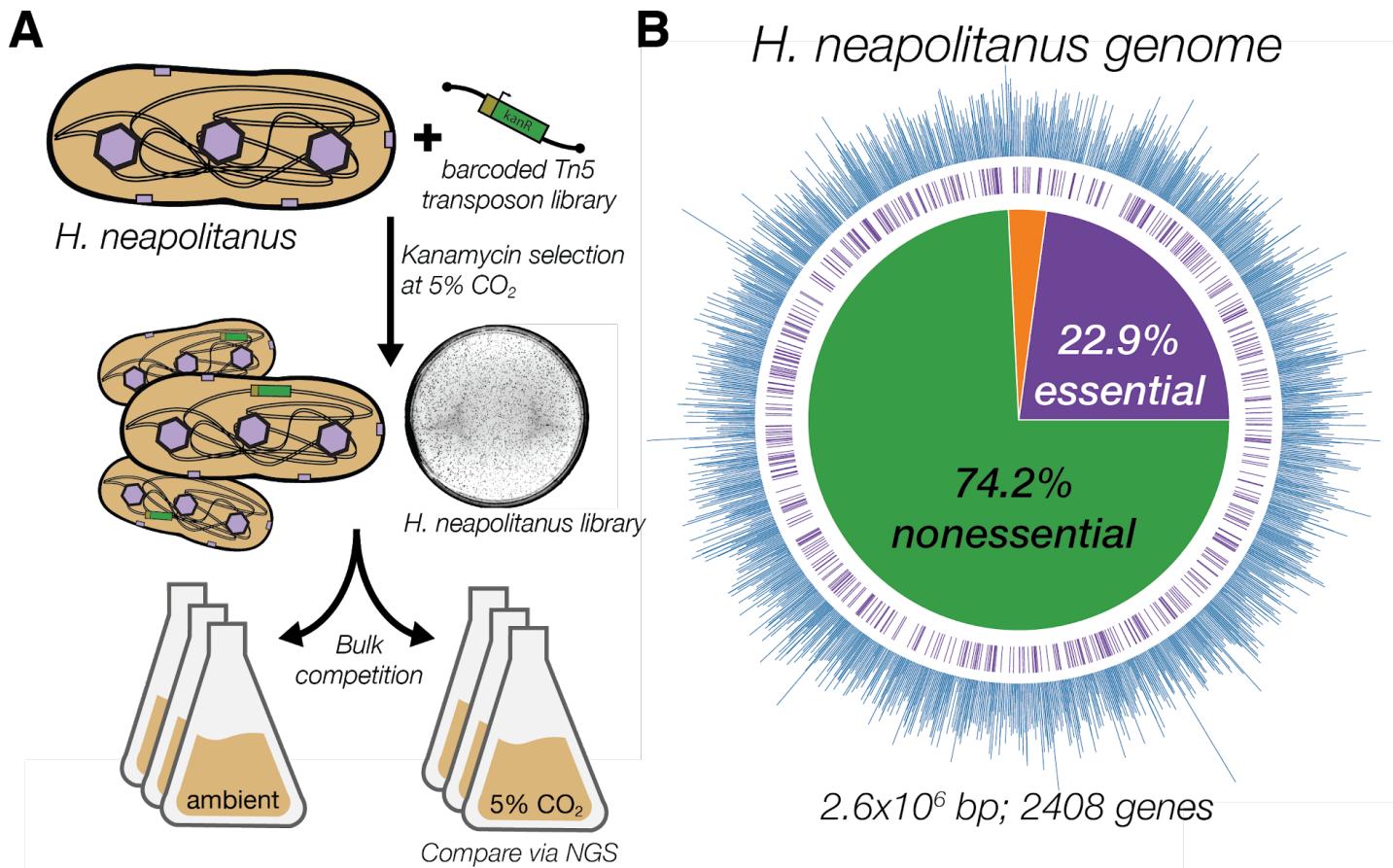
480 **Acknowledgements**

481 We thank Adam Deutschbauer and Morgan Price for assistance with RB-TnSeq experiments and analysis,  
482 respectively. We also thank Andreas Martin and Jared Bard for assistance with stopped flow experiments.  
483 Thanks to Emeric Charles, Woodward Fischer, Britta Forster, Ben Long, Robert Nichols, Dean Price and  
484 Patrick Shih for useful conversations and comments on the manuscript. X-ray-based experiments were  
485 performed at the Lawrence Berkeley National Laboratory Advanced Light Source Beamline 8.3.1. J.J.D.  
486 was supported by National Institute of General Medical Sciences grant-T32GM066698. A.F. and T.G.L.  
487 were supported by a National Science Foundation Graduate Research Fellowship. C.B. was supported by  
488 an International Postdoctoral grant from the Swedish Research Council. D.F.S. was supported by the US  
489 Department of Energy Grant DE-SC00016240.

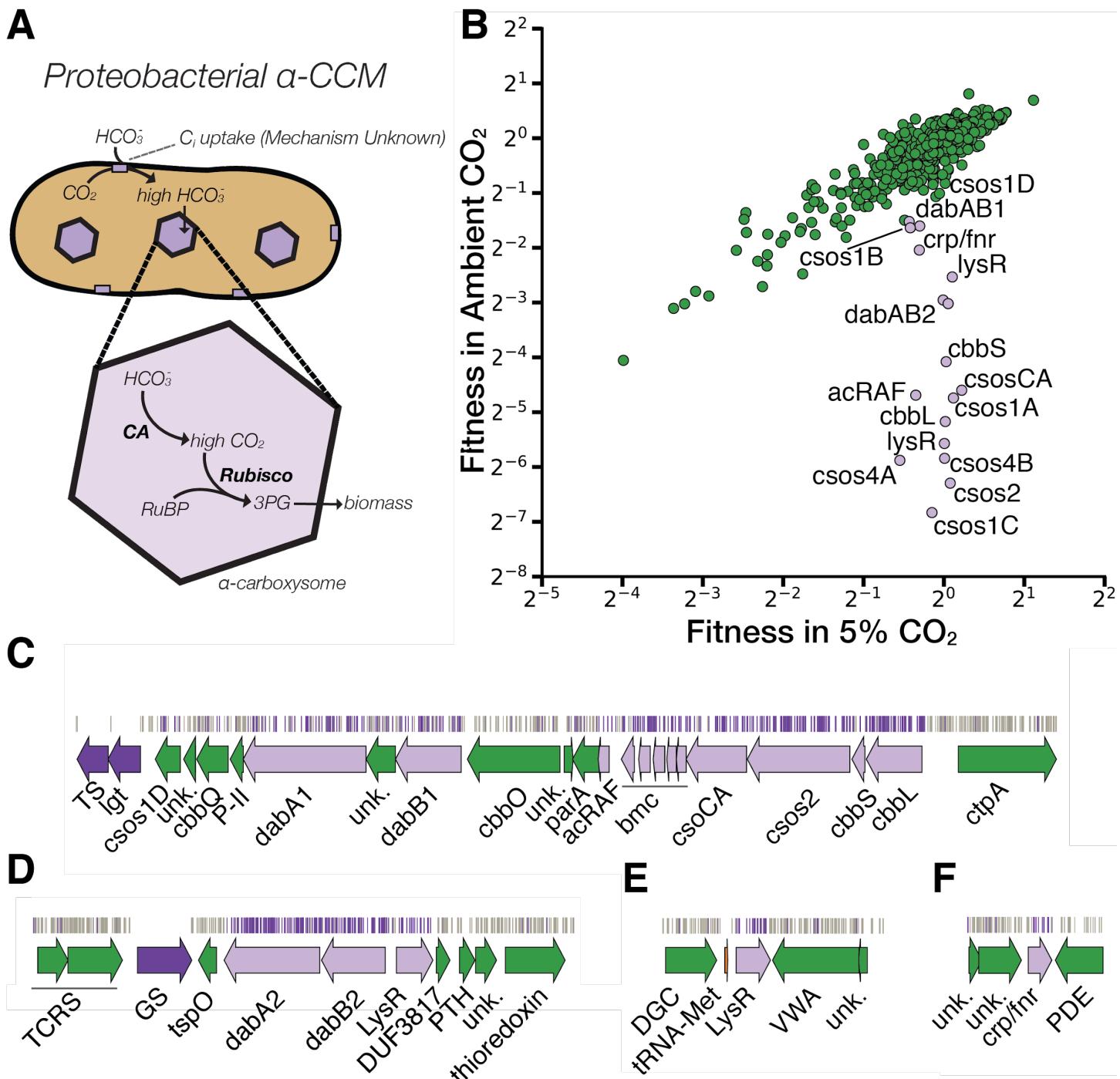
490

491 **Competing Interests**

492 UC Regents have filed a patent related to this work on which J.J.D., A.F.. and D.F.S. are inventors. D.F.S.  
493 is a co-founder of Scribe Therapeutics and a scientific advisory board member of Scribe Therapeutics and  
494 Mammoth Biosciences. All other authors declare no competing interests.



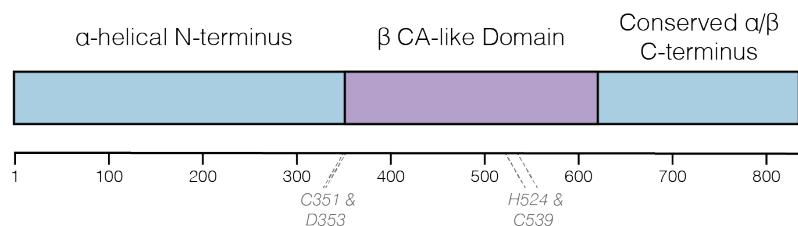
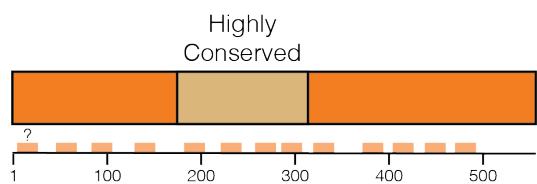
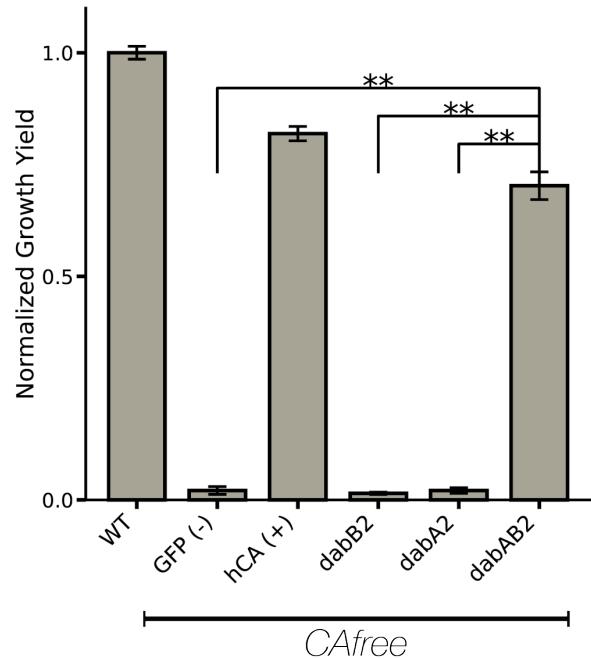
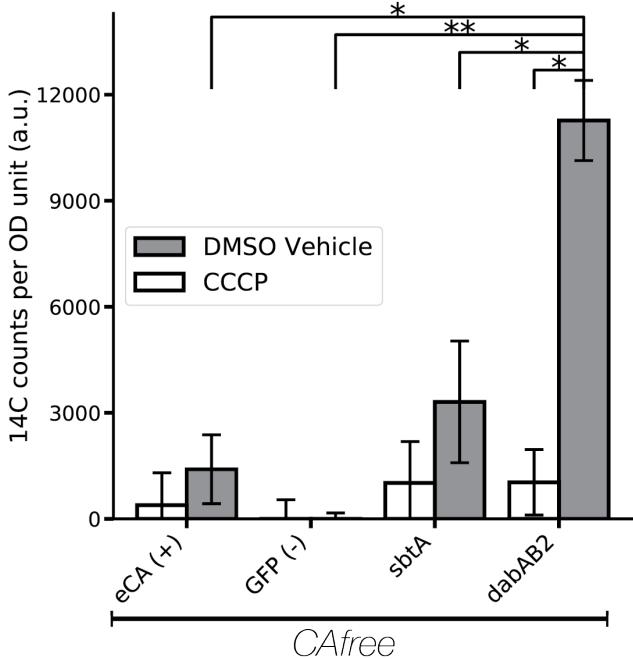
497 **Figure 1. Transposon mutagenesis reveals the essential gene set of a chemoautotrophic organism. A.**  
498 Schematic depicting the generation and screening of the RB-TnSeq library. Transposons were inserted into the *Hnea*  
499 genome by conjugation with an *E. coli* donor strain. The transposon contains a random 20 base pair barcode (yellow)  
500 and a kanamycin selection marker (green). Single colonies were selected for insertion in the presence of kanamycin at  
501 5% CO<sub>2</sub> and insertions were mapped by Illumina sequencing as described in the Methods. Subsequent screens were  
502 carried out as bulk competition assays and quantified by Illumina sequencing. **B.** Insertions and essential genes are  
503 well-distributed throughout the *Hnea* genome. The outer track (blue) is a histogram of the number of barcodes that  
504 were mapped to a 1 kb window. The inner track annotates essential genes in purple. The pie chart shows the  
505 percentages of the genome called essential (purple), ambiguous (orange), and nonessential (green). **C.**  
506 Representative essential genes and nonessential genes in the *Hnea* genome. The blue track indicates the presence of  
507 an insertion. Genes in purple were called essential and genes in green are nonessential. Genes labeled “unk.” are  
508 hypothetical proteins. The top operon contains 5 genes involved in glycolysis or the CBB cycle. The second operon  
509 contains genes encoding 30S and 50S subunits of the ribosome, the secY secretory channel, and an RNA polymerase  
510 subunit. The third operon contains genes involved in DNA replication. Acronyms: exopolyphosphatase (PP-ase),  
511 fructose-bisphosphate aldolase class II (fba), pyruvate kinase (pyk), phosphoglycerate kinase (pgk), type I  
512 glyceraldehyde-3-phosphate dehydrogenase (gapdh), transketolase (tkt), 30S ribosomal protein (30S), 50S ribosomal  
513 protein (50S), preprotein translocase subunit SecY (SecY), DNA-directed RNA polymerase subunit alpha (RNAP<sub>α</sub>),  
514 hypothetical protein (unk.), excinuclease ABC subunit UvrB (UvrB), chromosomal replication initiator protein dnaA  
515 (dnaA), DNA polymerase III subunit beta (DNAPIII $\beta$ ), DNA replication and repair protein recF (recF), DNA  
516 topoisomerase (ATP-hydrolyzing) subunit B (topoisomerase), lemA family protein (LemA).



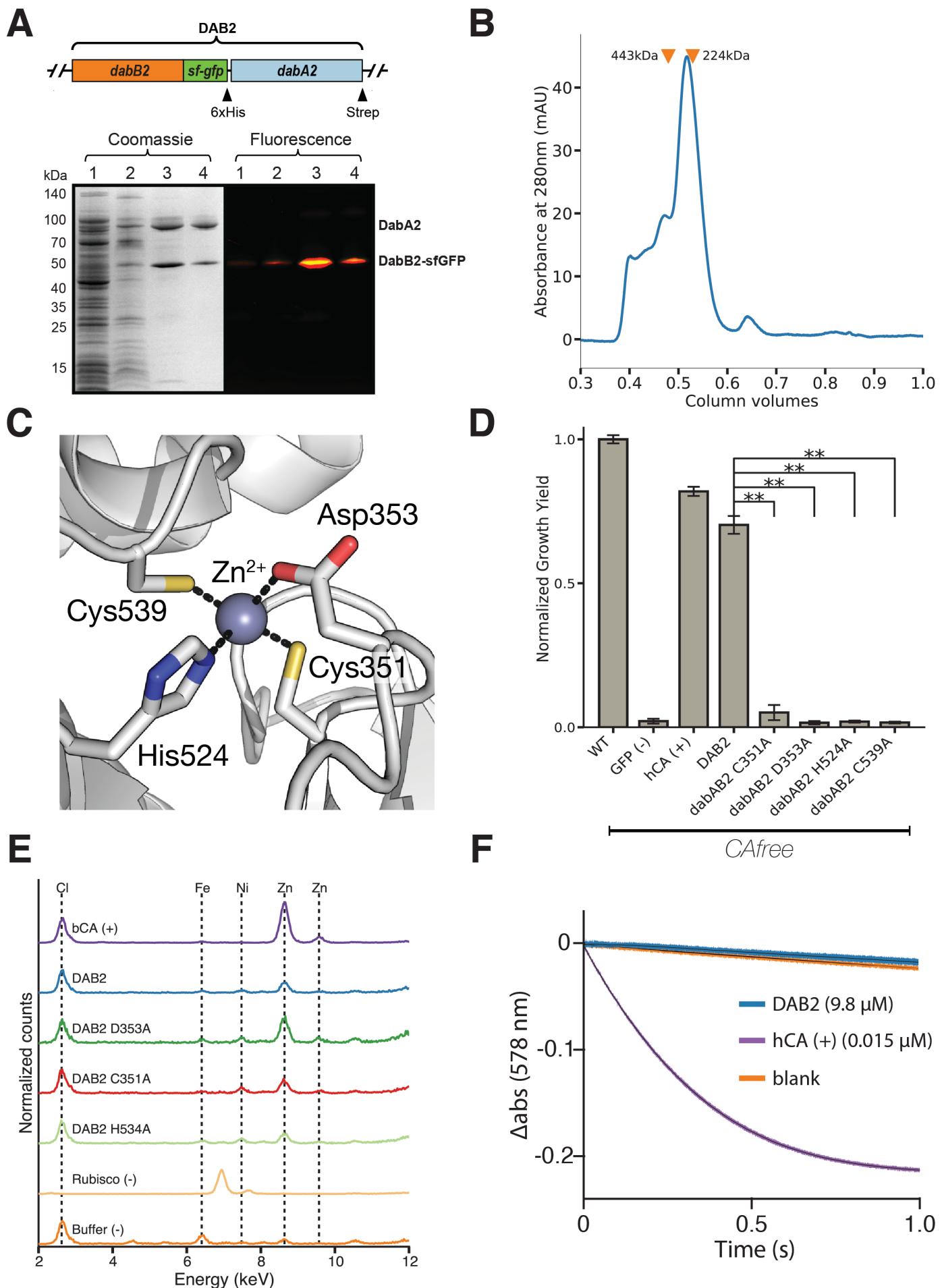
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530

**Figure 2. A systematic screen for high CO<sub>2</sub>-requiring mutants identifies genes putatively associated with the CCM.** **A.** Simplified model of the  $\alpha$ -CCM of chemotrophic proteobacteria. Inorganic carbon is concentrated via an unknown mechanism, producing a high cytosolic HCO<sub>3</sub><sup>-</sup> concentration. High cytosolic HCO<sub>3</sub><sup>-</sup> is converted into high carboxysomal CO<sub>2</sub> by CA, which is localized only to the carboxysome. **B.** Fitness effects of gene knockouts in 5% CO<sub>2</sub> as compared to ambient CO<sub>2</sub>. Data is from one of two replicates of the BarSeq - the second replicate gives consistent results. When the effect of single transposon insertions into a gene are mutually consistent, those effects are averaged to produce the gene-level fitness value plotted (Wetmore et al., 2015). We define HCR mutants as those displaying a twofold fitness defect in ambient CO<sub>2</sub> relative to 5% CO<sub>2</sub> (i.e. a fitness difference of 1 on the log<sub>2</sub> scale plotted). HCR genes are colored light purple. Panels **C-F** show regions of the *Hnea* genome containing genes annotated as HCR in panel A. Essential genes are in dark purple, HCR genes are in light purple, and other genes are in green. The top tracks show the presence of an insertion in that location. Insertions are colored light purple if they display a twofold fitness defect in ambient CO<sub>2</sub> relative to 5% CO<sub>2</sub>, otherwise they are colored grey. **C.** The gene cluster containing the carboxysome operon and a second CCM-associated operon annotated as in Figure 1C. This second operon contains

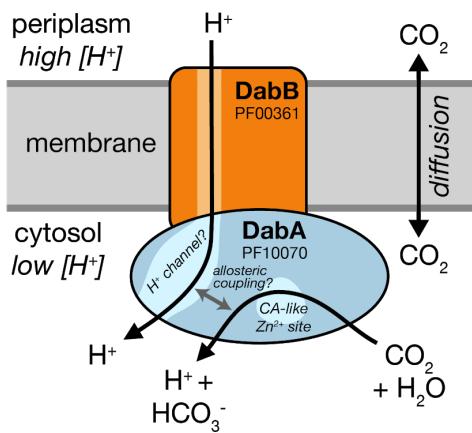
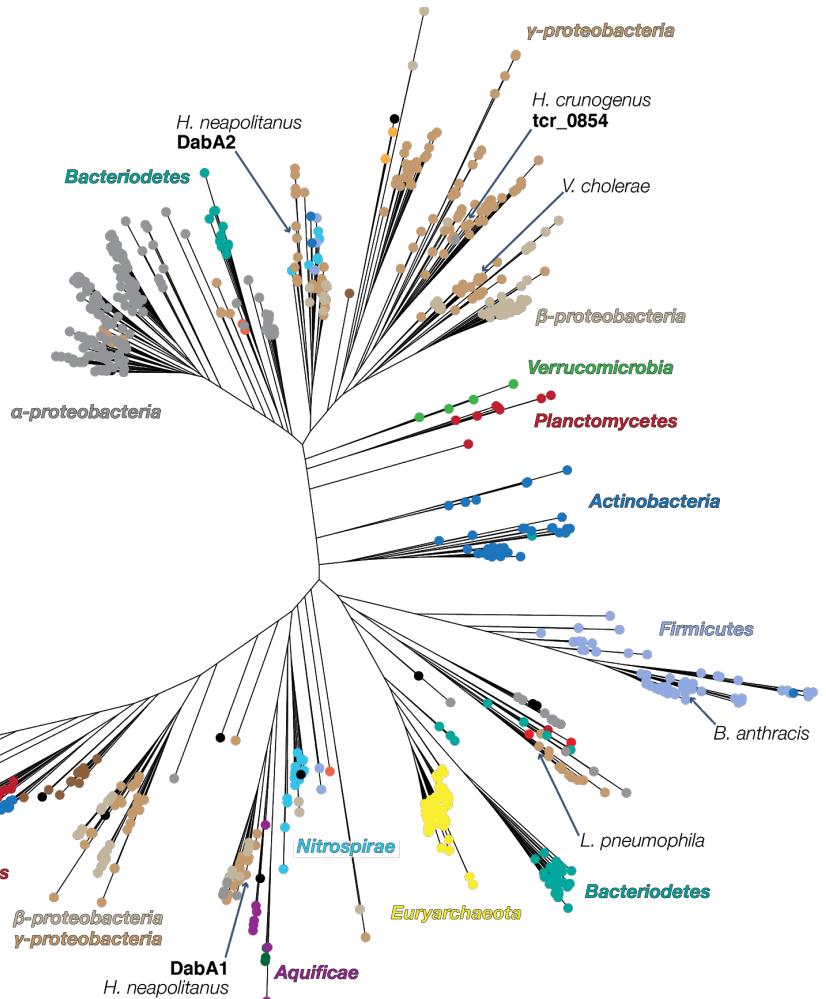
531 acRAF, a FormIC associated cbbOQ-type Rubisco activase and DAB1. **D**. The DAB2 operon and surrounding  
532 genomic context. **E**. The genomic context of a lysR-type transcriptional regulator that shows an HCR phenotype. **F** The  
533 genomic context of a crp/fnr-type transcriptional regulator that displays an HCR phenotype. Genes labeled “unk.” are  
534 hypothetical proteins. Acronyms: thymidylate synthase (TS), prolipoprotein diacylglycerol transferase (Igt), Rubisco  
535 activase Rubisco activase subunits (cbbOQ), nitrogen regulatory protein P-II (P-II), ParA family protein (parA),  
536 csos1CAB and csos4AB (bmc), copper-translocating P-type ATPase (ctpA), DNA-binding response regulator and two-  
537 component sensor histidine kinase (TCRS), glutamate--ammonia ligase (GS), tryptophan-rich sensory protein (tspO),  
538 DUF3817 domain-containing protein (DUF3817), aminoacyl-tRNA hydrolase (PTH), thioredoxin domain-containing  
539 protein (thioredoxin), sensor domain-containing diguanylate cyclase (DGC), methionine tRNA (tRNA-Met), VWA  
540 domain-containing protein (VWA), diguanylate phosphodiesterase (PDE).

**A***DabA2*  
827 Amino Acids*DabB2*  
551 Amino Acids**B****C**

**Figure 3. The DABs catalyze active transport of Ci energized by a cation gradient. A.** Diagrammatic representation of DabA2 and DabB2 based on bioinformatic annotation. DabA2 is an 827 amino acid protein with predicted homology to a type II β-CA enzyme. The four predicted active site residues (C351, D353, H524, C539) are marked on the primary amino acid sequence. DabB2 is a 551 amino acid protein with 12-13 transmembrane helices. There is a highly conserved region in the middle of the sequence. Predicted transmembrane helices are marked in light orange along the primary sequence. **B.** DAB2 was tested for ability to rescue growth of CAfree *E. coli* in ambient CO<sub>2</sub> conditions. The full operon (DabAB2) rescues growth as well as heterologous expression of the human carbonic anhydrase II (hCA), but rescue is contingent on the expression of both genes. Error bars represent standard deviations of 4 replicate cultures. **C.** CAfree *E. coli* were tested for Ci uptake using the silicone-oil centrifugation method. Expression of DabAB2 produced a large and statistically significant increase in <sup>14</sup>C uptake as compared to all controls. Moreover, treatment with the ionophore CCCP greatly reduces DabAB2-mediated <sup>14</sup>C uptake, suggesting that DabAB2 is coupled to a cation gradient. *canA* (eCA) was used as a control for a non-vectorial CA. *Synechococcus elongatus* PCC 7942 sbtA was used as a known Ci importer. GFP was used as a vector control. Error bars represent standard deviations of 3 technical replicates. In (B) and (C) “\*\*” denotes that the means are significantly different with P < 0.05 according to a two-tailed T-test. “\*\*\*” denotes P < 5X10<sup>-4</sup>.



559 **Figure 4. DabA contains a  $\beta$ -CA-like active site but is not constitutively active.** **A.** We purified the DabAB2  
560 complex from *E. coli* BL21(AI) cells using a purification construct in which DabB2 was C-terminally fused to sf-GFP  
561 and a 6xHis-tag and DabA2 was C-terminally fused to a Strep-tag. Progress in the purification was monitored using  
562 SDS-PAGE and gels were imaged for fluorescence (right view) before they were stained with coomassie (left view).  
563 Lane 1: clarified lysate; 2: solubilized membranes; 3: Ni resin eluent; 4: strep-tactin resin eluent. DabA2 and DabB2 co-  
564 purify as a single complex without any obvious interactors. **B.** Size-exclusion chromatography trace of His/Strep  
565 purified DabAB2 with retention volumes (orange arrows) and molecular weights (kDa) indicated for standard samples  
566 (apo ferritin, 443 kDa;  $\beta$ -amylase, 224 kDa). DabAB2 runs at an estimated mass of ~270 kDa, which must be an  
567 oligomer of DabA and DabB. Given the additional size contributed by the detergent-belt, a heterodimer is consistent  
568 with these data. **C.** Structural model of DabA2 active site based on the constitutive  $\beta$ -CA of *E. coli* (PDB 1I6P). Typical  
569  $\beta$ -CAs rely on two cysteine and one histidine residues to bind  $Zn^{2+}$ . A fourth residue - an aspartic acid - coordinates  
570  $Zn^{2+}$  in the structure but is thought to be displaced in order to enable catalysis (Cronk et al., 2006). **D.** Alanine mutants  
571 of the putative DabA2 active site residues (C351A, D353A, H524A, C539A) abrogate rescue of CAfree *E. coli*. “\*”  
572 denotes that means differ significantly with  $P < 0.05$  according to a two-tailed T-test, and “\*\*” denotes  $P < 5 \times 10^{-4}$ . Error  
573 bars represent standard deviations of four replicate cultures. **E.** Comparing to X-ray fluorescence of Bovine CA (bCA),  
574 DabAB2 is able to bind zinc as is expected based on the current model of type II  $\beta$ -CA activity. Active site point  
575 mutants retained their ability to bind zinc possibly because they still had three coordinating residues for the zinc, an  
576 amount that is sufficient in other carbonic anhydrases (Supuran, 2016). **F.** Purified DabAB2 does not display any  
577 obvious CA activity despite being present in 650-fold excess over the positive control (Human carbonic anhydrase II,  
578 hCA) in our assays. In (B) “\*\*” denotes that the means are significantly different with  $P < 5 \times 10^{-4}$  according to a two-  
579 tailed T-test.

**A****B**

**Figure 5. A model of the unidirectional energy-coupled CA activity of DAB complexes.** **A.** We propose that DabAB complexes couple the  $\beta$ -CA-like active site of DabA to a cation gradient across the cell membrane, thereby producing unidirectional hydration of  $\text{CO}_2$  to  $\text{HCO}_3^-$ . We draw this activity as being coupled to the  $\text{H}^+$  gradient (more generally, to the proton-motive force) for simplicity but our results are equally consistent with another cation gradient, e.g.  $\text{Na}^+$ . This model of energy-coupled CA activity is consistent with the DABs role as a  $\text{C}_i$  uptake system in the proteobacterial CCM since the CCM requires a high and, crucially, out-of-equilibrium  $\text{HCO}_3^-$  concentration in the cytosol in order for the carboxysomal CA to produce a high  $\text{CO}_2$  concentration near Rubisco. Because it appears that DabAB2 is not active as a purified complex, the protein must tightly couple the inflow of cations with  $\text{CO}_2$  hydration so that there is no “slippage.” Indeed, slippage - i.e., uncoupled CA activity - would be counterproductive from the perspective of the CCM (Mangan et al., 2016; Price and Badger, 1989a). Notably,  $\text{Zn}^{2+}$  binding by the active site aspartic acid of type II  $\beta$ -CAs (D353 in DabA2, Figure 4A) is thought to allosterically regulate activity (Cronk et al., 2006; Rowlett, 2010). This Asp-mediated activity switch could, therefore, provide a means for allosteric coupling of a  $\beta$ -CA active site to distal ion transport. **B.** Approximate maximum likelihood phylogenetic tree of DabA homologs associated with PF10070.9 (Methods). DabA homologs are found in > 15 prokaryotic clades, including archaea. *H. neapolitanus* DabA1 and DabA2 represent two different groupings that are commonly found in proteobacteria. The *tcr\_0854* gene of *H. crunogenus* is more closely related to DabA2 than DabA1 (Mangiapia et al., 2017). Inspecting the tree reveals several likely incidents of horizontal transfer, e.g. between proteobacteria and Firmicutes, Nitrospirae and Actinobacteria. Moreover, the genomes of several known pathogens contain a high-confidence DabA homolog, including *B. anthracis*, *L. pneumophila*, *V. cholerae*. Detailed annotations are given in Figure 5 S3.

**References**

- 603 Aguilera J, Van Dijken JP, De Winde JH, Pronk JT. 2005. Carbonic anhydrase (Nce103p): an essential  
604 biosynthetic enzyme for growth of *Saccharomyces cerevisiae* at atmospheric carbon dioxide pressure.  
605 *Biochem J* **391**:311–316. doi:10.1042/BJ20050556
- 606 Aigner H, Wilson RH, Bracher A, Calisse L, Bhat JY, Hartl FU, Hayer-Hartl M. 2017. Plant RuBisCo  
607 assembly in *E. coli* with five chloroplast chaperones including BSD2. *Science* **358**:1272–1278.  
608 doi:10.1126/science.aap9221
- 609 Antonovsky N, Gleizer S, Noor E, Zohar Y, Herz E, Barenholz U, Zelcbuch L, Amram S, Wides A, Tepper N,  
610 Davidi D, Bar-On Y, Bareia T, Wernick DG, Shani I, Malitsky S, Jona G, Bar-Even A, Milo R. 2016.  
611 Sugar Synthesis from CO<sub>2</sub> in *Escherichia coli*. *Cell* **166**:115–125. doi:10.1016/j.cell.2016.05.064
- 612 Artier J, Holland SC, Miller NT, Zhang M, Burnap RL. 2018. Synthetic DNA system for structure-function  
613 studies of the high affinity CO<sub>2</sub> uptake NDH-13 protein complex in cyanobacteria. *Biochim Biophys  
614 Acta Bioenerg*. doi:10.1016/j.bbabi.2018.06.015
- 615 Axen SD, Erbilgin O, Kerfeld CA. 2014. A taxonomy of bacterial microcompartment loci constructed by a  
616 novel scoring method. *PLoS Comput Biol* **10**:e1003898. doi:10.1371/journal.pcbi.1003898
- 617 Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko K a., Tomita M, Wanner BL, Mori H.  
618 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio  
619 collection. *Mol Syst Biol* **2**:2006.0008. doi:10.1038/msb4100050
- 620 Badger MR, Price GD. 2003. CO<sub>2</sub> concentrating mechanisms in cyanobacteria: molecular components,  
621 their diversity and evolution. *J Exp Bot* **54**:609–622.
- 622 Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, Milo R. 2011. The Moderately Efficient  
623 Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry*.  
624 doi:10.1021/bi2002289
- 625 Bathellier C, Tcherkez G, Lorimer GH, Farquhar GD. 2018. Rubisco is not really so bad. *Plant Cell Environ*  
626 **41**:705–716. doi:10.1111/pce.13149
- 627 Battchikova N, Eisenhut M, Aro EM. 2011. Cyanobacterial NDH-1 complexes: Novel insights and remaining  
628 puzzles. *Biochimica et Biophysica Acta - Bioenergetics* **1807**:935–944.  
629 doi:10.1016/j.bbabi.2010.10.017
- 630 Bauwe H, Hagemann M, Fernie AR. 2010. Photorespiration: players, partners and origin. *Trends Plant Sci*  
631 **15**:330–336. doi:10.1016/j.tplants.2010.03.006
- 632 Birungi M, Folea M, Battchikova N, Xu M, Mi H, Ogawa T, Aro EM, Boekema EJ. 2010. Possibilities of  
633 subunit localization with fluorescent protein tags and electron microscopy exemplified by a  
634 cyanobacterial NDH-1 study. *Biochimica et Biophysica Acta - Bioenergetics* **1797**:1681–1686.  
635 doi:10.1016/j.bbabi.2010.06.004
- 636 Bonacci W, Teng PK, Afonso B, Niederholtmeyer H, Grob P, Silver P a., Savage DF. 2012. Modularity of a  
637 carbon-fixing protein organelle. *Proc Natl Acad Sci U S A* **109**:478–483. doi:10.1073/pnas.1108557109
- 638 Buchanan BB, Grussem W, Jones RL. 2015. Biochemistry and Molecular Biology of Plants. Wiley.
- 639 Cai F, Menon BB, Cannon GC, Curry KJ, Shively JM, Heinhorst S. 2009. The pentameric vertex proteins  
640 are necessary for the icosahedral carboxysome shell to function as a CO<sub>2</sub> leakage barrier. *PLoS One*  
641 **4**:e7521. doi:10.1371/journal.pone.0007521
- 642 Cannon GC, Bradburne CE, Aldrich HC, Baker SH, Heinhorst S, Shively JM. 2001. Microcompartments in  
643 prokaryotes: carboxysomes and related polyhedra. *Appl Environ Microbiol* **67**:5351–5361.  
644 doi:10.1128/AEM.67.12.5351-5361.2001
- 645 Cronk JD, Endrizzi J a., Cronk MR, O’neill JW, Zhang KY. 2001. Crystal structure of *E. coli* beta-carbonic  
646 anhydrase, an enzyme with an unusual pH-dependent activity. *Protein Sci* **10**:911–922.  
647 doi:10.1110/ps.46301
- 648 Cronk JD, Rowlett RS, Zhang KYJ, Tu C, Endrizzi JA, Lee J, Gareiss PC, Preiss JR. 2006. Identification of  
649 a novel noncatalytic bicarbonate binding site in eubacterial beta-carbonic anhydrase. *Biochemistry*  
650 **45**:4351–4361. doi:10.1021/bi052272q
- 651 Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, Friedland GD, Huang KH, Keller K,  
652 Novichkov PS, Dubchak IL, Alm EJ, Arkin AP. 2010. MicrobesOnline: an integrated portal for  
653 comparative and functional genomics. *Nucleic Acids Res* **38**:D396–400. doi:10.1093/nar/gkp919
- 654 Dobrinski KP, Longo DL, Scott KM. 2005. The Carbon-Concentrating Mechanism of the Hydrothermal Vent  
655 Chemolithoautotroph *Thiomicrospira crunogena*. *J Bacteriol* **187**:5761–5766.

- 656 doi:10.1128/JB.187.16.5761-5766.2005
- 657 Dou Z, Heinhorst S, Williams EB, Murin CD, Shively JM, Cannon GC. 2008. CO<sub>2</sub> fixation kinetics of  
658 Halothiobacillus neapolitanus mutant carboxysomes lacking carbonic anhydrase suggest the shell acts  
659 as a diffusional barrier for CO<sub>2</sub>. *J Biol Chem* **283**:10377–10384. doi:10.1074/jbc.M709285200
- 660 Du J, Förster B, Rourke L, Howitt SM, Price GD. 2014. Characterisation of Cyanobacterial Bicarbonate  
661 Transporters in *E. coli* Shows that SbtA Homologs Are Functional in This Heterologous Expression  
662 System. *PLoS One* **9**:e115905. doi:10.1371/journal.pone.0115905
- 663 Fang Y, Huang F, Faulkner M, Jiang Q, Dykes GF, Yang M, Liu L-N. 2018. Engineering and Modulating  
664 Functional Cyanobacterial CO<sub>2</sub>-Fixing Organelles. *Front Plant Sci* **9**:739. doi:10.3389/fpls.2018.00739
- 665 Flamholz A, Prywes N, Moran U, Davidi D, Bar-On Y, Oltrogge L, Savage D, Milo R. 2018. Revisiting  
666 tradeoffs in Rubisco kinetic parameters. *bioRxiv*. doi:10.1101/470021
- 667 Han X, Sun N, Xu M, Mi H. 2017. Co-ordination of NDH and Cup proteins in CO<sub>2</sub> uptake in cyanobacterium  
668 *Synechocystis* sp. PCC 6803. *J Exp Bot* **68**:3869–3877. doi:10.1093/jxb/erx129
- 669 Heinhorst S, Cannon GC, Shively JM. 2006. Carboxysomes and Carboxysome-like Inclusions In: Shively  
670 JM, editor. Complex Intracellular Structures in Prokaryotes. Berlin, Heidelberg: Springer Berlin  
671 Heidelberg. pp. 141–165. doi:10.1007/7171\_023
- 672 Holthuijzen YA, van Dissel-Emiliani FFM, Kuenen JG, Konings WN. 1987. Energetic aspects of CO<sub>2</sub> uptake  
673 in *Thiobacillus neapolitanus*. *Arch Microbiol* **147**:285–290. doi:10.1007/BF00463489
- 674 Hopkinson BM, Young JN, Tansik a. L, Binder BJ. 2014. The Minimal CO<sub>2</sub>-Concentrating Mechanism of  
675 Prochlorococcus spp. MED4 Is Effective and Efficient. *Plant Physiol* **166**:2205–2217.  
676 doi:10.1104/pp.114.247049
- 677 Ippolito JA, Nair SK, Alexander RS, Kiefer LL, Fierke CA, Christianson DW. 1995. Structure of His94-->Asp  
678 carbonic anhydrase II in a new crystalline form reveals a partially occupied zinc binding site. *Protein*  
679 **8**:975–980.
- 680 Jorda J, Lopez D, Wheatley NM, Yeates TO. 2013. Using comparative genomics to uncover new kinds of  
681 protein-based metabolic organelles in bacteria. *Protein Sci* **22**:179–195. doi:10.1002/pro.2196
- 682 Kaplan A, Badger MR, Berry JA. 1980. Photosynthesis and the intracellular inorganic carbon pool in the  
683 bluegreen alga *Anabaena variabilis*: Response to external CO<sub>2</sub> concentration. *Planta* **149**:219–226.  
684 doi:10.1007/BF00384557
- 685 Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein  
686 modeling, prediction and analysis. *Nat Protoc* **10**:845–858. doi:10.1038/nprot.2015.053
- 687 Khalifah RG. 1971. The Carbon Dioxide Hydration Activity of Carbonic Anhydrase. *J Biol Chem* **246**:2561–  
688 2573.
- 689 Krishnamurthy VM, Kaufman GK, Urbach AR, Gitlin I, Gudiksen KL, Weibel DB, Whitesides GM. 2008.  
690 Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein-  
691 ligand binding. *Chem Rev* **108**:946–1051.
- 692 Krulwich TA, Hicks DB, Ito M. 2009. Cation/proton antiporter complements of bacteria: why so large and  
693 diverse? *Mol Microbiol* **74**:257–260. doi:10.1111/j.1365-2958.2009.06842.x
- 694 Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of  
695 phylogenetic and other trees. *Nucleic Acids Res* **44**:W242–5. doi:10.1093/nar/gkw290
- 696 Lin MT, Occhialini A, Andralojc PJ, Parry MAJ, Hanson MR. 2014. A faster Rubisco with potential to  
697 increase photosynthesis in crops. *Nature* **513**:547–550. doi:10.1038/nature13776
- 698 Long BM, Hee WY, Sharwood RE, Rae BD, Kaines S, Lim Y-L, Nguyen ND, Massey B, Bala S, von  
699 Caemmerer S, Badger MR, Price GD. 2018. Carboxysome encapsulation of the CO<sub>2</sub>-fixing enzyme  
700 Rubisco in tobacco chloroplasts. *Nat Commun* **9**:3570. doi:10.1038/s41467-018-06044-0
- 701 Long BM, Rae BD, Rolland V, Förster B, Price GD. 2016. Cyanobacterial CO<sub>2</sub>-concentrating mechanism  
702 components: function and prospects for plant metabolic engineering. *Curr Opin Plant Biol* **31**:1–8.  
703 doi:10.1016/j.pbi.2016.03.002
- 704 Lutz R, Bujard H. 1997. Independent and tight regulation of transcriptional units in *Escherichia coli* via the  
705 LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* **25**:1203–1210.  
706 doi:10.1093/nar/25.6.1203
- 707 Mackinder LCM, Meyer MT, Mettler-Altmann T, Chen VK, Mitchell MC, Caspari O, Freeman Rosenzweig  
708 ES, Pallesen L, Reeves G, Itakura A, Roth R, Sommer F, Geimer S, Mühlhaus T, Schröder M,  
709 Goodenough U, Stitt M, Griffiths H, Jonikas MC. 2016. A repeat protein links Rubisco to form the  
710 eukaryotic carbon-concentrating organelle. *Proc Natl Acad Sci U S A* **113**:5958–5963.

- 711 doi:10.1073/pnas.1522866113  
712 Maeda S-I, Badger MR, Price GD. 2002. Novel gene products associated with NdhD3/D4-containing NDH-1  
713 complexes are involved in photosynthetic CO<sub>2</sub> hydration in the cyanobacterium, *Synechococcus* sp.  
714 PCC7942. *Mol Microbiol* **43**:425–435.
- 715 Mangan NM, Flamholz A, Hood RD, Milo R, Savage DF. 2016. pH determines the energetic efficiency of the  
716 cyanobacterial CO<sub>2</sub> concentrating mechanism. *Proc Natl Acad Sci U S A* **113**:E5354–62.  
717 doi:10.1073/pnas.1525145113
- 718 Mangiapia M, USF MCB4404L, Brown T-RW, Chaput D, Haller E, Harmer TL, Hashemy Z, Keeley R,  
719 Leonard J, Mancera P, Nicholson D, Stevens S, Wanjigi P, Zabinski T, Pan C, Scott KM. 2017.  
720 Proteomic and mutant analysis of the CO<sub>2</sub> concentrating mechanism of hydrothermal vent  
721 chemolithoautotroph *Thiomicrospira crunogena*. *J Bacteriol*. doi:10.1128/JB.00871-16
- 722 Marcus Y, Schwarz R, Friedberg D, Kaplan A. 1986. High CO(2) Requiring Mutant of *Anacystis nidulans*  
723 R(2). *Plant Physiol* **82**:610–612.
- 724 Marreiros BC, Batista AP, Duarte AMS, Pereira MM. 2013. A missing link between complex I and group 4  
725 membrane-bound [NiFe] hydrogenases. *Biochim Biophys Acta* **1827**:198–209.  
726 doi:10.1016/j.bbabi.2012.09.012
- 727 McGrath JM, Long SP. 2014. Can the cyanobacterial carbon-concentrating mechanism increase  
728 photosynthesis in crop species? A theoretical analysis. *Plant Physiol* **164**:2247–2261.  
729 doi:10.1104/pp.113.232611
- 730 Merlin C, Masters M. 2003. Why is carbonic anhydrase essential to *Escherichia coli*? *J Bacteriol* **185**.  
731 doi:10.1128/JB.185.21.6415
- 732 Mueller-Cajar O. 2017. The Diverse AAA+ Machines that Repair Inhibited Rubisco Active Sites. *Front Mol*  
733 *Biosci* **4**:31. doi:10.3389/fmolb.2017.00031
- 734 Murakami H, Marellich GP, Grubb JH, Kyle JW, Sly WS. 1987. Cloning, expression, and sequence  
735 homologies of cDNA for human carbonic anhydrase II. *Genomics* **1**:159–166.
- 736 Newby ZER, O'Connell JD 3rd, Gruswitz F, Hays FA, Harries WEC, Harwood IM, Ho JD, Lee JK, Savage  
737 DF, Miercke LJW, Stroud RM. 2009. A general protocol for the crystallization of membrane proteins for  
738 X-ray structural investigation. *Nat Protoc* **4**:619–637. doi:10.1038/nprot.2009.27
- 739 Occhialini A, Lin MT, Andralojc PJ, Hanson MR, Parry MAJ. 2016. Transgenic tobacco plants with improved  
740 cyanobacterial Rubisco expression but no extra assembly factors grow at near wild-type rates if  
741 provided with elevated CO<sub>2</sub>. *Plant J* **85**:148–160. doi:10.1111/tpj.13098
- 742 Ogawa T, Kaneda T, Omata T. 1987. A Mutant of *Synechococcus* PCC7942 Incapable of Adapting to Low  
743 CO<sub>2</sub> Concentration. *Plant Physiol* **84**:711–715. doi:10.1104/pp.84.3.711
- 744 Pédelacq J-D, Cabantous S, Tran T, Terwilliger TC, Waldo GS. 2006. Engineering and characterization of a  
745 superfolder green fluorescent protein. *Nat Biotechnol* **24**:79–88. doi:10.1038/nbt1172
- 746 Price GD. 2011. Inorganic carbon transporters of the cyanobacterial CO<sub>2</sub> concentrating mechanism.  
747 *Photosynth Res* **109**:47–57. doi:10.1007/s11120-010-9608-y
- 748 Price GD, Badger MR. 1989a. Expression of Human Carbonic Anhydrase in the Cyanobacterium  
749 *Synechococcus* PCC7942 Creates a High CO<sub>2</sub>-Requiring Phenotype Evidence for a Central Role for  
750 Carboxysomes in the CO<sub>2</sub> Concentrating Mechanism. *Plant Physiol* **91**:505–513.
- 751 Price GD, Badger MR. 1989b. Isolation and characterization of high CO<sub>2</sub>-requiring-mutants of the  
752 cyanobacterium *Synechococcus* PCC7942: two phenotypes that accumulate inorganic carbon but are  
753 apparently unable to generate CO<sub>2</sub> within the carboxysome. *Plant Physiol* **91**:514–525.
- 754 Price GD, Badger MR, von Caemmerer S. 2011. The prospect of using cyanobacterial bicarbonate  
755 transporters to improve leaf photosynthesis in C3 crop plants. *Plant Physiol* **155**:20–26.  
756 doi:10.1104/pp.110.164681
- 757 Rae BD, Long BM, Badger MR, Price GD. 2013. Functions, compositions, and evolution of the two types of  
758 carboxysomes: polyhedral microcompartments that facilitate CO<sub>2</sub> fixation in cyanobacteria and some  
759 proteobacteria. *Microbiol Mol Biol Rev* **77**:357–379. doi:10.1128/MMBR.00061-12
- 760 Raven JA, Beardall J, Sánchez-Baracaldo P. 2017. The possible evolution, and future, of CO<sub>2</sub>-  
761 concentrating mechanisms. *J Exp Bot*. doi:10.1093/jxb/erx110
- 762 Reinhold L, Kosloff R, Kaplan A. 1991. A model for inorganic carbon fluxes and photosynthesis in  
763 cyanobacterial carboxysomes. *Can J Bot* **69**:984–988. doi:10.1139/b91-126
- 764 Roberts EW, Cai F, Kerfeld CA, Cannon GC, Heinhorst S. 2012. Isolation and characterization of the  
765 Prochlorococcus carboxysome reveal the presence of the novel shell protein CsoS1D. *J Bacteriol*

- 766                   **194**:787–795. doi:10.1128/JB.06444-11  
767 Robertson LA, Kuenen JG. 2006. The Genus *Thiobacillus* In: Dworkin M, Falkow S, Rosenberg E, Schleifer  
768 K-H, Stackebrandt E, editors. *The Prokaryotes: Volume 5: Proteobacteria: Alpha and Beta Subclasses*.  
769 New York, NY: Springer New York. pp. 812–827. doi:10.1007/0-387-30745-1\_37  
770 Rowlett RS. 2010. Structure and catalytic mechanism of the  $\beta$ -carbonic anhydrases. *Biochimica et  
771 Biophysica Acta (BBA) - Proteins and Proteomics* **1804**:362–373. doi:10.1016/j.bbapap.2009.08.002  
772 Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and  
773 function prediction. *Nat Protoc* **5**:725–738. doi:10.1038/nprot.2010.5  
774 Rubin BE, Wetmore KM, Price MN, Diamond S, Shultzaberger RK, Lowe LC, Curtin G, Arkin AP,  
775 Deutschbauer A, Golden SS. 2015. The essential gene set of a photosynthetic organism. *Proceedings  
776 of the National Academy of Sciences* 201519220. doi:10.1073/pnas.1519220112  
777 Savir Y, Noor E, Milo R, Tlusty T. 2010. Cross-species analysis traces adaptation of Rubisco toward  
778 optimality in a low-dimensional landscape. *Proc Natl Acad Sci U S A* **107**:3475–3480.  
779 doi:10.1073/pnas.0911663107  
780 Scott KM, Williams J, Porter CMB, Russel S, Harmer TL, Paul JH, Antonen KM, Bridges MK, Camper GJ,  
781 Campla CK, Casella LG, Chase E, Conrad JW, Cruz MC, Dunlap DS, Duran L, Fahsbender EM,  
782 Goldsmith DB, Keeley RF, Kondoff MR, Kussy BI, Lane MK, Lawler S, Leigh BA, Lewis C, Lostal LM,  
783 Marking D, Mancera PA, McClenahan EC, McIntyre EA, Mine JA, Modi S, Moore BD, Morgan WA,  
784 Nelson KM, Nguyen KN, Ogburn N, Parrino DG, Pedapudi AD, Pelham RP, Preece AM, Rampersad  
785 EA, Richardson JC, Rodgers CM, Schaffer BL, Sheridan NE, Solone MR, Staley ZR, Tabuchi M, Waide  
786 RJ, Wanjugi PW, Young S, Clum A, Daum C, Huntemann M, Ivanova N, Kyriides N, Mikhailova N,  
787 Palaniappan K, Pillay M, Reddy TBK, Shapiro N, Stamatis D, Varghese N, Woyke T, Boden R,  
788 Freyermuth SK, Kerfeld CA. 2018. Genomes of ubiquitous marine and hypersaline Hydrogenovibrio,  
789 *Thiomicrosrhabdus* and *Thiomicrospira* spp. encode a diversity of mechanisms to sustain  
790 chemolithoautotrophy in heterogeneous environments. *Environ Microbiol* **20**:2686–2708.  
791 doi:10.1111/1462-2920.14090  
792 Shibata M, Katoh H, Sonoda M, Ohkawa H, Shimoyama M, Fukuzawa H, Kaplan A, Ogawa T. 2002a.  
793 Genes essential to sodium-dependent bicarbonate transport in cyanobacteria: function and  
794 phylogenetic analysis. *J Biol Chem* **277**:18658–18664. doi:10.1074/jbc.M112468200  
795 Shibata M, Ohkawa H, Kaneko T, Fukuzawa H, Tabata S, Kaplan A, Ogawa T. 2001. Distinct constitutive  
796 and low-CO<sub>2</sub>-induced CO<sub>2</sub> uptake systems in cyanobacteria: genes involved and their phylogenetic  
797 relationship with homologous genes in other organisms. *Proc Natl Acad Sci U S A* **98**:11789–11794.  
798 doi:10.1073/pnas.191258298  
799 Shibata M, Ohkawa H, Katoh H, Shimoyama M, Ogawa T. 2002b. Two CO<sub>2</sub> uptake systems in  
800 cyanobacteria: four systems for inorganic carbon acquisition in *Synechocystis* sp. strain PCC6803.  
801 *Funct Plant Biol* **29**:123–129. doi:10.1071/PP01188  
802 Shih PM, Occhialini A, Cameron JC, Andralojc PJ, Parry MAJ, Kerfeld CA. 2016. Biochemical  
803 characterization of predicted Precambrian RuBisCO. *Nat Commun* **7**:10382.  
804 doi:10.1038/ncomms10382  
805 Sievers F, Higgins DG. 2018. Clustal Omega for making accurate alignments of many protein sequences.  
806 *Protein Sci* **27**:135–145. doi:10.1002/pro.3290  
807 Slabinski L, Jaroszewski L, Rychlewski L, Wilson IA, Lesley SA, Godzik A. 2007. XtalPred: a web server for  
808 prediction of protein crystallizability. *Bioinformatics* **23**:3403–3405. doi:10.1093/bioinformatics/btm477  
809 Supuran CT. 2016. Structure and function of carbonic anhydrases. *Biochem J* **473**:2023–2032.  
810 doi:10.1042/BCJ20160115  
811 Supuran CT. 2008. Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. *Nat  
812 Rev Drug Discov* **7**:168–181. doi:10.1038/nrd2467  
813 Tabita FR, Hanson TE, Satagopan S, Witte BH, Kreel NE. 2008. Phylogenetic and evolutionary  
814 relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse  
815 molecular forms. *Philos Trans R Soc Lond B Biol Sci* **363**:2629–2640. doi:10.1098/rstb.2008.0023  
816 Tcherkez G. 2016. The mechanism of Rubisco-catalysed oxygenation. *Plant Cell Environ* **39**:983–997.  
817 Tcherkez GGB, Farquhar GD, Andrews TJ. 2006. Despite slow catalysis and confused substrate specificity,  
818 all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc Natl Acad Sci U S A*.  
819 doi:10.1073/pnas.0600605103  
820 Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, Blow MJ, Bristow J, Butland G, Arkin

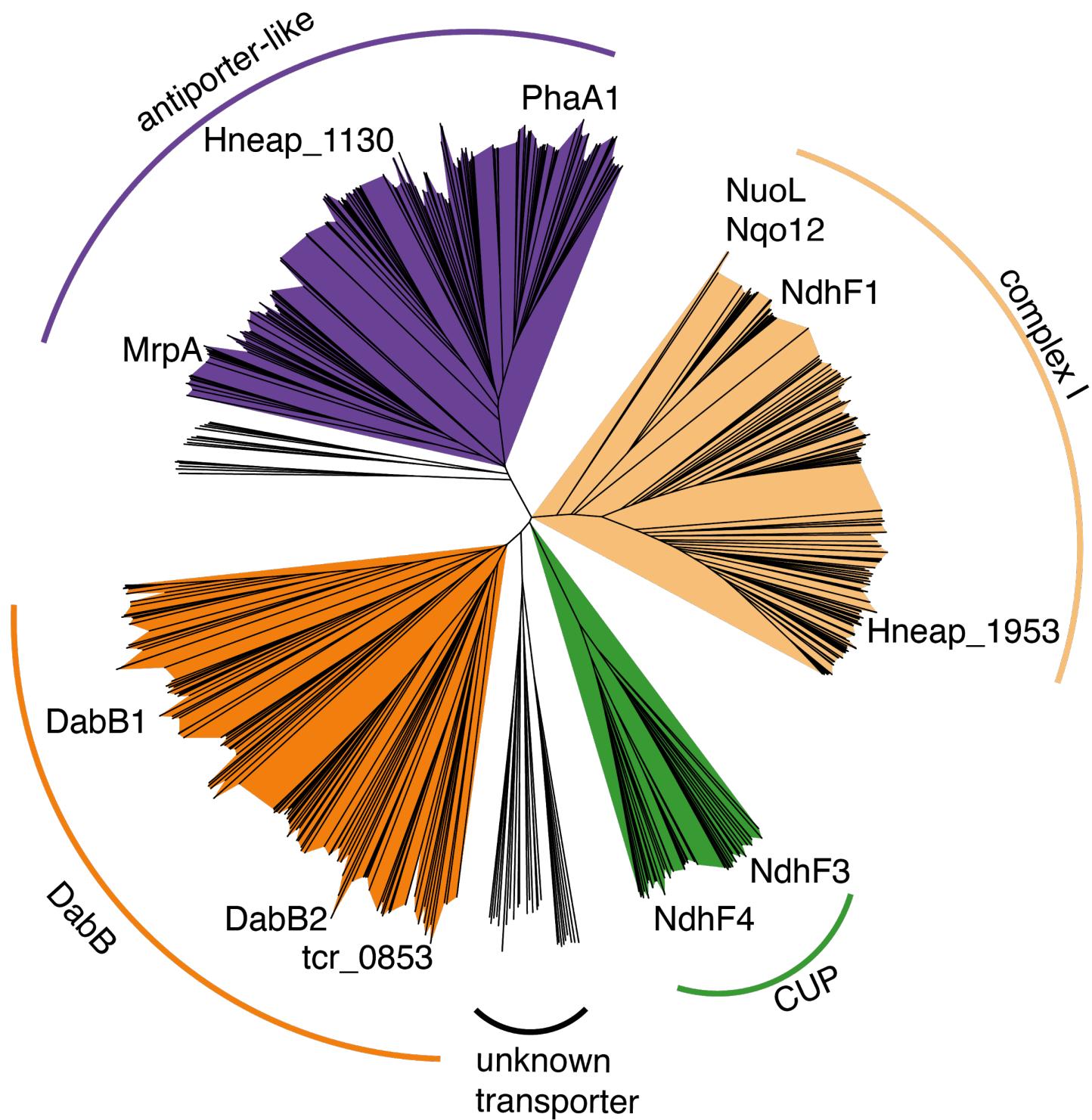
- 821 AP, Deutschbauer A. 2015. Rapid quantification of mutant fitness in diverse bacteria by sequencing  
822 randomly bar-coded transposons. *MBio* **6**:e00306–15. doi:10.1128/mBio.00306-15
- 823 Wheatley NM, Sundberg CD, Gidaniyan SD, Cascio D, Yeates TO. 2014. Structure and identification of a  
824 pterin dehydratase-like protein as a ribulose-bisphosphate carboxylase/oxygenase (RuBisCO)  
825 assembly factor in the  $\alpha$ -carboxysome. *J Biol Chem* **289**:7973–7981. doi:10.1074/jbc.M113.531236
- 826 Whitehead L, Long BM, Price GD, Badger MR. 2014. Comparing the in Vivo Function of  $\alpha$ -Carboxysomes  
827 and  $\beta$ -Carboxysomes in Two Model Cyanobacteria. *Plant Physiol* **165**:398–411.  
828 doi:10.1104/pp.114.237941

829

830  
831  
832  
833  
834

## Figure Supplements

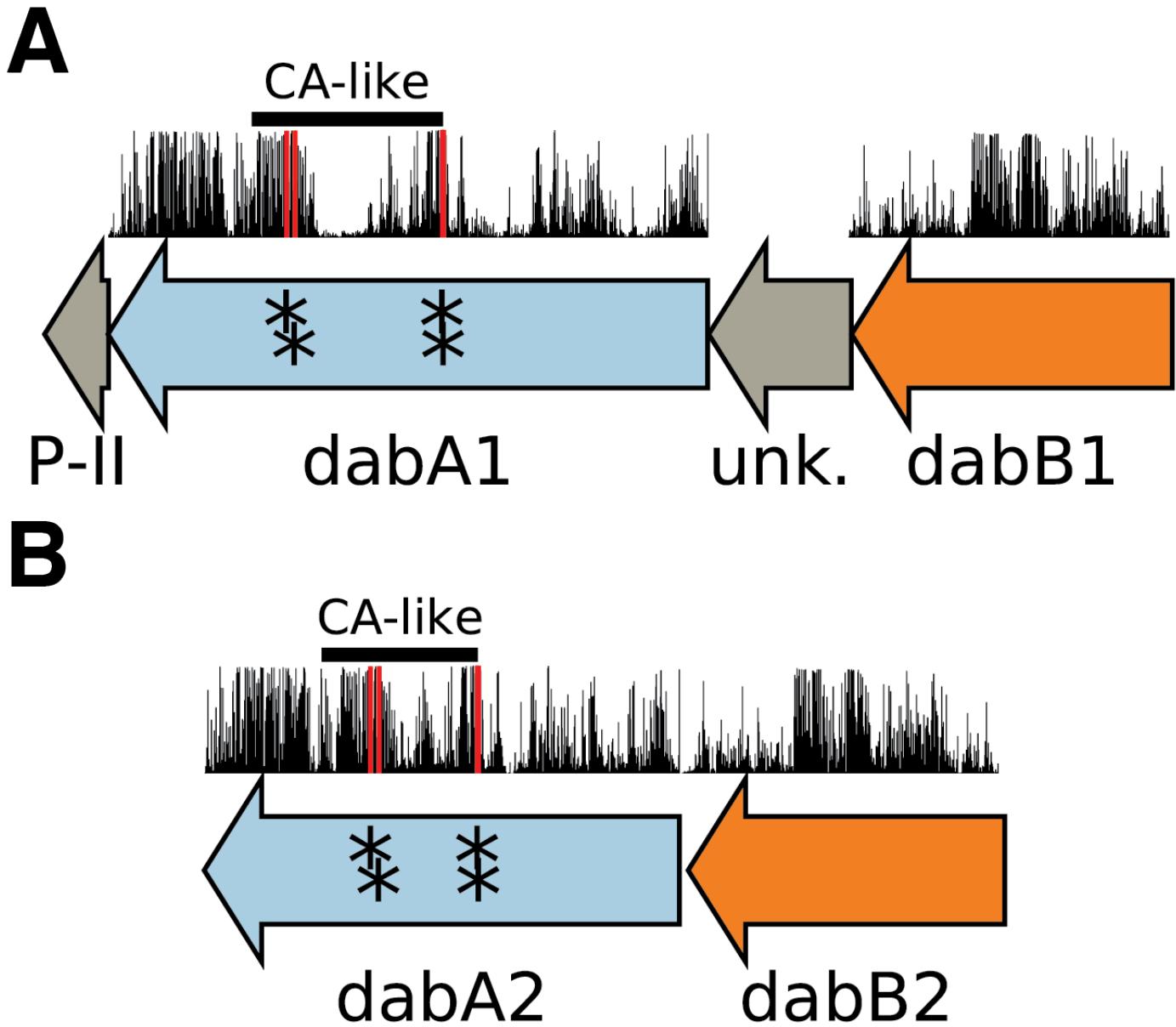
- Figure 1 Supplemental File 1. Transposon insertion information and essentiality determination by gene.  
Figure 2 Supplemental File 1. Fitness effects and HCR phenotype by gene.  
Figure 3 Supplemental File 1. Genes used to generate figure 3S1.



835  
836  
837  
838  
839  
840

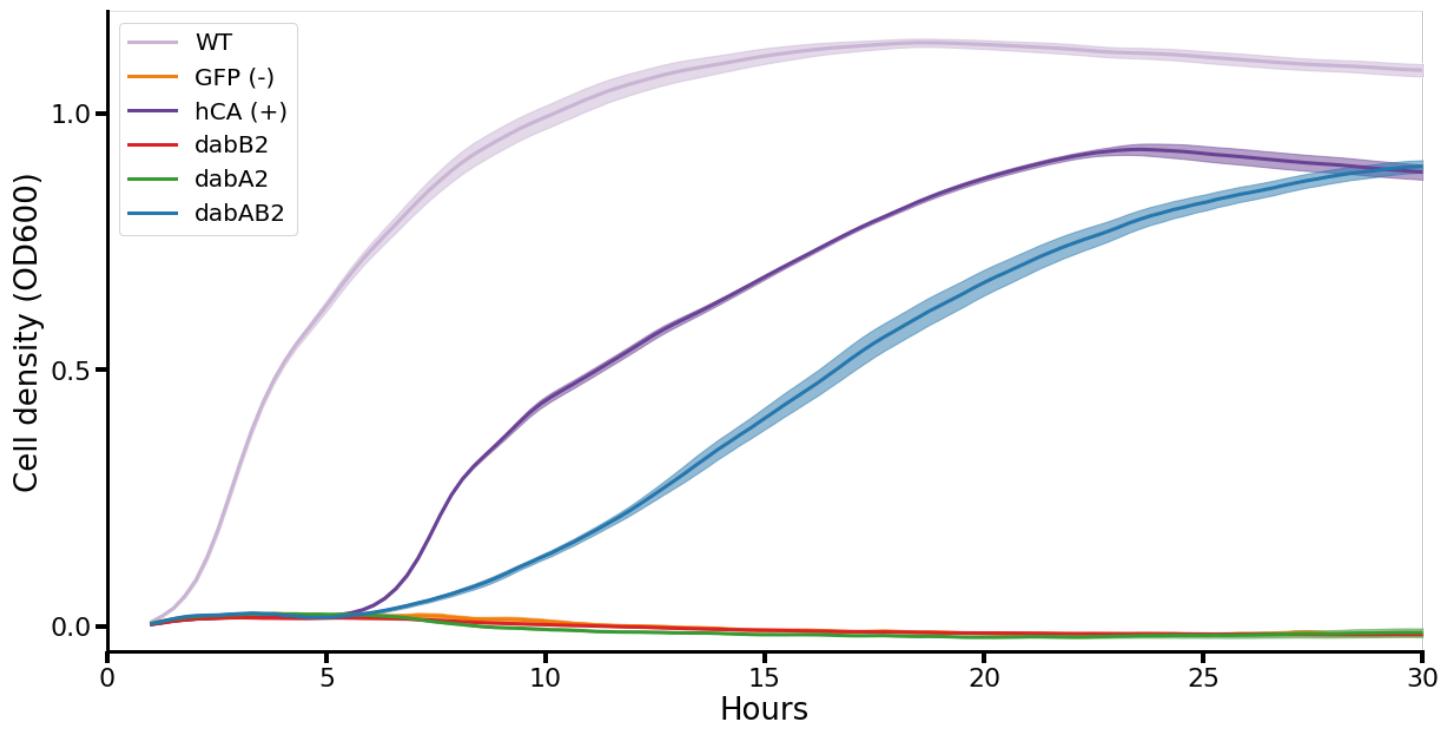
**Figure 3 S1 Nearest neighbor tree of PF0361 family proteins reveals multiple subfamilies.** PF0361 is a large and diverse protein family containing multiple subgroups with different documented activities. These subfamilies include Mrp-family antiporters, proton translocating subunits of complex I, membrane subunits of CUP ( $\text{CO}_2$  uptake protein) complexes, and DabB proteins. These subfamilies are highly diverged and perform a variety of activities. This means that it is not possible to draw conclusions about the mechanism of DAB complexes just from their homology to

841 PF0361. Clades were colored according to the presence of genes with known functions. The purple clade contains the  
842 *Bacillus subtilis* and *Staphylococcus aureus* MrpA cation antiporter subunits and the *Sinorhizobium meliloti* antiporter  
843 PhaA1. The light orange clade contains the known cation translocating subunits of complex I: nuoL from *Escherichia*  
844 *coli*, Nqo12 from *Thermus thermophilus*, and NdhF1 from both *Synechococcus elongatus* PCC7942 and  
845 *Thermosynechococcus elongatus* BP-1. The green clade contains CUP-associated membrane subunits ndhF3 from  
846 both *Synechococcus elongatus* PCC7942 and *Thermosynechococcus elongatus* BP-1 and ndhF4 from the same  
847 two species. The dark orange clade includes DabB1-2 and tcr\_0853 from *Thiomicrospira crunogena*. We note that  
848 DabB1-2 are clearly more closely related to each other and the cyanobacterial CUP-associated genes NdhF3-4 than  
849 they are to known complex I subunits or to mrp-family antiporters. This tree is consistent with our model, where DabB  
850 is not bound to a redox-coupled complex but rather couples redox-independent cation transport to CA activity (as  
851 shown in Figure 5). No conclusions should be drawn from the number of sequences in each clade as an exhaustive  
852 search for homologs was not performed to ensure that all members of each clade are represented.  
853

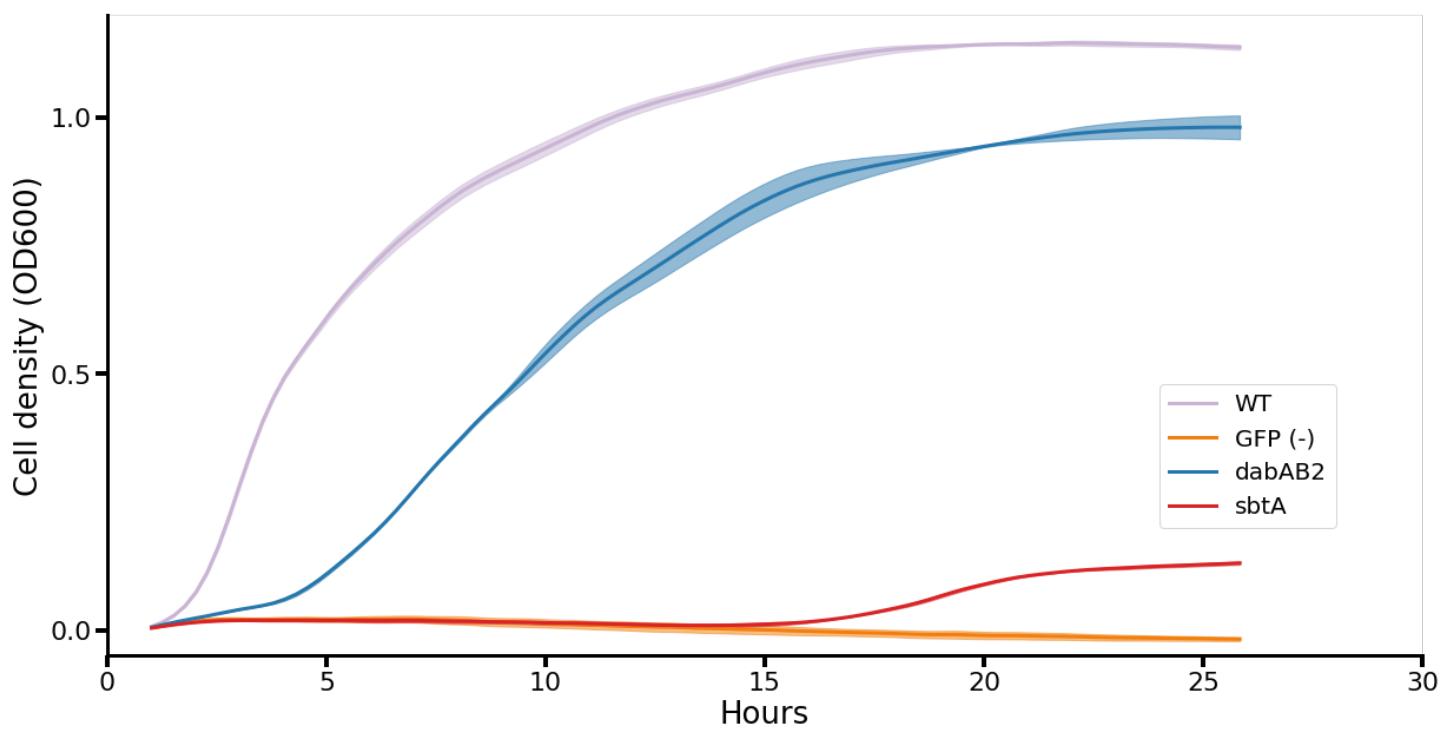


855

856 **Figure 3 S2 Operonic structure of the DAB1 and DAB2 operons.** As noted in the text and shown in Figure 2B,  
 857 DAB1 is actually a piece of a larger 11-gene operon directly downstream of the carboxysome operon and containing  
 858 CCM-associated genes. Both DAB1 (A) and DAB2 (B) “operons” contain two distinct genes that we label DabB and  
 859 DabA. DabA is annotated as Domain of Unknown Function 2309 (DUF2309, PFAM:PF10070) and appears to be a  
 860 soluble protein. Approximately one third of dabA is distantly homologous to a type II  $\beta$ -CA. CA-like regions are marked  
 861 with a line, and the four residues expected to be involved in binding the catalytic zinc ion are marked by asterisks.  
 862 The height of the asterisks has been varied to make them distinguishable despite proximity in sequence space. DabB is  
 863 homologous to a cation transporter in the same family as the H<sup>+</sup> pumping subunits of respiratory complex I  
 864 (PFAM:PF00361). The DAB1 operon also contains a protein of unknown function between DabA1 and DabB1. This  
 865 protein has distant homology to DabA1 but is truncated to half the length. Bars above the genes indicate percent  
 866 conservation of that particular amino acid position in a multiple sequence alignment (Methods). Active site residues are  
 867 in red. All active site residues are highly conserved with percent identities of greater than 99% and the active site  
 868 aspartate and one of the cysteines are the two most conserved residues in the protein with 99.89% identity each.

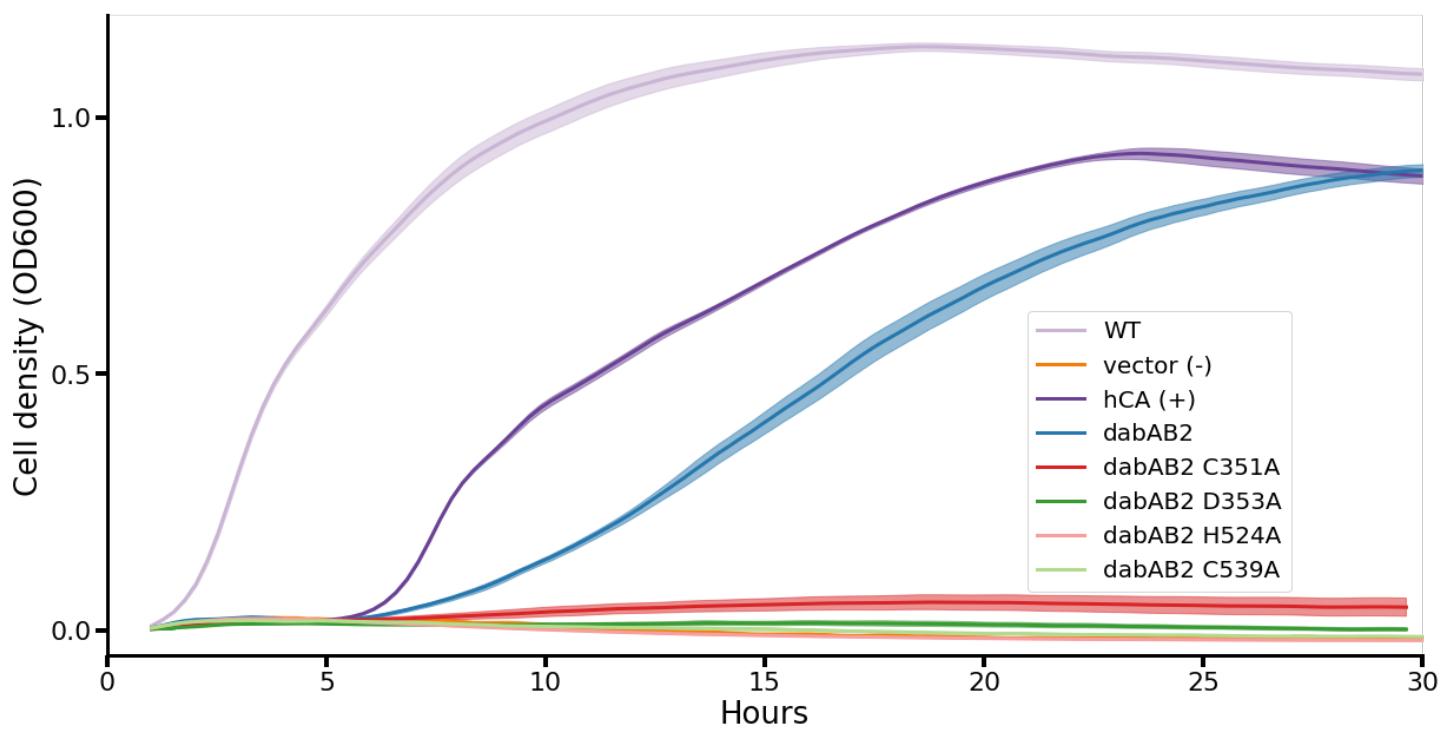


869  
870 **Figure 3 S3. Growth curves of CAfree *E. coli* growth rescue in ambient CO<sub>2</sub>.** These growth curves were used to  
871 generate the growth yield graph in figure 3B. Mean OD600 is graphed +/- standard error for four replicate cultures.  
872 Wild-type *E. coli* (BW25113) and CAfree strains expressing either dabAB2 or human carbonic anhydrase II (hCA) grow  
873 in ambient CO<sub>2</sub> while CAfree expressing GFP, dabB2 alone, or dabA2 alone fail to grow.

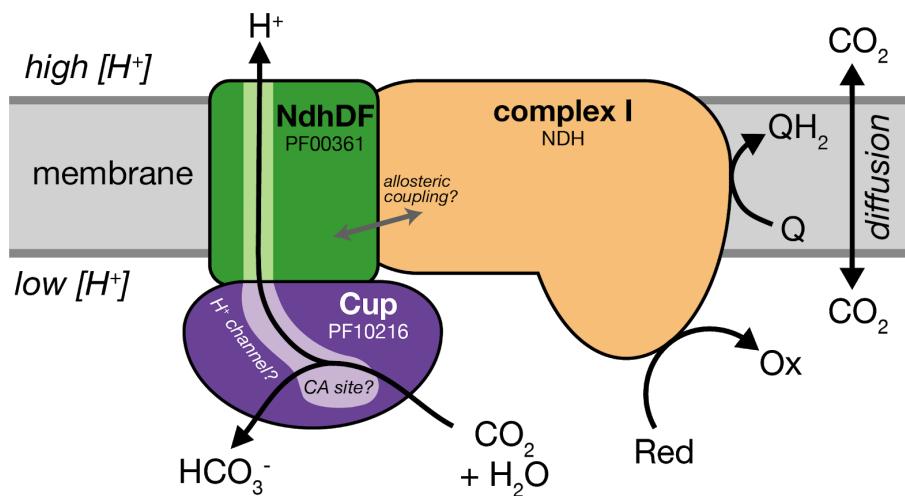
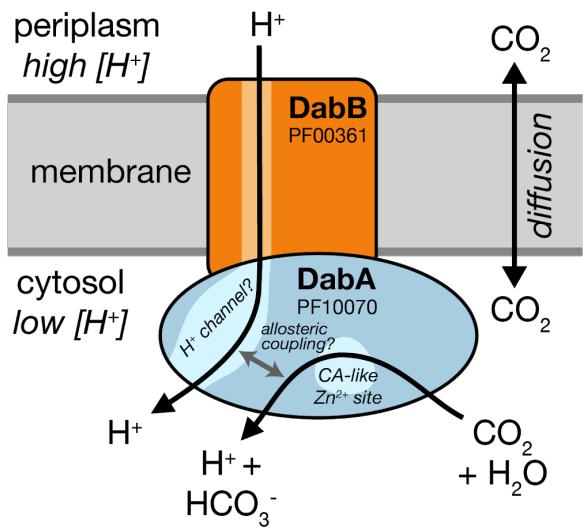
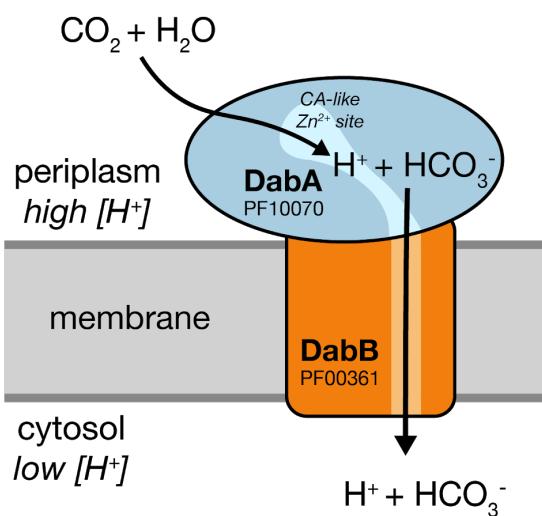


874  
875 **Figure 3 S4. Growth curves of CAfree *E. coli* rescued with *dabAB2* or the cyanobacterial HCO<sub>3</sub><sup>-</sup> transporter,**  
876 **sbtA.** Mean OD600 is graphed +/- standard error for four replicate cultures. Wild-type *E. coli* (BW25113) and CAfree  
877 strains expressing *dabAB2* grow in ambient CO<sub>2</sub> conditions, while a GFP-expressing negative control fails to grow.  
878 Expression of the cyanobacterial HCO<sub>3</sub><sup>-</sup> transporter, *sbtA*, is noticeably less effective at rescuing CAfree *E. coli* than  
879 *dabAB2* expression.

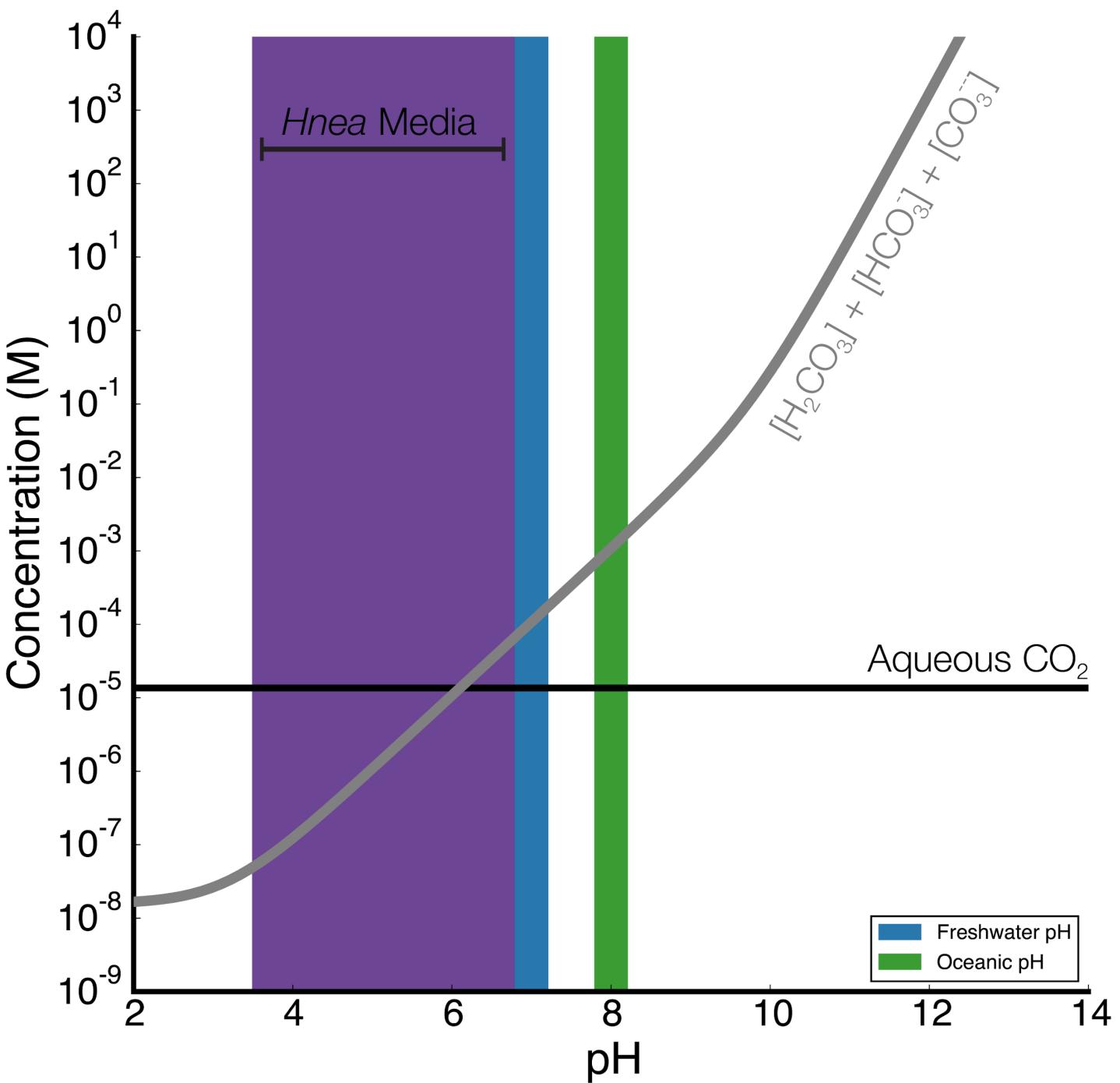
880  
881  
882  
883  
884  
885  
886  
887



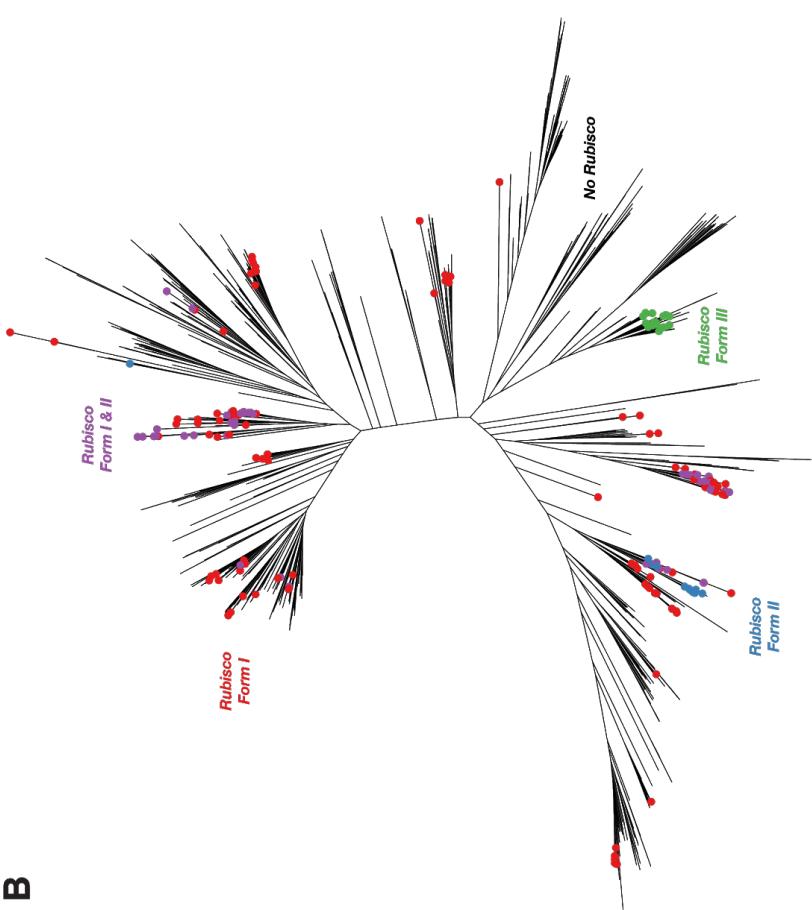
888  
889 **Figure 4 S1. Growth curves show that expressing dabAB2 active site mutants does not rescue**  
890 **CAfree**. These growth curves were used to generate the yield graph in figure 4B. The lines are mean plus and minus  
891 standard deviation of four replicate cultures. Wild type cells and those rescued with either dabAB2 or human carbonic  
892 anhydrase II (hCA) grow while those transformed with sfGFP or dabAB2 with mutations to potential active site residues  
893 C351, D353, H524, or C539 mutants fail to grow.

**Figure 5 Supplemental File 1.** Genes used to generate figure 5B.**A****B****C**

**Figure 5 S1. Comparison of models of vectorial CA activity for DABs and the Cyanobacterial Cup systems. A.** Cup proteins are CA-like subunits of a class of cyanobacterial Ci uptake systems. Cup-type systems are believed to couple electron transfer to vectorial CA activity and, potentially, outward-directed proton pumping. This model is based on the observation that Cup systems displace the two distal H<sup>+</sup>-pumping subunits of the cyanobacterial complex I and replace them with related subunits that bind CupA/B (illustrated in green as NdhDF). **B.** As our data are consistent with DAB2 functioning as a standalone complex (i.e. DabAB do not appear to bind the *E. coli* complex I), we propose a different model for DAB function where energy for unidirectional hydration of CO<sub>2</sub> is drawn from the movement of cations along their electrochemical gradient (right panel above). **C.** An alternative model for DAB activity is that DabA is localized to the periplasm and DabB is functioning as a H<sup>+</sup>: HCO<sub>3</sub><sup>-</sup> symporter. In this model DabA CA activity is made vectorial by removal of products. Energy is provided in the form of the PMF driving H<sup>+</sup> (and therefore HCO<sub>3</sub><sup>-</sup>) uptake. This model is not preferred because no secretion signals were observed in the DabA sequence and a homologous protein from *Acidimicrobium ferrooxidans* which appears to be a DabA:DabB fusion protein has a predicted architecture that would place DabA in the cytoplasm.

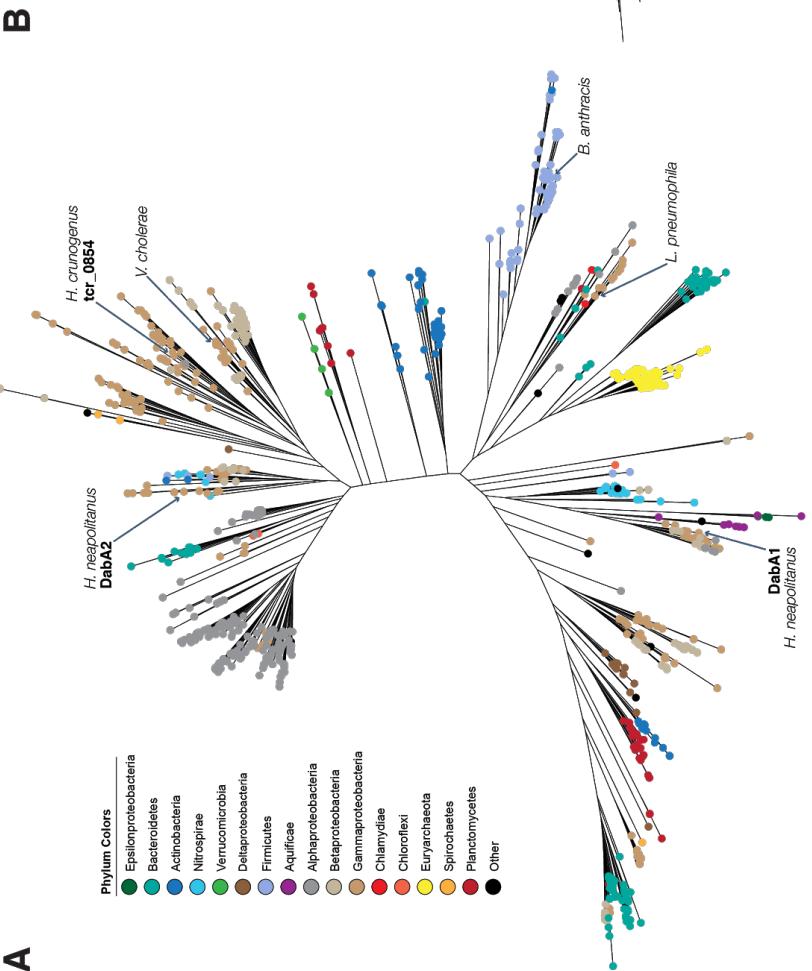


909  
910  
911 **Figure 5 S2. Equilibrium concentrations of dissolved inorganic carbon as a function of pH.** In this plot we  
912 assume the growth medium is in Henry's law equilibrium with present-day atmosphere (400 PPM  $\text{CO}_2$ ) at 25 °C giving  
913 a soluble  $\text{CO}_2$  concentration of roughly 15  $\mu\text{M}$ . The equilibrium concentrations of hydrated  $\text{C}_i$  species ( $\text{H}_2\text{CO}_3$ ,  $\text{HCO}_3^-$ ,  
914  $\text{CO}_3^{2-}$ ) is determined by the pH. As such, the organisms will "see" a  $\text{C}_i$  species in very different ratios depending on the  
915 environmental pH. In a oceanic pH near 8,  $\text{HCO}_3^-$  dominates the  $\text{C}_i$  pool.  $\text{HCO}_3^-$  is also the dominant constituent of the  
916  $\text{C}_i$  pool in freshwater, but less so (by a factor of ~10 since freshwater and oceanic environments differ by about 1 pH  
917 unit). In acid conditions (pH < 6.1)  $\text{CO}_2$  will be the dominant constituent of the  $\text{C}_i$  pool. The pH of our *Hnea* culture  
918 media ranges from 6.8 (when freshly made) to ~3.5 when cells reach stationary phase (*Hnea* make  $\text{H}_2\text{SO}_4$  as a  
919 product of their sulfur oxidizing metabolism). As such we expect that *Hnea* regularly experiences environments  
920 wherein it is advantageous to pump  $\text{CO}_2$  and not  $\text{HCO}_3^-$ .  
921



Phylum Colors

- Epsilonproteobacteria
- Bacteroidetes
- Actinobacteria
- Nitrospinae
- Verricomicrobia
- Deltaproteobacteria
- Firmicutes
- Aquificae
- Alphaproteobacteria
- Betaproteobacteria
- Gammaproteobacteria
- Chlamydiae
- Chloroflexi
- Euryarchaeota
- Spirochaetes
- Planctomycetes
- Other



923 **Figure 5 S3.** **A.** Fully annotated approximate maximum likelihood phylogenetic trees of DabA homologs associated  
924 with PF10070.9 (Methods). DabA homologs are found in > 15 prokaryotic clades, including archaea. *Hnea* DabA1 and  
925 DabA2 represent two different groupings that are commonly found in proteobacteria. The *tcr\_0854* gene of *H.  
926 crunogenus* is more closely related to DabA1 than DabA2. Inspecting the tree reveals several likely incidents of  
927 horizontal transfer, e.g. between proteobacteria and Firmicutes, Nitrospirae and Actinobacteria. Moreover, the  
928 genomes of several known pathogens contain a high-confidence DabA homolog, including *B. anthracis*, *L.  
929 pneumophila*, *V. cholerae*. **B.** Association of various Rubisco isoforms with DabA homologs. Many organisms that  
930 have DabA also have a Rubisco. However, there are numerous examples of DabA homologs that are found in  
931 genomes with no Rubisco, suggesting that this uptake system might play a role in heterotrophic metabolism. DabA is  
932 most-frequently associated with Form I Rubiscos (red and purple leaves in panel B), which is sensible because all  
933 known bacterial CCMs involve a Form I Rubisco exclusively. Some DabA-bearing genomes have only a Form II  
934 Rubisco (blue) and the Euryarchaeota genomes have that DabA have a Form III Rubisco (green) or none at all.