

1   **DABs are inorganic carbon pumps found throughout prokaryotic phyla**

2  
3   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>,  
4   Luke M. Oltrogge<sup>1</sup>, Allen W. Chen<sup>2</sup>, Kelly Wetmore<sup>3</sup>, Spencer Diamond<sup>4</sup>, Joy Y. Wang<sup>2</sup>, David F.  
5   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>Department of Chemistry, University of California, Berkeley, CA 94720

9   <sup>3</sup>Environmental Genomics and Systems Biology Division, Lawrence Berkeley National  
10 Laboratory, Berkeley, CA, USA

11   <sup>4</sup>Department of Earth and Planetary Science, University of California, Berkeley, CA 94720

12   \* Corresponding author

13

14   Correspondence should be addressed to: [savage@berkeley.edu](mailto:savage@berkeley.edu).

15    **Abstract**

16    Bacterial autotrophs often rely on CO<sub>2</sub> concentrating mechanisms (CCMs) to assimilate carbon.  
17    Although many CCM proteins have been identified, a systematic screen of CCM components  
18    has not been carried out. Here, we performed a genome-wide barcoded transposon screen  
19    to identify essential and CCM-related genes in the γ-proteobacterium *H. neapolitanus*.  
20    Screening revealed that the CCM comprises at least 17 and likely no more than 25 genes  
21    most of which are encoded in 3 operons. Two of these operons contain a two-gene locus  
22    encoding a domain of unknown function (PFAM:PF10070) and a putative cation transporter  
23    (PFAM:PF00361). Physiological and biochemical assays demonstrate that these proteins,  
24    which we name DabA and DabB for “DABs accumulate bicarbonate,” assemble into a  
25    heterodimeric complex, contain a putative β-carbonic anhydrase-like active site, and  
26    function as an energy-coupled inorganic carbon (C<sub>i</sub>) pump. Surprisingly, DabAB operons  
27    are found in diverse bacteria and archaea. We demonstrate that functional DABs are  
28    present in the human pathogens *B. anthracis* and *V. cholerae*. Based on these results, we  
29    propose that DABs constitute a new class of energized C<sub>i</sub> pump and play a critical role in C<sub>i</sub>  
30    metabolism throughout prokaryotic phyla.

31     **Introduction**

32         Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (Rubisco) is the primary  
33         carboxylase of the Calvin-Benson-Bassham (CBB) cycle and the major entry point of C<sub>i</sub> into the  
34         biosphere. Rubisco activity is critical to agriculture and a major flux removing anthropogenic  
35         CO<sub>2</sub> from the atmosphere. Despite its centrality and abundance, Rubisco is not a fast enzyme<sup>1,2</sup>.  
36         Nor is Rubisco specific - all known Rubiscos can use molecular oxygen (O<sub>2</sub>) as a substrate in  
37         place of CO<sub>2</sub><sup>3</sup>. Oxygenation does not fix carbon and produces a product that must be recycled  
38         through metabolically-expensive photorespiratory pathways<sup>4</sup>. Many studies support the  
39         hypothesis that improvements to Rubisco could improve crop yields, but Rubisco has proven  
40         recalcitrant to protein engineering. It remains unclear whether or how Rubisco can be  
41         improved<sup>2,5,6</sup>.

42         Organisms that depend on Rubisco for growth often employ CO<sub>2</sub> concentrating  
43         mechanisms (CCMs) that concentrate CO<sub>2</sub> near Rubisco so that carboxylation proceeds at high  
44         rate and specificity<sup>7,8</sup>. All cyanobacteria and many **chemolithoautotrophic** proteobacteria have a  
45         CCM<sup>9</sup>. The bacterial CCM has garnered particular interest among bioengineers because it is  
46         well-understood, thought to consist of relatively few genes and operates inside single cells<sup>10</sup>.  
47         Detailed modeling suggests that transplantation of the bacterial CCM into crops might improve  
48         yields<sup>11,12</sup> and efforts towards transplantation are already underway<sup>13</sup>.

49         Diverse experimental studies make it clear that the bacterial CCM requires two  
50         components to function: active C<sub>i</sub> transport driving accumulation of HCO<sub>3</sub><sup>-</sup> in the cytosol and  
51         organization of Rubisco with carbonic anhydrase (CA) in the lumen of a protein organelle called  
52         the carboxysome<sup>7,14,15</sup>. Energy-coupled C<sub>i</sub> pumps keep the cytosolic HCO<sub>3</sub><sup>-</sup> concentration high  
53         (> 10 mM) and, crucially, out-of-equilibrium with CO<sub>2</sub><sup>14,16,17</sup>. CA activity interconverts HCO<sub>3</sub><sup>-</sup> + H<sup>+</sup>  
54         with CO<sub>2</sub> + H<sub>2</sub>O, and thus, the carboxysomal CA converts a high cytosolic HCO<sub>3</sub><sup>-</sup> concentration  
55         into a high carboxysomal CO<sub>2</sub> concentration, promoting faster carboxylation by Rubisco and  
56         competitively inhibiting oxygenation<sup>7</sup>. Genetic lesions to either component - C<sub>i</sub> uptake systems  
57         or carboxysomes - disrupt the CCM and mutants have growth defects unless CO<sub>2</sub> is  
58         supplemented<sup>18,19</sup>. This high-CO<sub>2</sub> requiring (HCR) mutant phenotype is commonly-used to  
59         identify CCM components in screens<sup>18,19</sup>.

60         Despite early screens, a comprehensive list of bacterial CCM components remains  
61         unknown, leaving the possibility that additional activities are required for CCM function.  
62         Although well-assembled carboxysome structures can be produced in bacteria and plants<sup>13,20</sup>,  
63         the functionality of these carboxysomes in a heterologous CCM has not been demonstrated.  
64         Bioinformatic studies show that several non-carboxysomal genes are associated with

65 carboxysome operons<sup>21,22</sup>. Further, experimental<sup>14,23</sup> and modeling studies<sup>7,15</sup> make it clear that  
66 energy-coupled C<sub>i</sub> uptake systems are required for CCM function. Several different C<sub>i</sub> pump  
67 families, including transporters and facilitated uptake systems are known for cyanobacterial  
68 lineages, but mechanistic understanding of C<sub>i</sub> uptake is limited<sup>24</sup>.

69 Here we use a genome-wide barcoded transposon mutagenesis screen (RB-TnSeq) to  
70 interrogate the CCM of *Halothiobacillus neapolitanus* (henceforth *Hnea*). *Hnea* is a sulfur  
71 oxidizing γ-proteobacterial chemolithoautotroph and a model system for studying α-  
72 carboxysomes<sup>25–27</sup>. Older physiological measurements suggest that *Hnea* possesses an  
73 energized C<sub>i</sub> uptake system, but the molecular identity of this activity is unknown<sup>17</sup>. In addition to  
74 producing the first essential gene set for a bacterial chemolithoautotroph, we leverage our  
75 pooled mutant library to comprehensively screen for knockouts that produce an HCR  
76 phenotype. This screen identified all known CCM components and confirmed that a two-gene  
77 operon containing a large, conserved, poorly-characterized protein (PFAM:PF10070, hereafter  
78 DabA) and a cation transporter (PFAM:PF00361, hereafter DabB) is required for CCM  
79 function. Scott and colleagues have recently identified and validated homologs of these genes  
80 as a C<sub>i</sub> import system in hydrothermal vent chemolithoautotrophs<sup>28–30</sup>. Based on this work and  
81 results further described below, we propose naming this locus the **DAB operon** for “**DABs**  
82 **Accumulate Bicarbonate.**”

83 Here we show that the products of the DAB operon form a protein complex that is  
84 capable of energetically-coupled C<sub>i</sub> uptake. Both proteins are necessary for activity and  
85 treatment with an ionophore abrogates DAB-mediated C<sub>i</sub> uptake. Structural homology modeling  
86 suggests that DabA contains a domain distantly homologous to a type II β-CA. Indeed, DabA  
87 binds zinc, likely in a manner similar to β-CAs. These results are consistent with a model of  
88 activity dependant on unidirectional hydration of CO<sub>2</sub> to HCO<sub>3</sub><sup>-</sup> in the cytosol via a CA-like  
89 mechanism and energized by coupling to a cation gradient. Phylogenomic analysis  
90 demonstrates that DAB operons are widespread throughout prokaryotes including carbon-fixing  
91 bacteria and archaea. Surprisingly, DAB operons are also found in many heterotrophic bacteria.  
92 We demonstrate that functional operons are present in the notable pathogens *V. cholera* and *B.*  
93 *anthracis*. We therefore propose that DABs constitute a novel class of C<sub>i</sub> uptake pump whose  
94 biochemical tractability facilitates mechanistic analyses and whose widespread occurrence  
95 merits further investigation.

96 **Results**

97 *Transposon mutagenesis and gene essentiality*

98 We constructed a randomly-barcoded genome-wide pooled knockout library of *Hnea* by  
99 conjugation (Figure 1A). The conjugated vector contained a barcoded Tn5-derived transposon  
100 encoding a kanamycin resistance marker. The library was produced in 5% CO<sub>2</sub> enabling  
101 isolation of CCM gene knockouts.

102

103 Transposon barcodes simplify the use of the library for pooled screens using the ‘barseq’  
104 approach (Methods)<sup>31</sup>. Transposon insertion sites and cognate barcodes were mapped using  
105 standard TnSeq methods (Methods)<sup>31</sup>. The library was found to contain ~10<sup>5</sup> insertions, or one  
106 insertion for every ≈25 base pairs in the *Hnea* genome. Since the average gene contains ≈35  
107 insertions, genes with no insertions are very likely essential for growth. A simple statistical  
108 model identified 551 essential genes and 1787 nonessential genes out of 2408 genes in the  
109 *Hnea* genome (Methods, Figure 1A-B, File 2). The remaining 70 genes were classified as  
110 “ambiguous” due either to their short length or because replicate mapping experiments were  
111 discordant (Methods). Genes associated with known crucial functions including central carbon  
112 metabolism, ribosome production, and DNA replication were found to be essential (Figures 1C  
113 and S1). Importantly, known CCM genes, including carboxysome components, were not  
114 essential for growth at 5% CO<sub>2</sub> (Figure 2).

115

116 *Comprehensive screen for Hnea CCM components*

117 Based on the current model of the bacterial CCM (Figure 2A), knockouts of CCM genes are  
118 expected to have reduced fitness in atmospheric CO<sub>2</sub> conditions<sup>18,19</sup>. As our pooled library  
119 contains ~70,000 barcodes that map to exactly one position in the *Hnea* genome, we were able  
120 to use barseq to quantify the fitness effects of single gene knockouts for all nonessential *Hnea*  
121 genes in a pooled competition experiment (Methods, Figure 2B)<sup>31</sup>. Since the library contains  
122 roughly 20 uniquely-mapped knockouts per gene, this screen contains multiple internal  
123 biological replicates testing the effect of gene knockouts. Mutants in a particular gene were  
124 designated as HCR if the average effect of a knockout in that gene was a twofold (or greater)  
125 growth defect in ambient CO<sub>2</sub> as compared to 5% in two replicate experiments.

126

127 As expected, knockouts of carboxysome genes consistently produced large and specific fitness  
128 defects in ambient CO<sub>2</sub> (Figures 2B-C)<sup>27</sup>. These genes include *cbbLS* - the large and small

129 subunits of the  $\alpha$ -carboxysomal Rubisco<sup>9</sup>; *csoS2* - an intrinsically disordered protein required for  
130  $\alpha$ -carboxysome assembly<sup>32</sup>; *csoSCA* - the carboxysomal carbonic anhydrase<sup>9</sup>; *csoS4AB* - the  
131 pentameric proteins thought to form vertices of the  $\alpha$ -carboxysome<sup>33</sup>; and *csoS1CAB* - the  
132 hexamers that form the faces of the  $\alpha$ -carboxysome shell<sup>9,25</sup>. Knockouts of *csoS1D*, a shell  
133 hexamer with a large central pore<sup>20,34</sup>, had too weak a phenotype to be considered HCR  
134 (Figures 2B-C). The *Hnea* genome also contains a non-carboxysomal Form II Rubisco that is  
135 likely not involved in CCM activity as its disruption confers no fitness defect. A number of genes  
136 that are not associated with the carboxysome structure also exhibited HCR phenotypes. These  
137 include two LysR transcriptional regulators, a Crp/Fnr type transcriptional regulator, a protein  
138 called acRAF that is involved in Rubisco assembly<sup>35,36</sup>, and two paralogous loci encoding DAB  
139 genes (hereafter DAB1 and DAB2, Figure 2B-F).

140

#### 141 *DAB operon composition*

142 DAB1 is a cluster of 2 genes found in an operon directly downstream of the carboxysome  
143 operon (Figure 2C). Though DAB1 is part of an 11-gene operon containing several genes  
144 associated with Rubisco proteostasis, including acRAF<sup>35,36</sup> and a cbbOQ-type Rubisco  
145 activase<sup>37</sup>, we refer to DAB1 as an “operon” for simplicity. DAB2 is a true operon and is not  
146 proximal to the carboxysome operon in the *Hnea* genome. These “operons” are unified in that  
147 they both display HCR phenotypes and possess similar genes (Figures 2B-D).

148

149 Both operons contain a **conserved protein** of unknown function (PFAM:PF10070) that we term  
150 DabA. DabAs have no predicted transmembrane helices or signal peptides and appear to be  
151 large (DabA1: 118.5 kDa, DabA2: 91.7 kDa), soluble, cytoplasmic proteins (Methods, Figure  
152 3A). Both DAB operons also contain a member of the cation transporter family  
153 (PFAM:PF00361) that includes H<sup>+</sup>-pumping subunits of respiratory complex I and Mrp Na<sup>+</sup>:H<sup>+</sup>  
154 antiporters<sup>38</sup>. This protein, which we call DabB (DabB1: 62.2 kDa, DabB2: 59.3 kDa), is  
155 predicted to have 12-13 transmembrane helices (Figure 3A). The complex I subunits in  
156 PF00361 are H<sup>+</sup>-pumping proteins and do not contain redox active groups, e.g. iron-sulfur  
157 clusters or quinone binding sites. Phylogenetic analysis suggests DabB proteins form a clade  
158 among PF00361 members (Figure S4A) distinct from complex I subunits. Therefore, homology  
159 between DabB and complex I subunits (e.g. NuoL) suggests cation transport but does not imply  
160 redox activity. Importantly, operons of this type were recently demonstrated to be capable of C<sub>i</sub>  
161 uptake in the hydrothermal vent chemolithoautotroph *Hydrogenovibrio crunogenus*<sup>28-30</sup>.

162

163 *dabA2* and *dabB2* are necessary and sufficient for energy-coupled C<sub>i</sub> accumulation in *E. coli*  
164 In order to facilitate testing for C<sub>i</sub> transport activity, we generated an *E. coli* strain, CAfree, that  
165 has knockouts of both CA genes (Methods). It was previously shown that deletion of the  
166 constitutive CA, *can*, gene produces an HCR phenotype in *E. coli*<sup>39</sup> that is complemented by  
167 expression of cyanobacterial bicarbonate transporters<sup>40</sup>. Deleting both CA genes replicates this  
168 phenotype and greatly reduces the likelihood of escape mutants. Since DAB2 disruption is  
169 associated with a larger fitness defect than DAB1 (Figure 2B), we used CAfree to test DAB2 for  
170 C<sub>i</sub> uptake activity. DAB2 expression enables growth of CAfree in ambient CO<sub>2</sub> while expression  
171 of either gene alone is not sufficient (Figures 3B and S5). <sup>14</sup>C<sub>i</sub> uptake assays demonstrate that  
172 DAB2 facilitates import of extracellular C<sub>i</sub> to levels significantly above that of the appropriate  
173 control (Figure 3C). Moreover, DAB2-associated C<sub>i</sub> uptake is strongly inhibited by the ionophore  
174 CCCP (white bars in Figure 3C), indicating that DAB2 is energetically-coupled, either directly or  
175 indirectly, to a cation gradient (e.g. H<sup>+</sup> or Na<sup>+</sup>). This is consistent with previous observations that  
176 C<sub>i</sub> uptake in *Hnea* is powered by a membrane gradient<sup>17</sup>.

177

178 *DabA2 and DabB2 interact to form a complex*

179 In order to determine if the genetic interaction between *dabA2* and *dabB2* reflects a physical  
180 interaction, we attempted to co-purify the two proteins. DabA2 was fused to a C-terminal Strep-  
181 tag, DabB2 was fused to a C-terminal GFP with 6xHis-tag, and the genes were co-expressed in  
182 *E. coli* (Methods). Tandem-affinity purification following detergent solubilization revealed that  
183 DabA2 and DabB2 form a complex in *E. coli* (Figure 4A). The complex runs as a single major  
184 peak on size exclusion chromatography and has a retention volume consistent with a  
185 heterodimer of DabA2 and DabB2 (Figure 4B). We did not observe co-purification of any *E. coli*  
186 proteins suggesting that DAB2 operates as an independent complex within the membrane  
187 (Figure 4A). Moreover, DAB2 expression rescues CAfree growth even when complex I is  
188 knocked out ( $\Delta(nuoA\text{-}nuoN)$ ) (Figure S6), providing further evidence that DAB function is  
189 independent of complex I.

190

191 *pH independence of DAB2 rescue suggests that CO<sub>2</sub> is likely the true substrate*

192 Aqueous CO<sub>2</sub> spontaneously interconverts with the gas phase as well as hydrated C<sub>i</sub> species  
193 (H<sub>2</sub>CO<sub>3</sub>, HCO<sub>3</sub><sup>-</sup>, CO<sub>3</sub><sup>2-</sup>). The equilibrium of CO<sub>2</sub><sup>(aq)</sup> and CO<sub>2</sub><sup>(gas)</sup> is not affected by pH, but the  
194 conversion from CO<sub>2</sub> to hydrated C<sub>i</sub> is pH dependent. Thus, the equilibrium concentration of  
195 HCO<sub>3</sub><sup>-</sup> increases 100 fold between pH 5 and 7 without an accompanying change in CO<sub>2</sub>  
196 concentration (Figure S7A)<sup>7</sup>. Expression of SbtA, a known HCO<sub>3</sub><sup>-</sup> transporter, rescues CAfree

197 growth at pH 7 but not at pH 5, while DabAB2 rescues growth at both pHs (Figure S8). Since  
198 DabAB2 rescue is pH-independent in this range, its substrate is likely CO<sub>2</sub> and not H<sub>2</sub>CO<sub>3</sub>,  
199 HCO<sub>3</sub><sup>-</sup>, or CO<sub>3</sub><sup>-2</sup>. This is consistent with previous observations that CO<sub>2</sub> is the likely substrate of  
200 H<sub>ne</sub>a C<sub>i</sub> uptake<sup>17</sup>.

201

202 *Requirement of putative Zn-binding residues for DAB function*

203 Structural homology modeling software predicts that the middle of DabA2 has sequence  
204 elements related to a β-CA (Figure 3A). Phyre2 predictions identify C539 and H524 as part of a  
205 potential Zn<sup>2+</sup> binding site distantly homologous to a bacterial type II β-CA (10% coverage of  
206 DabA, 90.8% confidence). I-TASSER predicts a Zn<sup>2+</sup> binding site including the same residues  
207 along with an additional cysteine (C351), and aspartic acid (D353). As shown in Figure 4C,  
208 these residues could make up the active site of a type II β-CA<sup>41</sup>. We generated individual  
209 alanine mutants for each of these putative active site residues (C351A, D353A, H524A and  
210 C539A) and tested them in CAfree. All mutants failed to rescue CAfree growth in ambient CO<sub>2</sub>  
211 (Figure 4D). We proceeded to assay zinc binding of purified DabAB2 complex using X-ray  
212 fluorescence spectroscopy and found that wild-type DabAB2 and three of the single mutants  
213 (C351A, D353A, and H524A) bind zinc (Figure 4E). Single mutants retain three of four zinc-  
214 coordinating residues<sup>41</sup>, which could explain why the mutants bind zinc. Indeed, mutational  
215 studies of the human CA II show that single mutations to Zn<sup>2+</sup>-binding residues reduce but do  
216 not abrogate zinc binding<sup>42</sup>.

217

218 *Purified DabAB2 complex does not have conspicuous CA activity.*

219 The assay of detergent solubilized, purified DabAB2 did not show significant carbonic  
220 anhydrase activity over controls (Figure 4F). DabAB2 was assayed at high protein  
221 concentrations (> 650-fold more protein than the positive control) and under CO<sub>2</sub> concentrations  
222 that are typically saturating for CAs, but displayed no activity (Figure 4F). Absence of activity *in*  
223 *vitro* argues either that DabAB2 has extremely low CA activity or, *perhaps*, that DabAB2 must  
224 reside in a cell membrane holding a cation gradient to function as an energetically-activated  
225 carbonic anhydrase.

226

227 *DAB operons are widespread in prokaryotes and present in human pathogens*

228 A query of the Uniprot database with the DabA PFAM (PF10070) yielded 878 putative DabA  
229 sequences. DabAs were found in a wide variety of prokaryotes including bacteria and archaea  
230 (Figure 5A and S9), as is consistent with previous work<sup>28</sup>. Represented clades include not only

231 Proteobacteria, but also Euryarchaeota, Firmicutes, Planctomycetes, and Bacteroides.  
232 However, we were surprised to observe many *dabA* sequences were found in genomes of  
233 organisms that cannot fix CO<sub>2</sub> including the heterotrophic human pathogens *V. cholera*, *B.*  
234 *anthracis*, and *L. pneumophila* (Figure 5A). Notably, 843 (96%) of the identified *dabA*  
235 sequences were either within three genes of, or fused to, a *dabB*.

236

237 Finally, we assayed whether the DAB homologs from heterotrophic pathogens are functional C<sub>i</sub>  
238 pumps. Operons from *V. cholera* E7946 El Tor Ogawa and *B. anthracis* Sterne were cloned and  
239 expressed in CAfree. Both DAB operons rescued CAfree growth in ambient CO<sub>2</sub> (Fig. 5B and  
240 S10). Thus, DAB operons from heterotrophic, human pathogens are functional.

241

242

243 **Discussion**

244 Here, we generated a knockout library containing ≈35 individual knockouts for every gene in the  
245 genome of the proteobacterial **chemolithoautotroph** *H. neapolitanus*. Using these data, we  
246 compiled the essential gene set of a **chemolithoautotroph** (Figure 1) and were able to  
247 confidently identify 551 essential genes and 1787 nonessential genes. Mapping essential genes  
248 will provide insight into the metabolism and growth physiology of sulfur-oxidizing  
249 **chemolithoautotrophs**.

250

251 In addition to mapping essential genes, this library would be used to measure conditional  
252 phenotypes for nonessential genes. These mutants were isolated in high CO<sub>2</sub> and so we were  
253 able to disrupt all known components of the bacterial CCM (Figure 2). The resulting genome-  
254 wide knockout library was used to perform a comprehensive screen for novel bacterial CCM  
255 genes. This screen highlighted a small number of genes (17) as having the HCR phenotype  
256 associated with the CCM (Figure 2B-F), nearly all of these genes are known to be associated  
257 with the α-carboxysome. Though it is possible that genetic redundancy, conditional phenotypes,  
258 or **impairment only at sub-ambient CO<sub>2</sub>** permit some genes to escape notice, these data  
259 suggest that the proteobacterial CCM is composed of < 30 functionally distinct components.  
260 Moreover, none of the genes identified have unexpected functions, suggesting that current  
261 models of bacterial CCMs incorporate all necessary functions.

262

263 Our screen identified 3 transcriptional regulators as well as 3 distinct CCM operons (Figures 2B-  
264 F). Identification of transcriptional regulators with HCR phenotypes (Figures 2D-F) may inform  
265 the study of CCM regulation. The first operon contains nearly all known components of the α-  
266 carboxysome, all of which confer HCR phenotypes upon disruption (Figure 2C). The second  
267 operon is adjacent to the carboxysome operon and contains 11 genes. Only 3 of these genes -  
268 the Rubisco chaperone *acRAF* and *dabAB1* - displayed HCR phenotypes (Figure 2C). The  
269 remaining 8 genes had no associated phenotype but might nonetheless have roles in the CCM.  
270 These genes include *cbbOQ*, *csos1D*, *p-II*, and a *parA* homolog (Figure 2C). The third operon  
271 contains two genes, *dabAB2*, both with HCR phenotypes (Figure 2D).

272

273 A previous physiological study suggested that *Hnea* C<sub>i</sub> uptake is coupled to the membrane  
274 electrochemical potential and uses CO<sub>2</sub> as a substrate, but the protein(s) responsible for this  
275 activity were unknown<sup>17</sup>. DAB1 and DAB2 are homologous to C<sub>i</sub> pumps from hydrothermal vent

276 chemolithoautotrophs recently discovered by Scott and colleagues<sup>28,30</sup> and our screen suggests  
277 that DAB1 and DAB2 are likely the C<sub>i</sub> pumps in *Hnea*. These observations raise many  
278 mechanistic questions as to how DABs function, and we therefore sought to establish a  
279 biochemical system to investigate DAB structure-function.

280

281 We showed that the DAB2 operon encodes a two-component protein complex that has C<sub>i</sub>  
282 uptake activity when heterologously expressed in *E. coli* (Figures 3B-C & 4A). This complex is  
283 likely a heterodimer as suggested by size-exclusion chromatography (Figure 4B). As C<sub>i</sub> uptake  
284 is inhibited by the ionophore CCCP (Figure 3C), we suspect that DAB2 activity is energetically-  
285 coupled to a cation gradient (Figure 5A). Moreover, since DabAB2 shows pH-independent  
286 rescue of CAfree *E. coli*, CO<sub>2</sub> is likely the transported substrate (Figure 4C). This is further  
287 supported by the fact that DabA has distant homology to a type II β-CA and binds a zinc  
288 (Figures 3-4), suggesting that a CA active site hydrates transported CO<sub>2</sub>. Finally, mutations to  
289 the putative zinc-binding residues (C351A, D353A, H524A, and C539A) ablate function *in vivo*  
290 (Figure 4D). We therefore propose a speculative model of DAB activity wherein CO<sub>2</sub> is passively  
291 taken into the cell and then unidirectionally hydrated to HCO<sub>3</sub><sup>-</sup> by energy-coupled CA activity of  
292 DabA (Figure 6).

293

294 Model carbonic anhydrases are not coupled to any energy source (e.g., ATP, cation gradient).  
295 Rather, they equilibrate CO<sub>2</sub> and HCO<sub>3</sub><sup>-</sup><sup>42</sup>. However, energy-coupled CA activity could favor  
296 CO<sub>2</sub> hydration, allowing the DAB system to actively accumulate HCO<sub>3</sub><sup>-</sup> in the cytosol and power  
297 the CCM (Figure 2A). Given the similarity of DabB to H<sup>+</sup>-pumping proteins, we propose that  
298 DABs use the H<sup>+</sup> gradient, though our results are equally consistent with other cation gradients,  
299 e.g. Na<sup>+</sup>. This mechanism would require tight coupling of cation flow to CA activity by DabA,  
300 consistent with our observation that purified DabAB2 displays no measurable CA activity.  
301 Interestingly, type II β-CAs are the only CAs known to display allosteric regulation<sup>43</sup>. Allosteric  
302 control is thought to be mediated by Zn<sup>2+</sup> capping and uncapping by the active site aspartic acid  
303 (D353 in DabA2)<sup>43</sup>. A similar mechanism might couple cation flow through DabB to the active  
304 site of DabA.

305

306 Cyanobacteria possess vectorial CAs called CUPs, which may provide clues to the DAB  
307 mechanism<sup>24,44,45</sup>. Indeed, both DAB and CUP systems contain subunits in the Mrp protein  
308 family (DabB and NdhD/F are in PF00361) that also contains the H<sup>+</sup>-pumping subunits of  
309 complex I. This commonality might suggest similar mechanisms. CO<sub>2</sub> hydration by CUPs is

310 thought to be coupled to energetically-favorable electron flow because CUPs associate with  
311 complex I<sup>46</sup> (Figure S9B). However, the Mrp protein family (PF00361) is very diverse and  
312 contains many cation transporters that do not associate with complex I or any other redox-  
313 coupled complex<sup>38</sup>. Moreover, DabB sequences are only distantly related to complex I and CUP  
314 subunits (Figure S4A), DAB2 subunits do not co-purify with *E. coli* complex I (Figure 4A) and  
315 DAB2 rescues CAfree growth in a complex I knockout (Figure S6). We therefore propose that  
316 DAB activity is not coupled to electron flow through complex I but, rather, to a cation gradient  
317 across the membrane (Figure 6).

318

319 DabAB2 functions robustly, as demonstrated by complementation of CAfree (Figure 3B) and <sup>14</sup>C  
320 uptake measurements (Figure 3C). Indeed, we observed that DabAB2 functions substantially  
321 better in *E. coli* than SbtA, a C<sub>i</sub> transporter from cyanobacteria<sup>9,40</sup> (Figure 3C). As *E. coli* and  
322 *Hnea* are proteobacteria, this observation could result from greater “compatibility” of  
323 proteobacterial proteins with *E. coli* expression. It may also be the case, though, that the α-CCM  
324 of proteobacteria is more “portable” than the β-CCM of freshwater cyanobacteria. Indeed, α-  
325 CCM genes are typically found in a single gene cluster in **chemolithoautotrophs** throughout α-β-  
326 and γ-proteobacteria and the α-CCM was clearly horizontally transferred from proteobacteria to  
327 marine cyanobacteria<sup>9</sup>. DabA homologs are widespread in prokaryotes and were likely  
328 horizontally transferred multiple times (Figure 5A). Since DAB complexes are prevalent among  
329 prokaryotes and have superlative activity, DAB-family transporters are an attractive target for  
330 protein engineering and heterologous expression in plants and industrial microbes, where  
331 elevated intracellular C<sub>i</sub> could be useful<sup>47</sup>.

332

333 Finally, DABs are present in a wide variety of bacteria and archaea<sup>28</sup>. High-confidence DabA  
334 homologs are found not only in large numbers of autotrophs but also in heterotrophs (Figure 5A  
335 & S9). Moreover, homologs are present in the notable heterotrophic pathogens *V. cholerae*, *B.*  
336 *anthracis*, and *L. pneumophila* (Figure 5A). We showed that DABs from *V. cholerae* and *B.*  
337 *anthracis* are active in *E. coli* (Figure 5B). This leads us to wonder: what do heterotrophic  
338 pathogens use C<sub>i</sub> uptake systems for? Carbonic anhydrase activity is essential for growth of the  
339 heterotrophs *E. coli* and *S. cerevisiae* in ambient CO<sub>2</sub><sup>39,48</sup>. In the heterotrophic context, CA  
340 activity is thought to supply bicarbonate for biotin-dependent carboxylases in central  
341 metabolism, for which HCO<sub>3</sub><sup>-</sup> is the substrate<sup>39,48</sup>. Additionally, bicarbonate levels have been  
342 linked to virulence in both *V. cholera* and *B. anthracis*<sup>49,50</sup>. Perhaps DAB-family C<sub>i</sub> uptake  
343 systems play roles in the growth or virulence of these important pathogens? We hope that future

344 research will delineate the role of energetically-activated C<sub>i</sub> uptake in heterotrophic and  
345 pathogenic organisms.

346 **Materials and Methods**

347 *Important strains and reagents*

348 A detailed listing of key strains and reagents is given in Supplemental File 1.

349

350 *Bacterial strains and growth conditions*

351 *E. coli* strain APA766 was used as the conjugation donor to transfer the Tn5 transposon to  
352 *Halothiobacillus neapolitanus* C2 (*Hnea*) via conjugation<sup>31</sup>. The *E. coli* double CA deletion strain  
353 “CAfree” (BW25113  $\Delta$ can  $\Delta$ cynT) was generated by curing the KEIO collection cynT knockout  
354 (BW25113  $\Delta$ cynT, KEIO strain JW0330) of kanamycin resistance via pCP20-mediated FLP  
355 recombination and subsequent P1 transduction (and curing) of kanamycin resistance from the  
356 can knockout strain EDCM636 (MG1655  $\Delta$ can, Yale Coli Genomic Stock Center,<sup>39,51</sup>). Complex  
357 I knockout strains ( $\Delta$ (nuoA-nuoN)) were generated in the BW25113 and CAfree backgrounds.  
358 These strains were generated by lambda red mediated recombination of a Kan<sup>R</sup> resistance  
359 cassette flanked by FRT sites into the nuo locus such that the entire operon was removed. The  
360 pSIM5 plasmid carrying the lambda red recombinase was heat cured at 42 °C. Lysogeny broth  
361 (LB) and LB agar were used as *E. coli* growth media unless otherwise specified. *E. coli* strains  
362 were grown at 37 °C in the presence of 0.1 mg/ml carbenicillin, 0.06 mg/ml kanamycin, or 0.025  
363 mg/ml chloramphenicol as appropriate. *Hnea* was grown in DSMZ-68 media at 30 °C and in the  
364 presence of 0.03 mg/ml kanamycin when appropriate.

365

366 *Transposon mutagenesis and RB-TnSeq library production*

367 A barcoded library of *Hnea* transposon mutants was generated by adapting the methods of  
368 Wetmore *et al.*<sup>31</sup>. Conjugations were performed as follows. *Hnea* and APA766 were cultured  
369 and harvested by centrifugation. Both cultures were washed once in 10 mL antibiotic-free  
370 growth media per conjugation reaction and resuspended in 100  $\mu$ L 5 OD600 units of *Hnea* were  
371 mixed with 20 OD600 units of APA766 on a 0.45  $\mu$ M Millipore MCE membrane filter and  
372 cultured overnight at 30 °C in 5% CO<sub>2</sub> on an antibiotic-free LB agar plate containing 0.06 mg/ml  
373 diaminopimelic acid. Cells were scraped from the filter into 2 mL DSMZ-68 and collected in a 2  
374 mL microcentrifuge tube. Recovered cells were pelleted by centrifugation at 16000 x g for 1  
375 minute, washed in 2 mL DSMZ-68, pelleted again at 9000 x g for 1 minute, and resuspended in  
376 2 ml DSMZ-68 before 200  $\mu$ L was plated onto 10 separate DSMZ-68 kanamycin plates (per  
377 conjugation). Plates were incubated at 30 °C under 5% CO<sub>2</sub> until colonies formed (~ 7 days).  
378 Colonies were counted and scraped into 55 mL DSMZ-68. Two 1.4 OD600 unit samples were  
379 taken and used to prepare genomic DNA (Qiagen DNeasy blood and tissue kit). Transposon

380 insertions were amplified from gDNA and transposons were mapped after Illumina sequencing  
381 using protocols and software from Wetmore *et al.*<sup>31</sup> 1.6 OD600 unit aliquots were then flash  
382 frozen in 50% glycerol for subsequent BarSeq experiments.

383

#### 384 *Essential gene assignment*

385 Following the logic of Wetmore *et al.* and Rubin *et al.*<sup>31,52</sup>, we categorized genes as essential if  
386 we observed significantly fewer transposon insertions than would be expected by chance. If  
387 insertion occurred uniformly at random, the number of insertions per gene would be expected to  
388 follow a binomial distribution. The probability of observing at most  $k$  insertions into a gene of  
389 length  $n$  is therefore expressed as:

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

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

397

#### 398 *Gene fitness experiments*

399 Fitness experiments were performed according to a modification of the protocol in Wetmore *et*  
400 *al.*<sup>31</sup>. This method allows pooled library fitness experiments to be performed comparing different  
401 growth conditions by comparing barcode abundance changes in order to track changes in the  
402 abundance of the transposon mutants. In short, a library aliquot was thawed and used to  
403 inoculate three 33 mL cultures. Cultures were grown to OD600 ~0.08 in 5% CO<sub>2</sub>. At this point,  
404 20 mL were removed and harvested by centrifugation as two t<sub>0</sub> (input) samples. Cultures were  
405 back-diluted 1:64 into 128 mL and incubated for 6.5-7.5 doublings under 5% CO<sub>2</sub> or ambient  
406 conditions. 50 mL of culture was harvested by centrifugation. gDNA was prepared and barcodes  
407 were amplified for fitness determination via Illumina sequencing as described previously<sup>31</sup>.  
408 Fitness values were calculated using existing software<sup>31</sup>. Genes were assigned an HCR  
409 phenotype if they had a fitness defect of two fold or greater in ambient CO<sub>2</sub> compared to 5%  
410 CO<sub>2</sub> in two replicate experiments.

411

412 CAfree rescue experiments

413 Electrocompetent CAfree cells were prepared using standard protocols<sup>53</sup> and transformed with  
414 pFE plasmids expressing genes of interest by electroporation. CAfree pre-cultures were grown  
415 overnight in 10% CO<sub>2</sub> and diluted into 96 well plates (3 µl cells in 250 µl media). Growth curves  
416 were measured by culturing cells in a Tecan M1000 microplate reader under ambient conditions  
417 with continuous shaking, and measuring OD600 every 15 minutes. When samples are marked  
418 "induced," 200 nM anhydrotetracycline (aTc) was added to the media. Growth yields are  
419 calculated as the maximum OD600 achieved after 24 hours of growth and normalized to the  
420 yield of a wild type control. CFU experiments were performed by back diluting cultures to  
421 OD600 0.2 before performing 10X serial dilutions. 3 µl of the OD600 0.2 sample and each of the  
422 serial dilutions were then spotted on plates with 200 nM aTc and grown overnight in **ambient**  
423 **conditions (400 ppm CO<sub>2</sub>)**. The spot with the highest dilution that yielded more than one colony  
424 was counted and a minimum of six replicates were averaged for each strain.

425

426 *Silicone oil centrifugation measurement of C<sub>i</sub> uptake*

427 The silicone oil filtration method was modified from Dobrinski *et al.*<sup>54</sup> and used to measure  
428 uptake of radiolabeled inorganic carbon. Assay tubes were generated using 0.6 ml  
429 microcentrifuge tubes containing 20 µl of dense kill solution (66.7% v/v 1 M glycine pH 10,  
430 33.3% v/v triton X-100) covered by 260 µl of silicone oil (4 parts AR20:3.5 parts AR200).  
431 Electrocompetent CAfree cells were prepared using standard protocols and transformed with  
432 pFA-based plasmids containing genes of interest by electroporation. CAfree cultures were  
433 grown overnight in 10% CO<sub>2</sub>, back diluted to an OD600 of 0.1 and allowed to grow to mid-log  
434 phase in 10% CO<sub>2</sub> in the presence of 200 nM aTc for induction. Cells were then harvested by  
435 centrifugation, washed once in PBS (pH 7.55) and resuspended to OD600 0.6 in PBS + 0.4%  
436 glucose. <sup>14</sup>C-labeled sodium bicarbonate (PerkinElmer) was added to a final concentration of  
437 4.1 nM and an activity of 0.23 µCi. Cells were incubated with <sup>14</sup>C for 4 minutes before  
438 centrifugation at 17,000 x g for 4 minutes to separate cells from buffer. Pellets were clipped into  
439 scintillation vials containing 5 ml Ultima Gold scintillation fluid and 300 µl 3M NaOH using  
440 microcentrifuge tube clippers or medium dog toenail clippers. Counts were measured on a  
441 PerkinElmer scintillation counter. <sup>14</sup>C counts are normalized to 1 OD600 unit of cells added.  
442 During inhibition assays, cells were incubated in PBS pH 7.55 with 0.4% glucose + 0.4% DMSO  
443 and the inhibitor (100 µM CCCP) for 10 minutes before assay.

444

445 *Generation of DabA phylogenetic tree*  
446 We searched the Uniprot reference proteome database using the Pfam Hidden Markov Model  
447 PF10070.9 with a cutoff e-value of  $10^{-4}$ . Our search recovered 941 candidate DabA proteins.  
448 These sequences were aligned using MAFFT and manually pruned to remove fragments and  
449 poorly aligning sequences. The remaining 878 candidate DabA sequences were **re-aligned with**  
450 **MAFFT** and an approximate maximum likelihood phylogenetic tree was constructed using  
451 FastTree. Taxonomy was assigned to nodes in the tree based on NCBI taxonomy information  
452 for the genomes harboring each sequence. Genomic neighborhoods for each gene in the tree  
453 were determined using the EFIGNT online server<sup>55</sup> and genomes with a *dabB* gene within 3  
454 genes of *dabA* and oriented in the same direction were considered to have full DAB operons.  
455 *dabAB* fusions were found by visual inspection of genomic neighborhoods from those genomes  
456 that did not have separate *dabB* genes located close to *dabA*.

457

458 *Generation of DabB phylogenetic tree*

459 DabB homologs were collected manually by searching MicrobesOnline for close homologs of  
460 four PF00361 members in the *Hnea* genome (*dabB1*, *dabB2*, *Hneap\_1953*, *Hneap\_1130*) and  
461 other characterized PF00361 members including *Synechococcus elongatus ndhF1*,  
462 *Synechococcus elongatus ndhF3*, and *Synechococcus elongatus ndhF4*. Genes were clustered  
463 to 95% similarity and genes with divergent operon structure were removed manually using  
464 MicrobesOnline treeview<sup>56</sup>. *nuoL* from *Escherichia coli*, *nqo12* from *Thermus thermophilus*, and  
465 *ndhF1/3/4* from *Thermosynechococcus elongatus* BP-1 were added as markers. ClustalOmega  
466 was used to construct a multiple sequence alignment and an approximate maximum likelihood  
467 phylogenetic tree was constructed using FastTree<sup>57,58</sup>. The tree was visualized using the  
468 Interactive Tree of Life<sup>59</sup>.

469

470 *Protein annotation and structural homology modeling*

471 Secondary structural annotations for DabA and DabB were generated using XtalPred<sup>60</sup>.  
472 Structural Homology modeling of DabA was performed using Phyre2 and I-TASSER web  
473 servers with default parameters<sup>61,62</sup>. A list of close DabB homologs was assembled by searching  
474 MicrobesOnline for PF00361 members with similar operon structure. A ClustalOmega alignment  
475 was used to calculate residue-level conservation of DabB proteins while the MAFFT alignment  
476 generated during the creation of the DabA tree was used to calculate residue level conservation  
477 of DabA proteins (Figure S4B).

478

479 *Purification of DAB2*

480 Chemically competent BL21-AI *E. coli* were transformed with a pET14b-based vector containing  
481 the *dabAB* genes. 1 liter of 2xYT media was inoculated with 20 ml of an overnight culture of  
482 BL21-AI *E. coli* in LB+CARB and allowed to grow to mid log at 37 °C. When midlog was  
483 reached, cells were induced with 20 ml of 50 mg/ml arabinose and transitioned to 20 °C for  
484 overnight growth. Cultures were pelleted and resuspended in 10 ml TBS (50 mM Tris, 150 mM  
485 NaCl, pH 7.5) supplemented with 1.2 mM phenylmethylsulfonyl fluoride, 0.075 mg/ml lysozyme  
486 and 0.8 ug/ml DNase I per liter of starting culture and then incubated at room temperature on a  
487 rocker for 20 minutes. Cells were lysed with four passes through a homogenizer (Avestin).  
488 Lysate was clarified at 15,000 x g for 30 minutes. Membranes were pelleted at 140,000 x g for  
489 90 minutes. Membrane pellets were resuspended overnight in 25 ml TBS supplemented with 1  
490 mM phenylmethylsulfonyl fluoride and 1% β-dodecyl-maltoside (DDM, Anatrace) per liter of  
491 culture following<sup>63</sup>. Membranes were then re-pelleted at 140,000 - 200,000 x g for 60 minutes  
492 and the supernatant was incubated with Ni-NTA beads (Thermo Fisher) for 90 min at 4 °C. The  
493 resin was washed with “Ni buffer” (20 mM Tris + 300 mM NaCl + 0.03% DDM, pH 7.5)  
494 supplemented with 30 mM imidazole and eluted with Ni buffer supplemented with 300 mM  
495 imidazole. Eluent was then incubated with Strep-Tactin (Millipore) resin for 90 min at 4 °C.  
496 Resin was washed with “strep buffer” (TBS + 0.03% DDM) and eluted with strep buffer  
497 supplemented with 2.5 mM desthiobiotin. Eluent was concentrated using Vivaspin 6 100 kDa  
498 spin concentrators and buffer exchanged into strep buffer by either spin concentration or using  
499 Econo-Pac 10DG (Biorad) desalting columns. For analytical purposes, 300 µg of strep-purified  
500 protein was injected onto a Superdex 200 Increase 3.2/300 size-exclusion column pre-  
501 equilibrated in strep buffer and eluted isocratically in the same buffer.

502

503 *Carbonic anhydrase assays*

504 CA-catalyzed CO<sub>2</sub> hydration of purified DAB2 complex and human carbonic anhydrase (hCA)  
505 was measured using the buffer/indicator assay of Khalifah<sup>64</sup> on a KinTek AutoSF-120 stopped-  
506 flow spectrophotometer at 25 °C. The buffer/indicator pair used was TAPS/*m*-cresol purple  
507 measured at a wavelength of 578 nm using a pathlength of 0.5 cm. Final buffer concentration  
508 after mixing was 50 mM TAPS, pH 8.0 with the ionic strength adjusted to 50 mM with Na<sub>2</sub>SO<sub>4</sub>,  
509 and 50 µM of pH-indicator. Final protein concentration used was: 9.8 µM DAB2 (His-elution) and  
510 0.015 µM hCA (positive control; Sigma Aldrich C6624). Saturated solution of CO<sub>2</sub> (32.9 mM)  
511 was prepared by bubbling CO<sub>2</sub> gas into milli-Q water at 25 °C. The saturated solution was  
512 injected into the stopped-flow using a gas-tight Hamilton syringe, and measurements were

513 performed in a final CO<sub>2</sub> concentration of 16.5 mM. Progression curves were measured in 7  
514 replicates.

515

516 *X-ray fluorescence spectroscopy for metal analysis*

517 50-100 µg of protein in 20-200 µl of TBS + 0.03% DDM was precipitated by addition of 4  
518 volumes of acetone and incubation at -20 °C for 1 hour. Samples were centrifuged at 21,130 x g  
519 for 15 minutes in a benchtop centrifuge and the supernatant was removed. Pellets were stored  
520 at 4 °C until analysis. Fluorescence analysis was performed by breaking up the pellet into 5 µl of  
521 TBS + 0.03% DDM with a pipette tip. Small pieces of the pellet were looped with a nylon loop  
522 and flash frozen at the beamline under a nitrogen stream. The sample was excited with a 14  
523 keV X-ray beam and a fluorescence spectrum was collected. Sample emission spectra were  
524 then used to identify metals. Metal analysis was performed on wild-type DAB2, Zn-binding  
525 mutants C351A, D353A, and H524A, bovine CA (positive control; Sigma Aldrich C7025), and a  
526 buffer blank was used as a negative control. A Rubisco crystal containing cobalt salts was also  
527 used as a zinc free control. Displayed traces are averages of at least two experiments.  
528 Experiments were performed at the Lawrence Berkeley National Laboratory Advanced Light  
529 Source Beamline 8.3.1.

530 **Materials & Correspondence**

531 Correspondence should be addressed to: [savage@berkeley.edu](mailto:savage@berkeley.edu). Materials will be available  
532 upon reasonable request.

533

534 **Acknowledgements**

535 We thank Adam Deutschbauer and Morgan Price for assistance with RB-TnSeq experiments  
536 and analysis, respectively. Genomic DNA samples were kindly provided by Zoe Netter and  
537 Kimberly Seed (*V. cholera*) and Dan Portnoy and Richard Calendar (*B. anthracis* Sterne). We  
538 thank Andreas Martin and Jared Bard for assistance with stopped flow experiments. Thanks to  
539 Emeric Charles, Woodward Fischer, Britta Forster, Ben Long, Robert Nichols, Dean Price and  
540 Patrick Shih for useful conversations and comments on the manuscript. X-ray-based  
541 experiments were performed at the Lawrence Berkeley National Laboratory Advanced Light  
542 Source Beamline 8.3.1. J.J.D. was supported by National Institute of General Medical Sciences  
543 grant-T32GM066698. A.F. and T.G.L. were supported by a National Science Foundation  
544 Graduate Research Fellowship. C.B. was supported by an International Postdoctoral grant from  
545 the Swedish Research Council 637-2014-6914. D.F.S. was supported by the US Department of  
546 Energy Grant DE-SC00016240.

547

548 **Data availability**

549 All illumina sequencing data will be made publicly available upon acceptance of the paper  
550 (accession number: XXXXXX). All other data is available on github at:  
551 <https://github.com/jackdesmarais/DabTransporterPaper>.

552

553 **Code availability**

554 All custom code is available on github at:  
555 <https://github.com/jackdesmarais/DabTransporterPaper>.

556

557 **Author contributions**

558 J.J.D., A.I.F., and D.F.S. conceived and designed this study, and wrote the final manuscript with  
559 input from all authors; J.J.D., A.I.F., C.B., E.J.D., T.G.L., L.M.O., A.W.C., S.D., K.W., J.Y.W.,  
560 and D.F.S. conducted the research or interpreted results. All authors reviewed and approved  
561 the final manuscript.

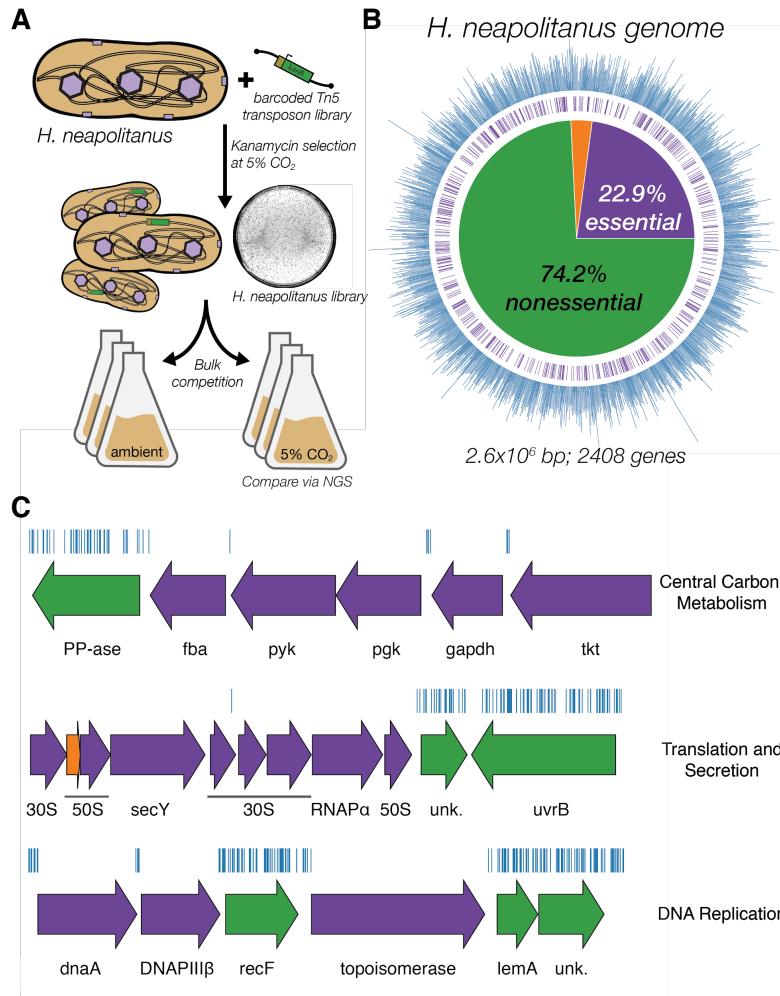
562

563    **Competing Interests**

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

568

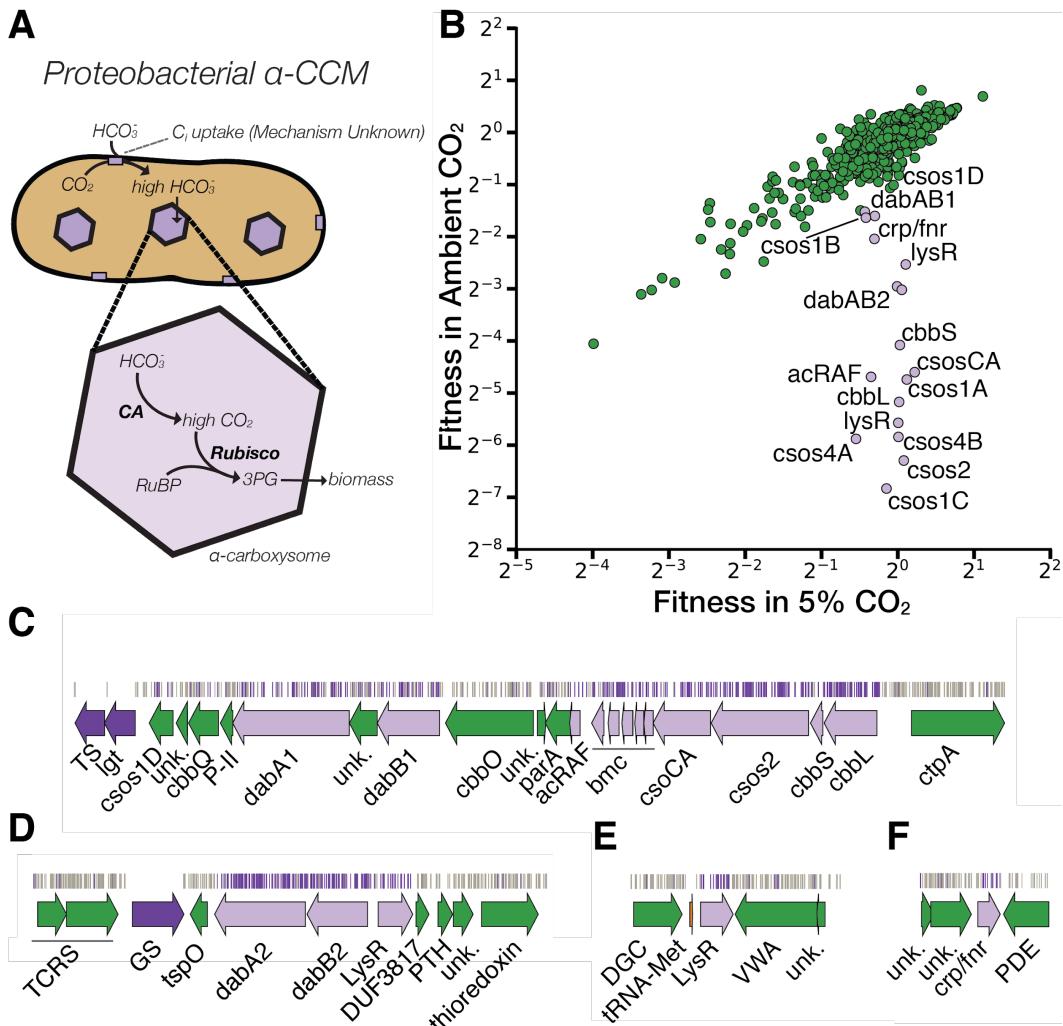
569



570

571 **Figure 1. Transposon mutagenesis reveals the essential gene set of a chemolithoautotrophic**  
 572 **organism. A.** Schematic depicting the generation and screening of the RB-TnSeq library. Transposons  
 573 were inserted into the *Hne*a genome by conjugation with an *E. coli* donor strain. The transposon contains  
 574 a random 20 base pair barcode (yellow) and a kanamycin selection marker (green). Selection for colonies  
 575 containing insertions was performed in the presence of kanamycin at 5% CO<sub>2</sub> and insertions were  
 576 mapped by sequencing as described in the Methods. Subsequent screens were carried out as bulk  
 577 competition assays and quantified by BarSeq. **B.** Insertions and essential genes are well-distributed  
 578 throughout the *Hne*a genome. The outer track (blue) is a histogram of the number of barcodes that were  
 579 mapped to a 1 kb window. The inner track annotates essential genes in purple. The pie chart shows the  
 580 percentages of the genome called essential (purple), ambiguous (orange), and nonessential (green). **C.**  
 581 Representative essential genes and nonessential genes in the *Hne*a genome. The blue track indicates  
 582 the presence of an insertion. Genes in purple were called essential and genes in green are nonessential.  
 583 Genes labeled “unk.” are hypothetical proteins. The first genomic locus contains 5 essential genes  
 584 involved in glycolysis or the CBB cycle including pyruvate kinase (pyk) and transketolase (tkt). The 8  
 585 essential genes in the second locus encoding 30S and 50S subunits of the ribosome, the secY secretory

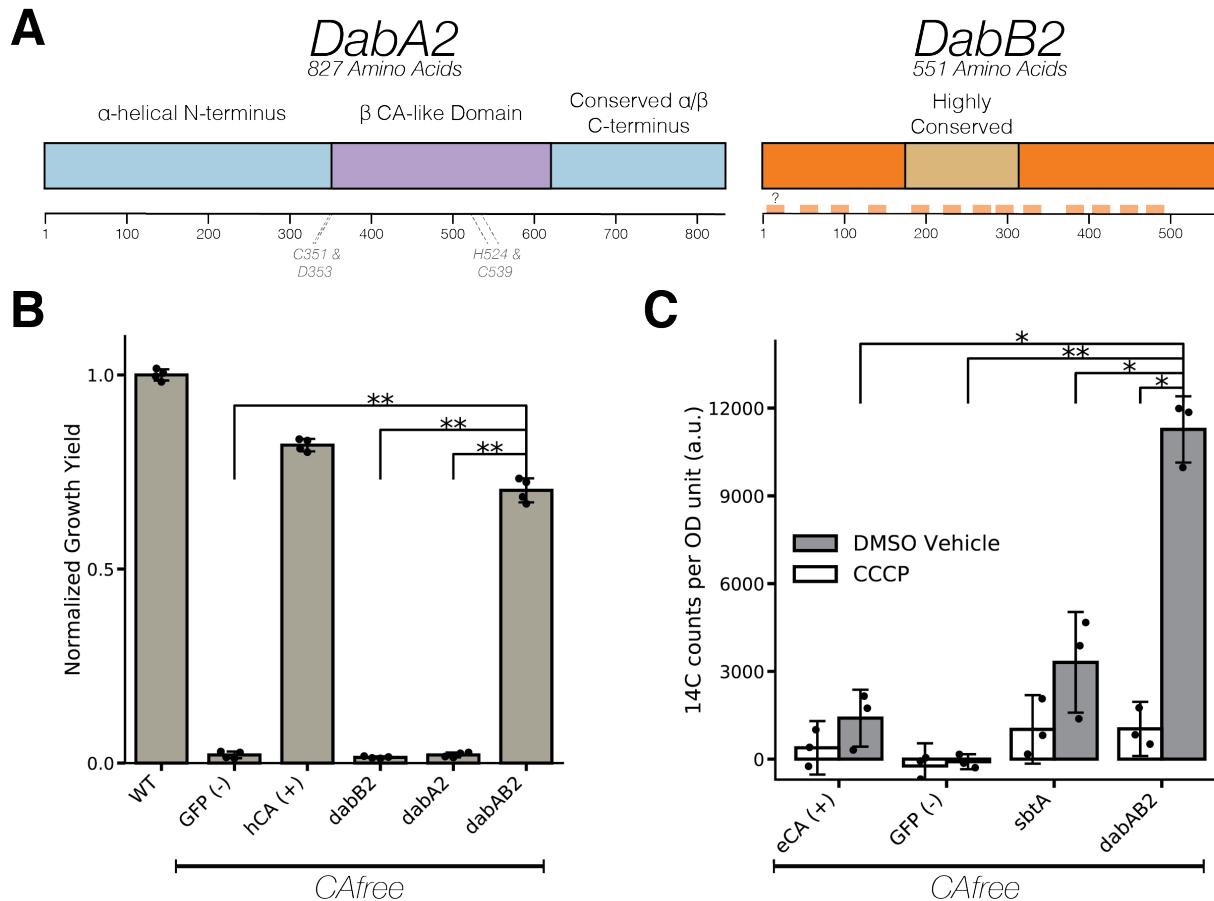
586 channel, and an RNA polymerase subunit. Essential genes in the third example locus include  
 587 topoisomerase and DNA polymerase III  $\beta$ . A full analysis with gene names is in Figure S1 and essentiality  
 588 information for every gene can be found in supplemental file 2.



589

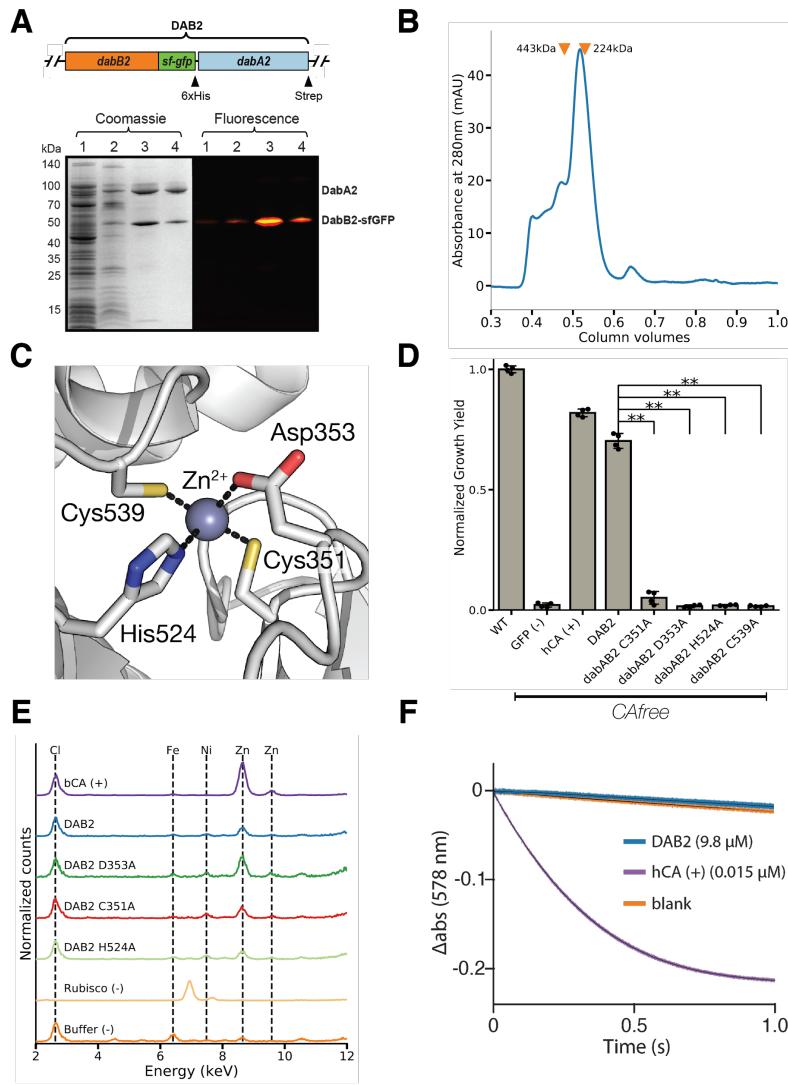
590 **Figure 2. A systematic screen for high  $CO_2$ -requiring mutants identifies genes putatively**  
 591 **associated with the CCM. A.** Simplified model of the  $\alpha$ -CCM of chemolithoautotrophic proteobacteria.  
 592 Inorganic carbon is concentrated via an unknown mechanism, producing a high cytosolic  $HCO_3^-$   
 593 concentration. High cytosolic  $HCO_3^-$  is converted into high carboxysomal  $CO_2$  by CA, which is localized  
 594 only to the carboxysome. **B.** Fitness effects of gene knockouts in 5%  $CO_2$  as compared to ambient  $CO_2$ .  
 595 Data is from one of two replicates of BarSeq. The effects of single transposon insertions into a gene are  
 596 averaged to produce the gene-level fitness value plotted. We define HCR mutants as those displaying a  
 597 twofold fitness defect in ambient  $CO_2$  relative to 5%  $CO_2$  in both replicates. HCR genes are colored light  
 598 purple. Data from both replicates and the associated standard errors are shown in Figure S2 and in  
 599 supplemental file 3. Panels **C-F** show regions of the *Hnea* genome containing genes annotated as HCR  
 600 in panel A. Essential genes are in dark purple, HCR genes are in light purple, and other genes are in

601 green. The top tracks show the presence of an insertion in that location. Insertions are colored grey  
602 unless they display a twofold or greater fitness defect in ambient CO<sub>2</sub>, in which case they are colored light  
603 purple. **C.** The gene cluster containing the carboxysome operon and a second CCM-associated operon.  
604 This second operon contains acRAF, a Form IC associated cbbOQ-type Rubisco activase and *dabAB1*.  
605 **D.** The DAB2 operon and surrounding genomic context. **E.** The genomic context of a lysR-type  
606 transcriptional regulator that shows an HCR phenotype. **F** Genomic context of a crp/fnr-type  
607 transcriptional regulator that displays an HCR phenotype. Genes labeled “unk.” are hypothetical proteins.  
608 Full gene names are given in Figure S3. **Accession numbers and gi numbers for selected genes can be**  
609 **found in Table S1.**



610  
611 **Figure 3. The DABs catalyze active transport of  $\text{C}_i$  and are energized by a cation gradient. A.**  
612 Diagrammatic representation of DabA2 and DabB2 based on bioinformatic annotation. The four predicted  
613 active site residues (C351, D353, H524, C539) are marked on the primary amino acid sequence. Amino  
614 acid numbers are marked below each gene and predicted transmembrane helices are marked in light  
615 orange. **B.** DAB2 was tested for ability to rescue growth of CAfree *E. coli* in ambient  $\text{CO}_2$  conditions.  
616 Expression of the full operon (DabAB2) rescues growth, as does the positive control, and human carbonic  
617 anhydrase II (hCA). Error bars represent standard deviations of 4 replicate cultures. **C.** CAfree *E. coli*  
618 were tested for  $\text{C}_i$  uptake using the silicone-oil centrifugation method. Expression of DabAB2 produced a  
619 large and statistically significant increase in  $^{14}\text{C}$  uptake as compared to all controls. Moreover, treatment  
620 with the ionophore CCCP greatly reduces DabAB2-mediated  $^{14}\text{C}$  uptake, suggesting that DabAB2 is  
621 coupled to a cation gradient. *E. coli* CA (eCA) was used as a control for a non-vectorial CA.  
622 *Synechococcus elongatus* PCC 7942 sbtA was used as a known  $\text{C}_i$  transporter. GFP was used as a  
623 vector control. Error bars represent standard deviations of 3 technical replicates. In (B) and (C) “\*\*”  
624 denotes that the means are significantly different with Bonferroni corrected  $p < 0.05$  according to a two-  
625 tailed t-test. “\*\*\*” denotes  $p < 5 \times 10^{-4}$ . In panel B, dabAB2 has a larger rescue than GFP ( $t=42.6$ , corrected  
626  $p=3.4 \times 10^{-8}$ ), dabA2 ( $t=43.4$ , corrected  $p=3 \times 10^{-8}$ ), and dabB2 ( $t=44.5$ , corrected  $p=2.6 \times 10^{-8}$ ). In panel C,  
627 dabAB2 expressing cells treated with DMSO have greater uptake than dabAB2 expressing cells treated

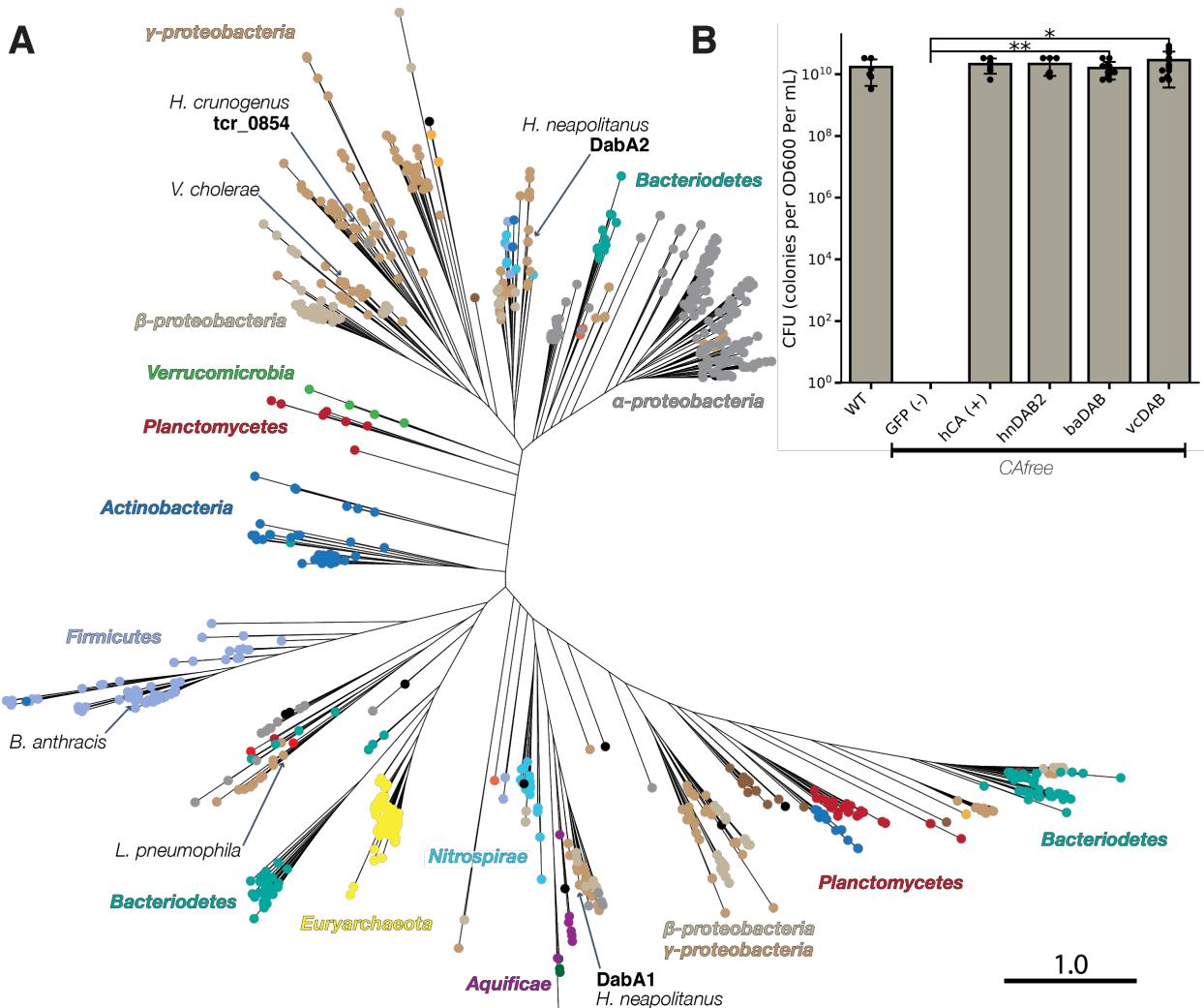
628 with CCCP ( $t=13.6$ , corrected  $p=6.8 \times 10^{-4}$ ), sbtA expressing cells treated with DMSO ( $t = 6.7$ , corrected  
 629  $p=10^{-2}$ ), GFP expressing cells treated with DMSO ( $t=17.1$ , corrected  $p=2.8 \times 10^{-4}$ ), or eCA expressing cells  
 630 treated with DMSO ( $t=11.5$ , corrected  $p=1.3 \times 10^{-3}$ ).



631

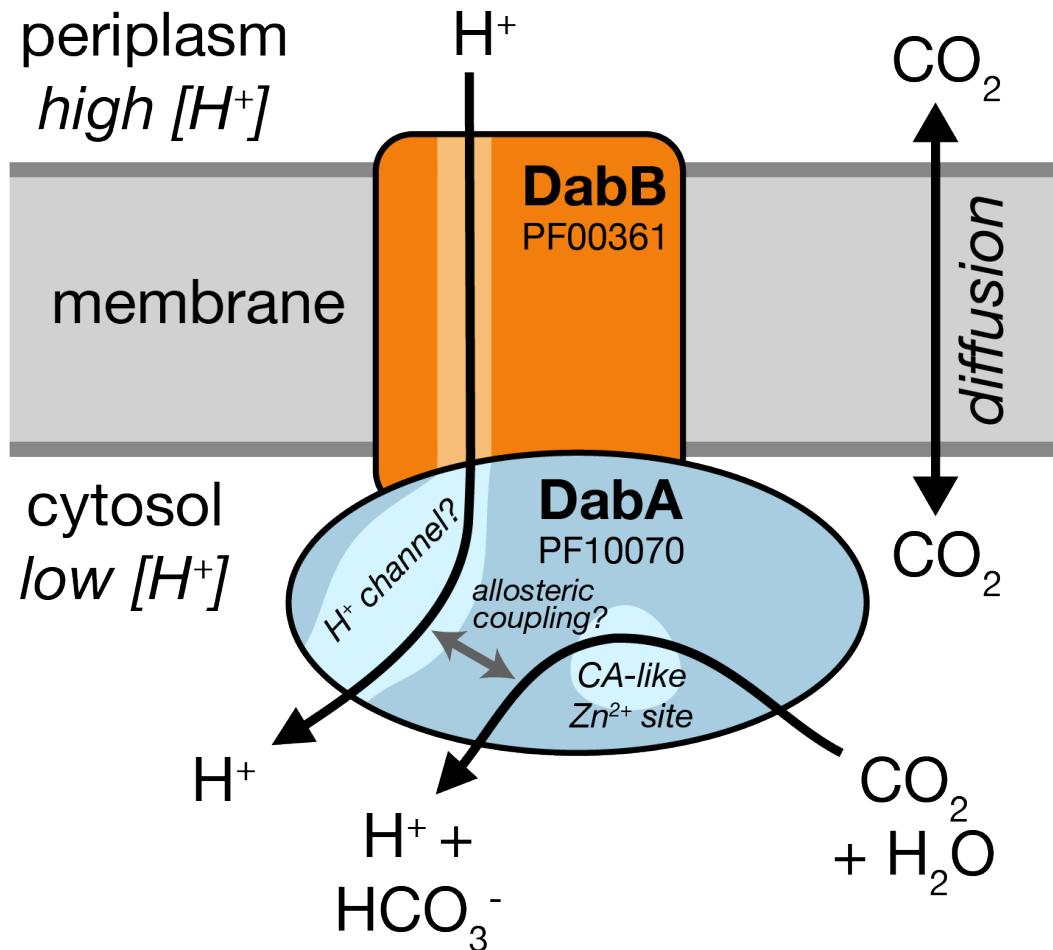
632 **Figure 4. DabA contains a β-CA-like active site but is not active outside of the membrane. A.**  
 633 Purification of DabAB2 complex from *E. coli*. DabA2 was C-terminally tagged to a Strep-tag and DabB2  
 634 was C-terminally tagged with sf-GFP and a 6xHis-tag. Purification was monitored using SDS-PAGE  
 635 imaged with fluorescence (right view) before coomassie staining (left view). Lane 1: clarified lysate; 2:  
 636 solubilized membranes; 3: Ni-NTA resin eluent; 4: strep-tactin resin eluent. DabA2 and DabB2 co-purify  
 637 as a single complex without any obvious interactors. **B.** Size-exclusion chromatogram of His/Strep  
 638 purified DabAB2 with retention volumes (orange arrows) and molecular weights (kDa) indicated for  
 639 standard samples (apo ferritin, 443 kDa; β-amylase, 224 kDa). DabAB2 runs with a mass of ~270 kDa,  
 640 which is likely an oligomer of DabA and DabB. **C.** Structural model of the DabA2 active site based on a β-  
 641 CA of *E. coli* (PDB 1I6P). Typical β-CAs rely on two cysteine and one histidine residues to bind Zn<sup>2+</sup>. The  
 642 aspartic acid coordinates Zn<sup>2+</sup> but is likely displaced during catalysis<sup>43</sup>. **D.** Alanine mutants of the putative

643 DabA2 active site residues (C351A, t=54.3, p=1.1\*10<sup>-8</sup>; D353A, t=144, p=3.1\*10<sup>-11</sup>; H524A, t=44,  
644 p=3.7\*10<sup>-8</sup>; C539A, t=44.3, p=3.5\*10<sup>-8</sup>;) abrogate rescue of CAfree *E. coli* compared to wild-type dabAB2.  
645 Error bars give standard deviations of four replicates. “\*\*” denotes that means differ with bonferroni  
646 corrected p < 0.05 by a two-tailed **t-test**, and “\*\*\*” denotes p < 5X10<sup>-4</sup>. **E.** X-ray fluorescence data indicate  
647 that DabAB2 binds zinc like all known β-CAs. Single mutations to the active site do not abrogate zinc  
648 binding. **F.** Purified DabAB2 does not display any obvious CA activity despite being present in 650-fold  
649 excess over the positive control (Human carbonic anhydrase II, hCA) in our assays.  
650



653 **Figure 5. DAB operons are widespread among prokaryotes. A.** Approximate maximum likelihood  
 654 phylogenetic tree of DabA homologs associated with PF10070.9 (Methods). DabA homologs are found in  
 655 > 15 prokaryotic clades, including archaea. *Hne*a DabA1 and DabA2 represent two different groupings  
 656 that are commonly found in proteobacteria. Inspecting the tree reveals several likely incidents of  
 657 horizontal transfer, e.g. between Proteobacteria and Firmicutes, Nitrospirae and Actinobacteria.  
 658 Moreover, the genomes of several known pathogens contain a high-confidence DabA homolog, including  
 659 *B. anthracis*, *V. cholerae*, and *L. pneumophila*. Detailed annotations are given in Figure S9. Scale bar  
 660 indicates one substitution per site. **B.** Functional DABs are found in human pathogens. Colony forming  
 661 units per OD<sub>600</sub> per ml were measured on LB plates with induction in air. DAB operons from *B. anthracis*  
 662 (baDAB,  $t=5.98$ ,  $p=1.8 \times 10^{-4}$ ) and *V. cholerae* (vcDAB,  $t=3.97$ ,  $p=4.4 \times 10^{-3}$ ) rescued growth of CAfree cells.  
 663 The *Hne*a operon DAB2 is abbreviated as hnDAB2. Error bars represent the standard deviation of 6  
 664 replicate platings for WT, GFP (-), hCA (+), and hnDAB2. Error bars represent standard deviations of 12

665 replicate platings for baDAB and vcDAB. “\*” denotes that means differ with Bonferroni corrected  $p < 0.05$   
666 by a two-tailed t-test, and “\*\*” denotes  $p < 5 \times 10^{-4}$ . CFU plates are shown in Figure S10.  
667



668  
669 **Figure 6. A speculative model of the unidirectional energy-coupled CA activity of DAB complexes.**  
670 We propose that DabAB complexes couple CA activity of DabA to a cation gradient across the cell  
671 membrane, producing unidirectional hydration of  $CO_2$  to  $HCO_3^-$ . The cation gradient could be  $H^+$  or  $Na^+$ .  
672 Energy-coupled CA activity is required for the DABs role as a  $C_i$  uptake system in the proteobacterial  
673 CCM, as discussed in the text. Because it appears that DabAB2 is not active as a purified complex  
674 outside of the membrane, it is assumed protein tightly couples the inflow of cations with  $CO_2$  hydration so  
675 that there is no “slippage.” Indeed, slippage - i.e., uncoupled CA activity - would be counterproductive for  
676 CCM function<sup>7,14</sup>. Notably,  $Zn^{2+}$  binding by the active site aspartic acid of type II  $\beta$ -CAs (D353 in DabA2,  
677 Figure 4A) is thought to allosterically regulate activity<sup>43</sup>. This Asp-mediated activity switch could,  
678 therefore, provide a means for allosteric coupling of a  $\beta$ -CA active site to distal ion transport.

679 **References**

- 680 1. Bathellier, C., Tcherkez, G., Lorimer, G. H. & Farquhar, G. D. Rubisco is not really so bad.  
681 *Plant Cell Environ.* **41**, 705–716 (2018).
- 682 2. Flamholz, A. *et al.* Revisiting tradeoffs in Rubisco kinetic parameters. *bioRxiv* 470021  
683 (2018). doi:10.1101/470021
- 684 3. Tcherkez, G. The mechanism of Rubisco-catalysed oxygenation. *Plant Cell Environ.* **39**,  
685 983–997 (2016).
- 686 4. Bauwe, H., Hagemann, M. & Fernie, A. R. Photorespiration: players, partners and origin.  
687 *Trends Plant Sci.* **15**, 330–336 (2010).
- 688 5. Tcherkez, G. G. B., Farquhar, G. D. & Andrews, T. J. Despite slow catalysis and confused  
689 substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly  
690 optimized. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 7246–7251 (2006).
- 691 6. Savir, Y., Noor, E., Milo, R. & Tlusty, T. Cross-species analysis traces adaptation of  
692 Rubisco toward optimality in a low-dimensional landscape. *Proc. Natl. Acad. Sci. U. S. A.*  
693 **107**, 3475–3480 (2010).
- 694 7. Mangan, N. M., Flamholz, A., Hood, R. D., Milo, R. & Savage, D. F. pH determines the  
695 energetic efficiency of the cyanobacterial CO<sub>2</sub> concentrating mechanism. *Proc. Natl. Acad.*  
696 *Sci. U. S. A.* **113**, E5354–62 (2016).
- 697 8. Raven, J. A., Beardall, J. & Sánchez-Baracaldo, P. The possible evolution and future of  
698 CO<sub>2</sub>-concentrating mechanisms. *J. Exp. Bot.* **68**, 3701–3716 (2017).
- 699 9. Rae, B. D., Long, B. M., Badger, M. R. & Price, G. D. Functions, compositions, and  
700 evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate  
701 CO<sub>2</sub> fixation in cyanobacteria and some proteobacteria. *Microbiol. Mol. Biol. Rev.* **77**, 357–  
702 379 (2013).
- 703 10. Long, B. M., Rae, B. D., Rolland, V., Förster, B. & Price, G. D. Cyanobacterial CO<sub>2</sub>-

- 704       concentrating mechanism components: function and prospects for plant metabolic  
705       engineering. *Curr. Opin. Plant Biol.* **31**, 1–8 (2016).
- 706       11. Price, G. D., Badger, M. R. & von Caemmerer, S. The prospect of using cyanobacterial  
707       bicarbonate transporters to improve leaf photosynthesis in C3 crop plants. *Plant Physiol.*  
708       **155**, 20–26 (2011).
- 709       12. McGrath, J. M. & Long, S. P. Can the cyanobacterial carbon-concentrating mechanism  
710       increase photosynthesis in crop species? A theoretical analysis. *Plant Physiol.* **164**, 2247–  
711       2261 (2014).
- 712       13. Long, B. M. *et al.* Carboxysome encapsulation of the CO<sub>2</sub>-fixing enzyme Rubisco in  
713       tobacco chloroplasts. *Nat. Commun.* **9**, 3570 (2018).
- 714       14. Price, G. D. & Badger, M. R. Expression of Human Carbonic Anhydrase in the  
715       Cyanobacterium Synechococcus PCC7942 Creates a High CO<sub>2</sub>-Requiring Phenotype  
716       Evidence for a Central Role for Carboxysomes in the CO<sub>2</sub> Concentrating Mechanism. *Plant*  
717       *Physiol.* **91**, 505–513 (1989).
- 718       15. Hopkinson, B. M., Young, J. N., Tansik, a. L. & Binder, B. J. The Minimal CO<sub>2</sub>-  
719       Concentrating Mechanism of Prochlorococcus spp. MED4 Is Effective and Efficient. *Plant*  
720       *Physiol.* **166**, 2205–2217 (2014).
- 721       16. Whitehead, L., Long, B. M., Price, G. D. & Badger, M. R. Comparing the in Vivo Function of  
722       α-Carboxysomes and β-Carboxysomes in Two Model Cyanobacteria. *Plant Physiol.* **165**,  
723       398–411 (2014).
- 724       17. Holthuijzen, Y. A., van Dissel-Emiliani, F. F. M., Kuenen, J. G. & Konings, W. N. Energetic  
725       aspects of CO<sub>2</sub> uptake in *Thiobacillus neapolitanus*. *Arch. Microbiol.* **147**, 285–290 (1987).
- 726       18. Price, G. D. & Badger, M. R. Isolation and characterization of high CO<sub>2</sub>-requiring-mutants  
727       of the cyanobacterium Synechococcus PCC7942: two phenotypes that accumulate  
728       inorganic carbon but are apparently unable to generate CO<sub>2</sub> within the carboxysome. *Plant*  
729       *Physiol.* **91**, 514–525 (1989).

- 730 19. Marcus, Y., Schwarz, R., Friedberg, D. & Kaplan, A. High CO<sub>2</sub> Requiring Mutant of  
731 *Anacystis nidulans* R(2). *Plant Physiol.* **82**, 610–612 (1986).
- 732 20. Bonacci, W. *et al.* Modularity of a carbon-fixing protein organelle. *Proc. Natl. Acad. Sci. U.*  
733 *S. A.* **109**, 478–483 (2012).
- 734 21. Jorda, J., Lopez, D., Wheatley, N. M. & Yeates, T. O. Using comparative genomics to  
735 uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci.* **22**, 179–  
736 195 (2013).
- 737 22. Axen, S. D., Erbilgin, O. & Kerfeld, C. A. A taxonomy of bacterial microcompartment loci  
738 constructed by a novel scoring method. *PLoS Comput. Biol.* **10**, e1003898 (2014).
- 739 23. Shibata, M., Ohkawa, H., Katoh, H., Shimoyama, M. & Ogawa, T. Two CO<sub>2</sub> uptake  
740 systems in cyanobacteria: four systems for inorganic carbon acquisition in *Synechocystis*  
741 sp. strain PCC6803. *Funct. Plant Biol.* **29**, 123–129 (2002).
- 742 24. Price, G. D. Inorganic carbon transporters of the cyanobacterial CO<sub>2</sub> concentrating  
743 mechanism. *Photosynth. Res.* **109**, 47–57 (2011).
- 744 25. Heinhorst, S., Cannon, G. C. & Shively, J. M. Carboxysomes and Carboxysome-like  
745 Inclusions. in *Complex Intracellular Structures in Prokaryotes* (ed. Shively, J. M.) 141–165  
746 (Springer Berlin Heidelberg, 2006).
- 747 26. Shively, J. M., Ball, F., Brown, D. H. & Saunders, R. E. Functional organelles in  
748 prokaryotes: polyhedral inclusions (carboxysomes) of *Thiobacillus neapolitanus*. *Science*  
749 **182**, 584–586 (1973).
- 750 27. Cannon, G. C. *et al.* Microcompartments in prokaryotes: carboxysomes and related  
751 polyhedra. *Appl. Environ. Microbiol.* **67**, 5351–5361 (2001).
- 752 28. Mangiapia, M. *et al.* Proteomic and mutant analysis of the CO<sub>2</sub> concentrating mechanism  
753 of hydrothermal vent chemolithoautotroph *Thiomicrospira crunogena*. *J. Bacteriol.* (2017).  
754 doi:10.1128/JB.00871-16
- 755 29. Scott, K. M. *et al.* Genomes of ubiquitous marine and hypersaline *Hydrogenovibrio*,

- 756        Thiomicrorhabdus and Thiomicrospira spp. encode a diversity of mechanisms to sustain  
757        chemolithoautotrophy in heterogeneous environments. *Environ. Microbiol.* **20**, 2686–2708  
758        (2018).
- 759        30. Scott, K. M. *et al.* Diversity in CO<sub>2</sub>-Concentrating Mechanisms among  
760        Chemolithoautotrophs from the Genera Hydrogenovibrio, Thiomicrorhabdus, and  
761        Thiomicrospira, Ubiquitous in Sulfidic Habitats Worldwide. *Appl. Environ. Microbiol.* **85**,  
762        (2019).
- 763        31. Wetmore, K. M. *et al.* Rapid quantification of mutant fitness in diverse bacteria by  
764        sequencing randomly bar-coded transposons. *MBio* **6**, e00306–15 (2015).
- 765        32. Chaijaraspong, T. *et al.* Programmed Ribosomal Frameshifting Mediates Expression of  
766        the  $\alpha$ -Carboxysome. *J. Mol. Biol.* **428**, 153–164 (2016).
- 767        33. Cai, F. *et al.* The pentameric vertex proteins are necessary for the icosahedral  
768        carboxysome shell to function as a CO<sub>2</sub> leakage barrier. *PLoS One* **4**, e7521 (2009).
- 769        34. Roberts, E. W., Cai, F., Kerfeld, C. A., Cannon, G. C. & Heinhorst, S. Isolation and  
770        characterization of the Prochlorococcus carboxysome reveal the presence of the novel  
771        shell protein CsoS1D. *J. Bacteriol.* **194**, 787–795 (2012).
- 772        35. Wheatley, N. M., Sundberg, C. D., Gidaniyan, S. D., Cascio, D. & Yeates, T. O. Structure  
773        and Identification of a Pterin Dehydratase-like Protein as a Ribulose-bisphosphate  
774        Carboxylase/Oxygenase (RuBisCO) Assembly Factor in the  $\alpha$ -Carboxysome. *J. Biol.*  
775        *Chem.* **289**, 7973–7981 (2014).
- 776        36. Aigner, H. *et al.* Plant RuBisCo assembly in *E. coli* with five chloroplast chaperones  
777        including BSD2. *Science* **358**, 1272–1278 (2017).
- 778        37. Mueller-Cajar, O. The Diverse AAA+ Machines that Repair Inhibited Rubisco Active Sites.  
779        *Front Mol Biosci* **4**, 31 (2017).
- 780        38. Krulwich, T. A., Hicks, D. B. & Ito, M. Cation/proton antiporter complements of bacteria: why  
781        so large and diverse? *Mol. Microbiol.* **74**, 257–260 (2009).

- 782 39. Merlin, C. & Masters, M. Why is carbonic anhydrase essential to *Escherichia coli*? *J.*  
783 *Bacteriol.* **185**, (2003).
- 784 40. Du, J., Förster, B., Rourke, L., Howitt, S. M. & Price, G. D. Characterisation of  
785 Cyanobacterial Bicarbonate Transporters in *E. coli* Shows that SbtA Homologs Are  
786 Functional in This Heterologous Expression System. *PLoS One* **9**, e115905 (2014).
- 787 41. Cronk, J. D., Endrizzi, J. a., Cronk, M. R., O'neill, J. W. & Zhang, K. Y. Crystal structure of  
788 *E. coli* beta-carbonic anhydrase, an enzyme with an unusual pH-dependent activity. *Protein*  
789 *Sci.* **10**, 911–922 (2001).
- 790 42. Krishnamurthy, V. M. *et al.* Carbonic anhydrase as a model for biophysical and physical-  
791 organic studies of proteins and protein- ligand binding. *Chem. Rev.* **108**, 946–1051 (2008).
- 792 43. Cronk, J. D. *et al.* Identification of a novel noncatalytic bicarbonate binding site in  
793 eubacterial beta-carbonic anhydrase. *Biochemistry* **45**, 4351–4361 (2006).
- 794 44. Shibata, M. *et al.* Distinct constitutive and low-CO<sub>2</sub>-induced CO<sub>2</sub> uptake systems in  
795 cyanobacteria: genes involved and their phylogenetic relationship with homologous genes  
796 in other organisms. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 11789–11794 (2001).
- 797 45. Maeda, S.-I., Badger, M. R. & Price, G. D. Novel gene products associated with NdhD3/D4-  
798 containing NDH-1 complexes are involved in photosynthetic CO<sub>2</sub> hydration in the  
799 cyanobacterium, *Synechococcus* sp. PCC7942. *Mol. Microbiol.* **43**, 425–435 (2002).
- 800 46. Battchikova, N., Eisenhut, M. & Aro, E. M. Cyanobacterial NDH-1 complexes: Novel  
801 insights and remaining puzzles. *Biochimica et Biophysica Acta - Bioenergetics* **1807**, 935–  
802 944 (2011).
- 803 47. Antonovsky, N. *et al.* Sugar Synthesis from CO<sub>2</sub> in *Escherichia coli*. *Cell* **166**, 115–125  
804 (2016).
- 805 48. Aguilera, J., Van Dijken, J. P., De Winde, J. H. & Pronk, J. T. Carbonic anhydrase  
806 (Nce103p): an essential biosynthetic enzyme for growth of *Saccharomyces cerevisiae* at  
807 atmospheric carbon dioxide pressure. *Biochem. J.* **391**, 311–316 (2005).

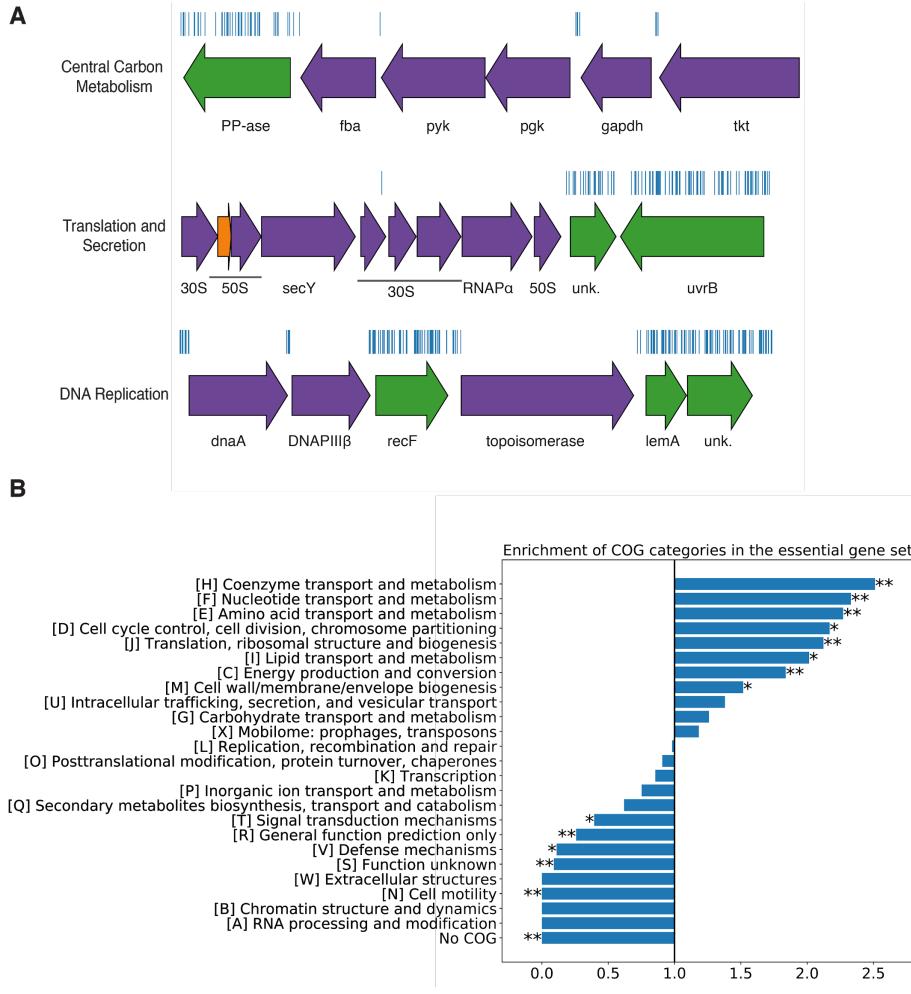
- 808 49. Sirard, J. C., Mock, M. & Fouet, A. The three *Bacillus anthracis* toxin genes are  
809 coordinately regulated by bicarbonate and temperature. *J. Bacteriol.* **176**, 5188–5192  
810 (1994).
- 811 50. Abuaita, B. H. & Withey, J. H. Bicarbonate Induces *Vibrio cholerae* virulence gene  
812 expression by enhancing ToxT activity. *Infect. Immun.* **77**, 4111–4120 (2009).
- 813 51. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout  
814 mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
- 815 52. Rubin, B. E. *et al.* The essential gene set of a photosynthetic organism. *Proceedings of the*  
816 *National Academy of Sciences* 201519220 (2015).
- 817 53. Oakes, B. L., Nadler, D. C. & Savage, D. F. Chapter Twenty-Three - Protein Engineering of  
818 Cas9 for Enhanced Function. in *Methods in Enzymology* (ed. Sontheimer, J. A. D. A. E. J.)  
819 **546**, 491–511 (Academic Press, 2014).
- 820 54. Dobrinski, K. P., Longo, D. L. & Scott, K. M. The Carbon-Concentrating Mechanism of the  
821 Hydrothermal Vent Chemolithoautotroph *Thiomicrospira crunogena*. *J. Bacteriol.* **187**,  
822 5761–5766 (2005).
- 823 55. Zallot, R., Oberg, N. O. & Gerlt, J. A. ‘Democratized’ genomic enzymology web tools for  
824 functional assignment. *Curr. Opin. Chem. Biol.* **47**, 77–85 (2018).
- 825 56. Dehal, P. S. *et al.* MicrobesOnline: an integrated portal for comparative and functional  
826 genomics. *Nucleic Acids Res.* **38**, D396–400 (2010).
- 827 57. Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein  
828 sequences. *Protein Sci.* **27**, 135–145 (2018).
- 829 58. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees  
830 with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
- 831 59. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and  
832 annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–5 (2016).
- 833 60. Slabinski, L. *et al.* XtalPred: a web server for prediction of protein crystallizability.

- 834        *Bioinformatics* **23**, 3403–3405 (2007).
- 835        61. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web  
836            portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
- 837        62. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein  
838            structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).
- 839        63. Newby, Z. E. R. *et al.* A general protocol for the crystallization of membrane proteins for X-  
840            ray structural investigation. *Nat. Protoc.* **4**, 619–637 (2009).
- 841        64. Khalifah, R. G. The Carbon Dioxide Hydration Activity of Carbonic Anhydrase. *J. Biol.*  
842            *Chem.* **246**, 2561–2573 (1971).

843

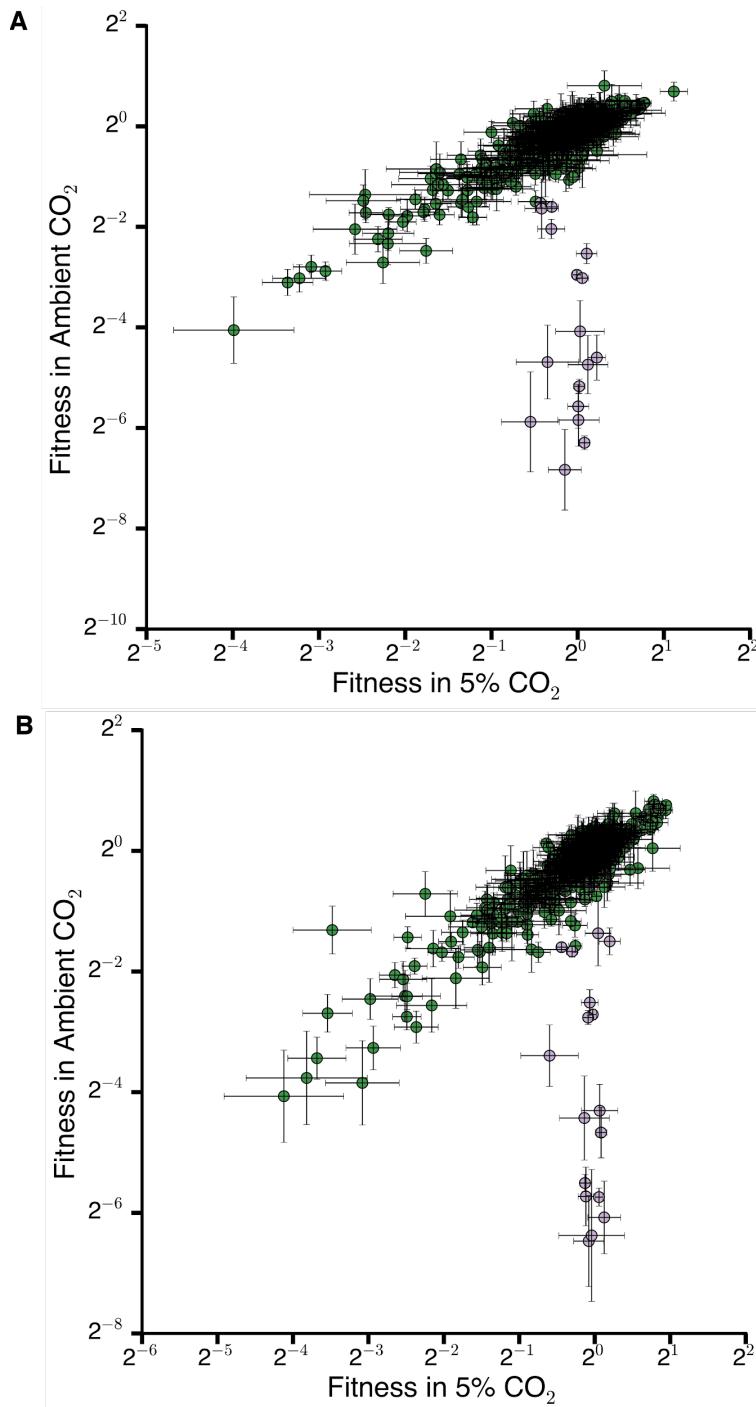
844 **Supplementary Information:**

- 845 **Supplemental File 1.** Important strains and reagents.
- 846 **Supplemental File 2.** Transposon insertion information and essentiality determination by gene.
- 847 **Supplemental File 3.** Fitness effects and HCR phenotype by gene.
- 848 **Supplemental File 4.** Genes used to generate figure S4A.
- 849 **Supplemental File 5.** Genes used to generate figure 5A.



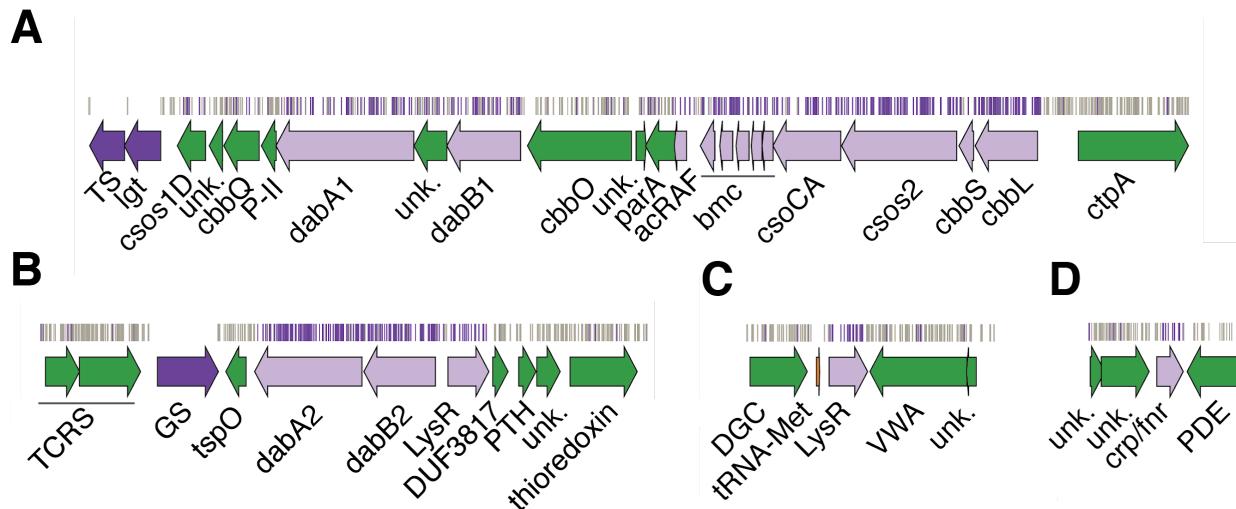
850  
851 **Figure S1 The essential gene set is enriched for COGs associated with essential cellular  
852 processes.** A. Representative essential genes and nonessential genes in the *Hnea* genome. The blue  
853 track indicates the presence of an insertion. Genes in purple were called essential and genes in green are  
854 nonessential. Genes labeled "unk." are hypothetical proteins. The first genomic locus contains 5 essential  
855 genes involved in glycolysis or the CBB cycle including pyruvate kinase (pyk) and transketolase (tkt). The  
856 8 essential genes in the second locus encode 30S and 50S subunits of the ribosome, the secY secretory  
857 channel, and an RNA polymerase subunit. Essential genes in the third example locus include  
858 topoisomerase and DNA polymerase III β. B. COG enrichments were calculated by dividing the fraction of  
859 genes in the essential gene set associated with this COG category by the fraction of genes in the genome  
860 associated with this category. \*\* denotes that this COG is enriched (or depleted) with Bonferroni  
861 corrected  $p < 0.05$  by a hypergeometric test, and \*\*\* denotes  $p < 5 \times 10^{-4}$ . Exact p values are as follows  
862 for each category, No COG:  $7 \times 10^{-68}$ , C:  $1.1 \times 10^{-5}$ , D:  $1 \times 10^{-2}$ , E:  $< 7 \times 10^{-68}$ , F:  $4.3 \times 10^{-6}$ , H:  $< 7 \times 10^{-68}$ , I:  $6.3 \times 10^{-4}$ ,  
863 J:  $< 7 \times 10^{-68}$ , M:  $6.9 \times 10^{-3}$ , N:  $6.5 \times 10^{-6}$ , R:  $6.17 \times 10^{-5}$ , S:  $7.7 \times 10^{-8}$ , T:  $2.8 \times 10^{-2}$ , V:  $1.1 \times 10^{-2}$ . In panel A, the following  
864 abbreviations are used: exopolyphosphatase (PP-ase), fructose-bisphosphate aldolase class II (fba),  
865 pyruvate kinase (pyk), phosphoglycerate kinase (pgk), type I glyceraldehyde-3-phosphate dehydrogenase  
866 (gapdh), transketolase (tkt), 30S ribosomal protein (30S), 50S ribosomal protein (50S), preprotein

867 translocase subunit SecY (SecY), DNA-directed RNA polymerase subunit alpha (RNAP $\alpha$ ), hypothetical  
 868 protein (unk.), excinuclease ABC subunit UvrB (UvrB), chromosomal replication initiator protein dnaA  
 869 (dnaA), DNA polymerase III subunit beta (DNAPIII $\beta$ ), DNA replication and repair protein recF (recF), DNA  
 870 topoisomerase (ATP-hydrolyzing) subunit B (topoisomerase), lemA family protein (LemA).



871  
 872 **Figure S2 Gene fitnesses measurements for each replicates.** Fitness effects of gene knockouts in  
 873  $5\% \text{CO}_2$  as compared to ambient  $\text{CO}_2$ . The effects of single transposon insertions into a gene are

874 averaged to produce the gene-level fitness value plotted. Error bars represent one standard error of the  
 875 mean. We define HCR mutants as those displaying a twofold fitness defect in ambient CO<sub>2</sub> relative to 5%  
 876 CO<sub>2</sub>. HCR genes are colored light purple. Panel **A** contains data from the first replicate experiment and  
 877 panel **B** contains data from the second replicate experiment.

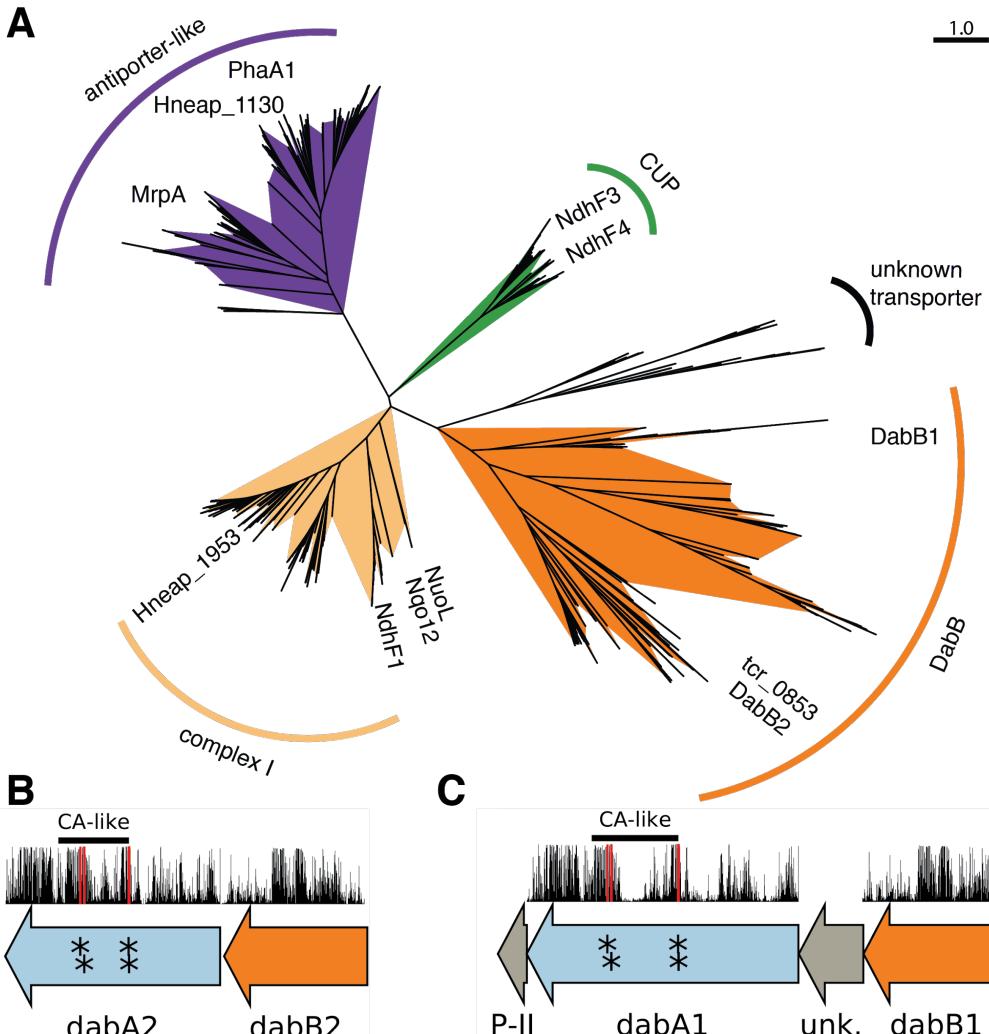


878  
 879 **Figure S3 Genomic context of *Hneaa* HCR genes identified in our genome-wide screen.** Panels **A-D**  
 880 show regions of the *Hneaa* genome containing genes annotated as HCR. Essential genes are in dark  
 881 purple, HCR genes are in light purple, and other genes are in green. The top tracks show the presence of  
 882 an insertion in that location. Insertions are colored grey unless they display a twofold or greater  
 883 fitness defect in ambient CO<sub>2</sub>, in which case they are colored purple. **A.** The gene cluster containing the  
 884 carboxysome operon (HNEAP\_RS04660-HNEAP\_RS04620) and a second CCM-associated operon. This  
 885 second operon contains acRAF (HNEAP\_RS04615), a FormIC associated cbbOQ-type Rubisco activase  
 886 (HNEAP\_RS04575 and HNEAP\_RS04600), parA (HNEAP\_RS04610), P-II (HNEAP\_RS04580) and  
 887 dabAB1 (dabA1: HNEAP\_RS04585 and dabB1: HNEAP\_RS04620). **B.** The DAB2 operon and  
 888 surrounding genomic context (lysR: HNEAP\_RS01040, dabA2: HNEAP\_RS01030, and dabB2:  
 889 HNEAP\_RS01035). **C.** The genomic context of a lysR-type transcriptional regulator (HNEAP\_RS05490)  
 890 that shows an HCR phenotype. **D** Genomic context of a crp/fnr-type transcriptional regulator that displays  
 891 an HCR phenotype (HNEAP\_RS07320). Accession numbers and gi numbers for selected genes can be  
 892 found in Table S1. Abbreviations for Figure S3: thymidylate synthase (TS), prolipoprotein diacylglycerol  
 893 transferase (lgt), Rubisco activase Rubisco activase subunits (cbbOQ), nitrogen regulatory protein P-II (P-  
 894 II), ParA family protein (parA), csos1CAB and csos4AB (bmc), copper-translocating P-type ATPase  
 895 (ctpA), DNA-binding response regulator and two-component sensor histidine kinase (TCRS), glutamate--  
 896 ammonia ligase (GS), tryptophan-rich sensory protein (tspO), DUF3817 domain-containing protein  
 897 (DUF3817), aminoacyl-tRNA hydrolase (PTH), thioredoxin domain-containing protein (thioredoxin),  
 898 sensor domain-containing diguanylate cyclase (DGC), methionine tRNA (tRNA-Met), VWA domain-  
 899 containing protein (VWA), diguanylate phosphodiesterase (PDE).

Locus Id	NCBI Accession number	NCBI gi number	Gene description	Has HCR phenotype
HNEAP_RS01030	WP_012823110.1	502585319	DabA2	TRUE
HNEAP_RS01035	WP_012823111.1	502585320	DabB2	TRUE
HNEAP_RS01040	WP_012823112.1	502585321	LysR	TRUE
HNEAP_RS04565	WP_012823782.1	502586009	Csos1D	FALSE
HNEAP_RS04570	WP_012823783.1	502586011	unk.	FALSE
HNEAP_RS04575	WP_012823784.1	502586012	CbbQ	FALSE
HNEAP_RS04580	WP_012823785.1	502586013	p-II	FALSE
HNEAP_RS04585	WP_012823786.1	502586014	DabA1	TRUE
HNEAP_RS04590	WP_012823787.1	502586015	unk.	FALSE
HNEAP_RS04595	WP_012823788.1	502586016	DabB1	TRUE
HNEAP_RS04600	WP_012823789.1	502586017	CbbO	FALSE
HNEAP_RS04605	WP_041600361.1	753844744	unk.	FALSE
HNEAP_RS04610	WP_049772467.1	908628434	ParA	FALSE
HNEAP_RS04615	WP_012823792.1	502586020	acRAF	TRUE
HNEAP_RS04620	WP_012823793.1	502586021	Csos1B	TRUE
HNEAP_RS04625	WP_012823794.1	502586022	Csos1A	TRUE
HNEAP_RS04630	WP_012823795.1	502586023	Csos1C	TRUE
HNEAP_RS04635	WP_012823796.1	502586024	Csos4B	TRUE
HNEAP_RS04640	WP_012823797.1	502586025	Csos4A	TRUE
HNEAP_RS04645	WP_012823798.1	502586026	CsosCA	TRUE
HNEAP_RS04650	WP_081441107.1	1174219926	Csos2	TRUE
HNEAP_RS04655	WP_012823800.1	502586028	CbbS	TRUE
HNEAP_RS04660	WP_012823801.1	502586029	CbbL	TRUE
HNEAP_RS05490	WP_012823963.1	502586200	LysR	TRUE
HNEAP_RS07320	WP_081441122.1	1174219941	Crp/Fnr	TRUE

900  
901  
902

**Table S1 Genes from HCR operons.** This table includes genes from the the HCR operons with their phenotype and identifying information. "unk." indicates a hypothetical protein.



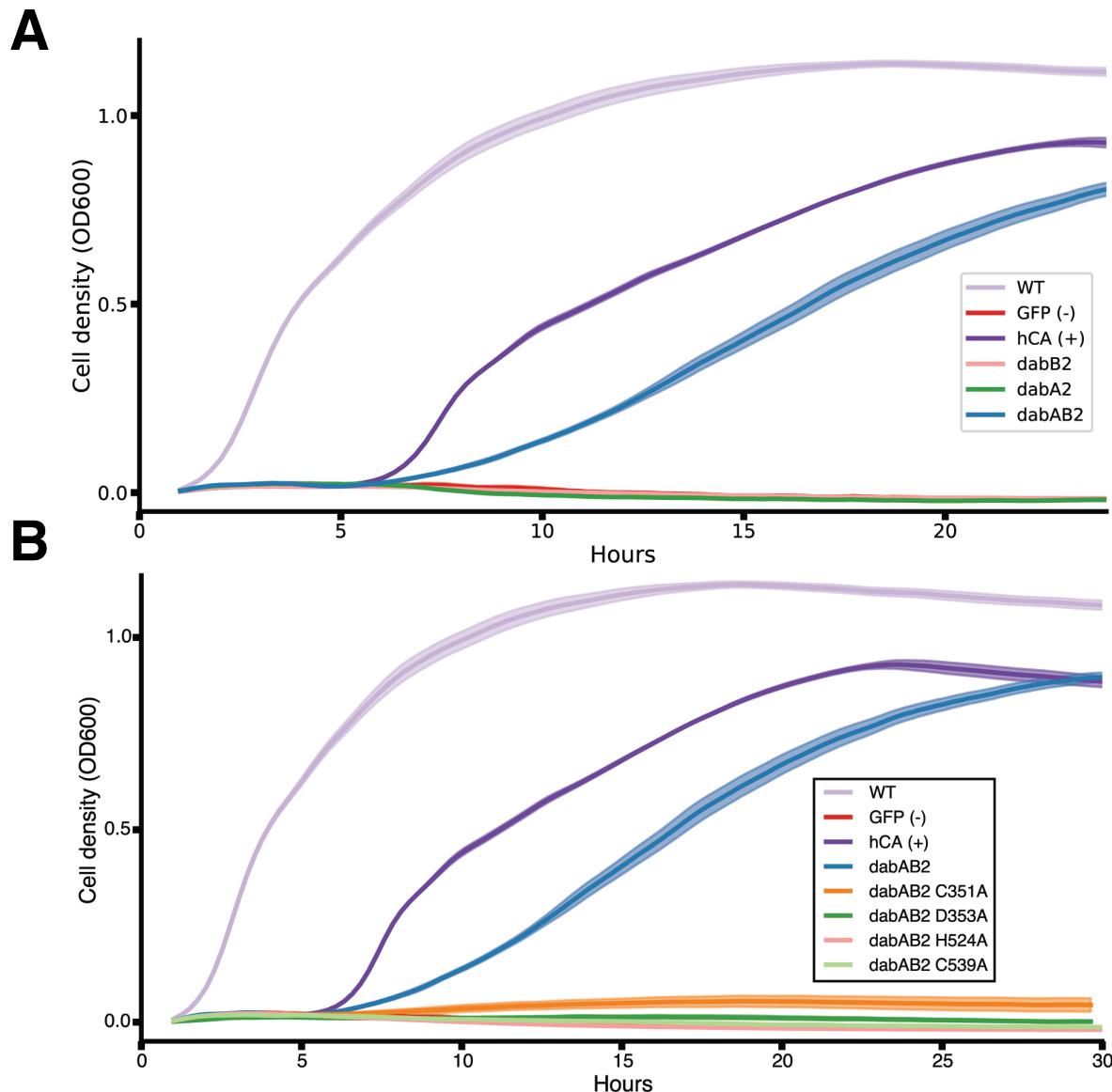
903  
904 **Figure S4 PF0361 contains multiple subfamilies, but some regions of DAB subunits are highly**  
905 **conserved.** A. PF0361 is a large and diverse protein family containing multiple subgroups with different  
906 documented activities. These subfamilies include Mrp-family cation antiporters, proton translocating  
907 subunits of complex I, membrane subunits of CUP (CO<sub>2</sub> uptake protein) complexes, and DabB proteins.  
908 These subfamilies are highly diverged and perform a variety of activities. This means that it is not  
909 possible to draw conclusions about the mechanism of DAB complexes just from their homology to  
910 PF0361. This panel contains an approximate maximum likelihood tree of PF0361 genes. Clades were  
911 colored according to the presence of genes with known functions. The purple clade contains the *Bacillus*  
912 *subtilis* and *Staphylococcus aureus* MrpA cation antiporter subunits and the *Sinorhizobium meliloti*  
913 antiporter PhaA1. The light orange clade contains the known cation translocating subunits of complex I:  
914 nuoL from *Escherichia coli*, Nqo12 from *Thermus thermophilus*, and NdhF1 from both *Synechococcus*  
915 *elongatus* PCC7942 and *Thermosynechococcus elongatus* BP-1. The green clade contains CUP-  
916 associated membrane subunits ndhF3 from both *Synechococcus elongatus* PCC7942 and  
917 *Thermosynechococcus elongatus* BP-1 and ndhF4 from the same two species. The dark orange  
918 clade includes DabB1-2 and tcr\_0853 from *Thiomicrospira crunogena*. We note that the clade containing  
919 DabB1-2 is distinct from that containing known complex I subunits or to mrp-family antiporters. This tree is

920 consistent with our model, where DabB is not bound to a redox-coupled complex but rather couples  
921 redox-independent cation transport to CA activity (as shown in Figure 5). No conclusions should be drawn  
922 from the number of sequences in each clade as an exhaustive search for homologs was not performed to  
923 ensure that all members of each clade are represented. Scale bar indicates one substitution per site. **B**  
924 and **C** As noted in the text and shown in Figure 2B, DAB1 is a segment of an 11-gene operon directly  
925 downstream of the carboxysome operon that contains CCM-associated genes. Both DAB1 (**B**) and DAB2  
926 (**C**) “operons” contain two distinct genes that we label DabB and DabA. DabA is annotated as Domain of  
927 Unknown Function 2309 (DUF2309, PFAM:PF10070) and appears to be a soluble protein. Approximately  
928 one third of dabA is distantly homologous to a type II β-CA. CA-like regions are marked with a line, and  
929 the four residues expected to be involved in binding the catalytic zinc ion are marked by asterisks. The  
930 height of the asterisks has been varied to make them distinguishable despite proximity in sequence  
931 space. DabB is homologous to a cation transporter in the same family as the H<sup>+</sup> pumping subunits of  
932 respiratory complex I (PFAM:PF00361). The DAB1 operon also contains a protein of unknown function  
933 between DabA1 and DabB1. This protein has distant homology to DabA1 but is truncated to half the  
934 length. Vertical bars above the genes indicate percent conservation of that particular amino acid position  
935 in a multiple sequence alignment (Methods). Active site residues are in red. All active site residues are  
936 highly conserved with percent identities of greater than 99%. One active site cysteine and the active site  
937 aspartate residue are the two most conserved residues in DabA with 99.9% identity each.

938

939

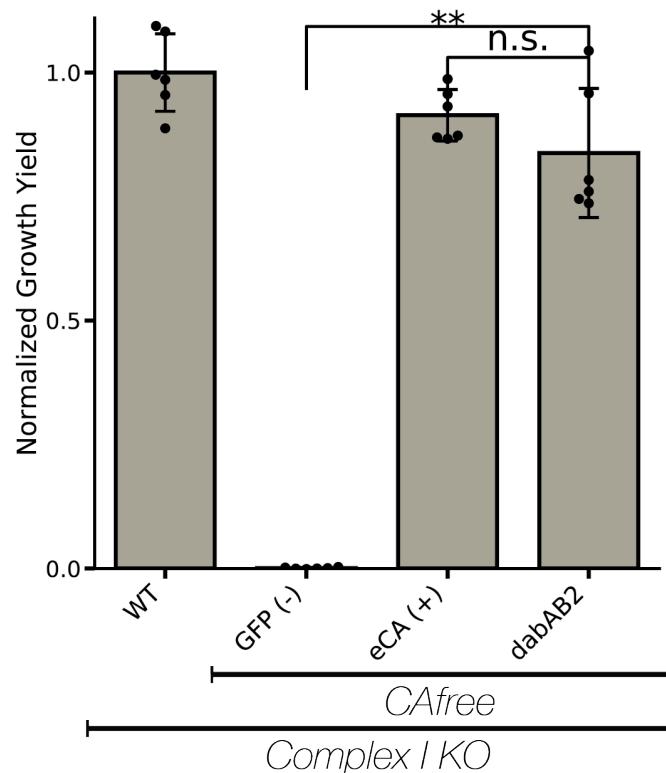
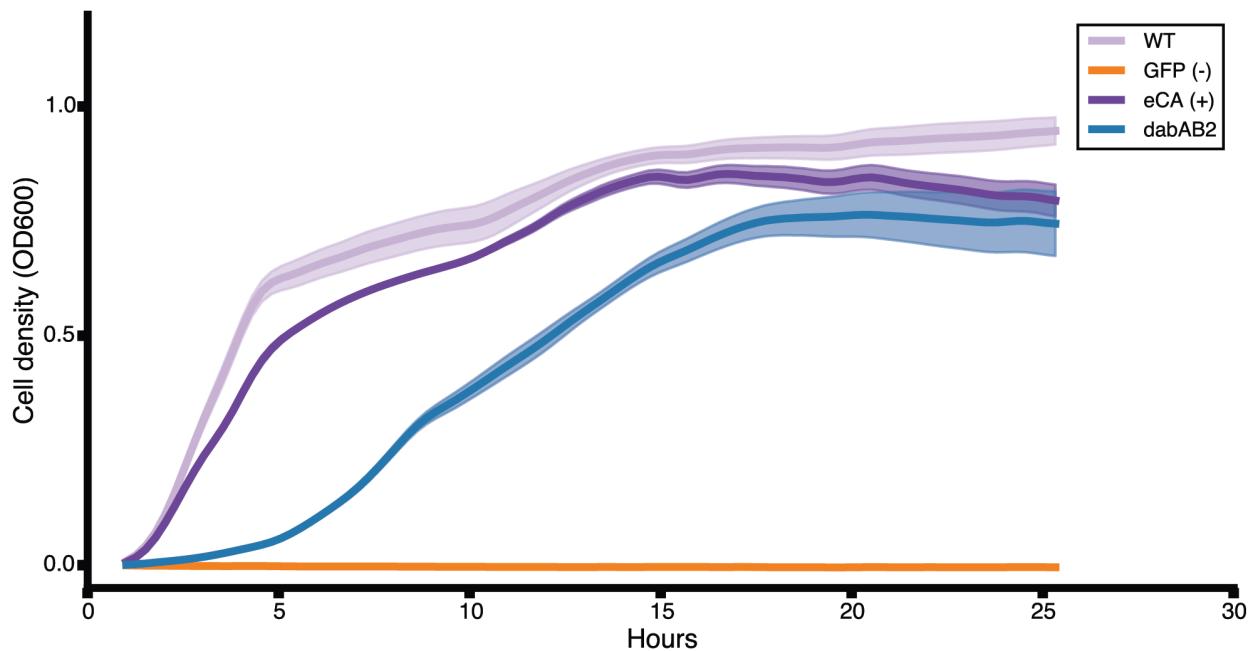
940



941

942 **Figure S5. Expression of DabAB2 rescues growth of CAfree *E. coli* in ambient CO<sub>2</sub>. A.** These  
 943 growth curves were used to generate the growth yield values in Figure 3B. Mean OD<sub>600</sub> is plotted +/-  
 944 standard error for four replicate cultures. Wild-type *E. coli* (BW25113) and CAfree strains expressing  
 945 either dabAB2 or human carbonic anhydrase II (hCA) grow in ambient CO<sub>2</sub> while CAfree expressing GFP,  
 946 dabB2 alone, or dabA2 alone fail to grow. **B.** These growth curves were used to generate the growth yield  
 947 values in Figure 4B. Mean OD<sub>600</sub> is plotted +/- standard error of four replicate cultures. Wild type cells  
 948 and CAfree expressing either DabAB2 or human carbonic anhydrase II (hCA) grow robustly. CAfree cells  
 949 expressing putative active site mutants of DabAB2 (C351, D353, H524, or C539) grow as poorly as the  
 950 negative control – CAfree expressing superfolder GFP in the same plasmid backbone.

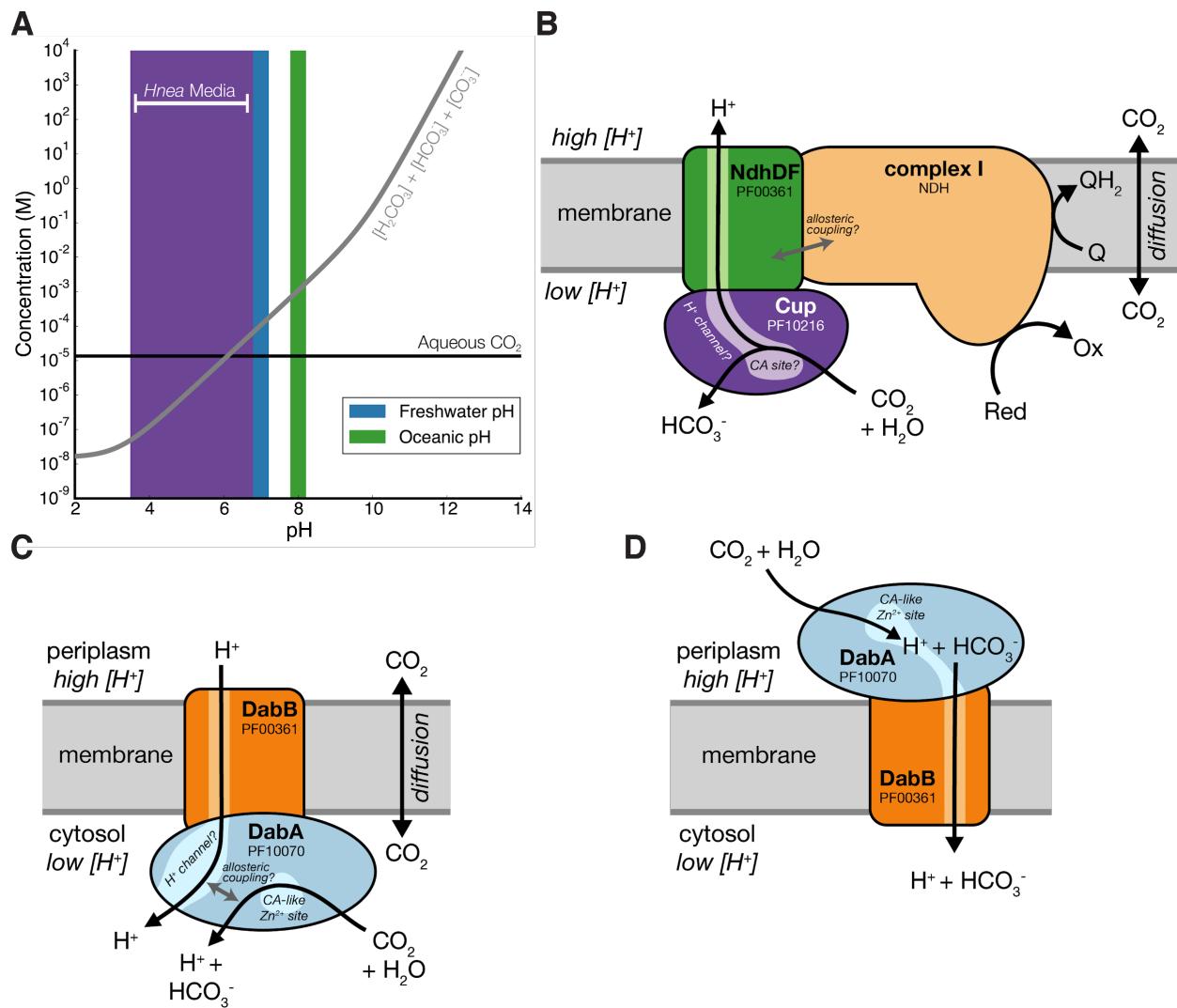
951

**A****B**

952

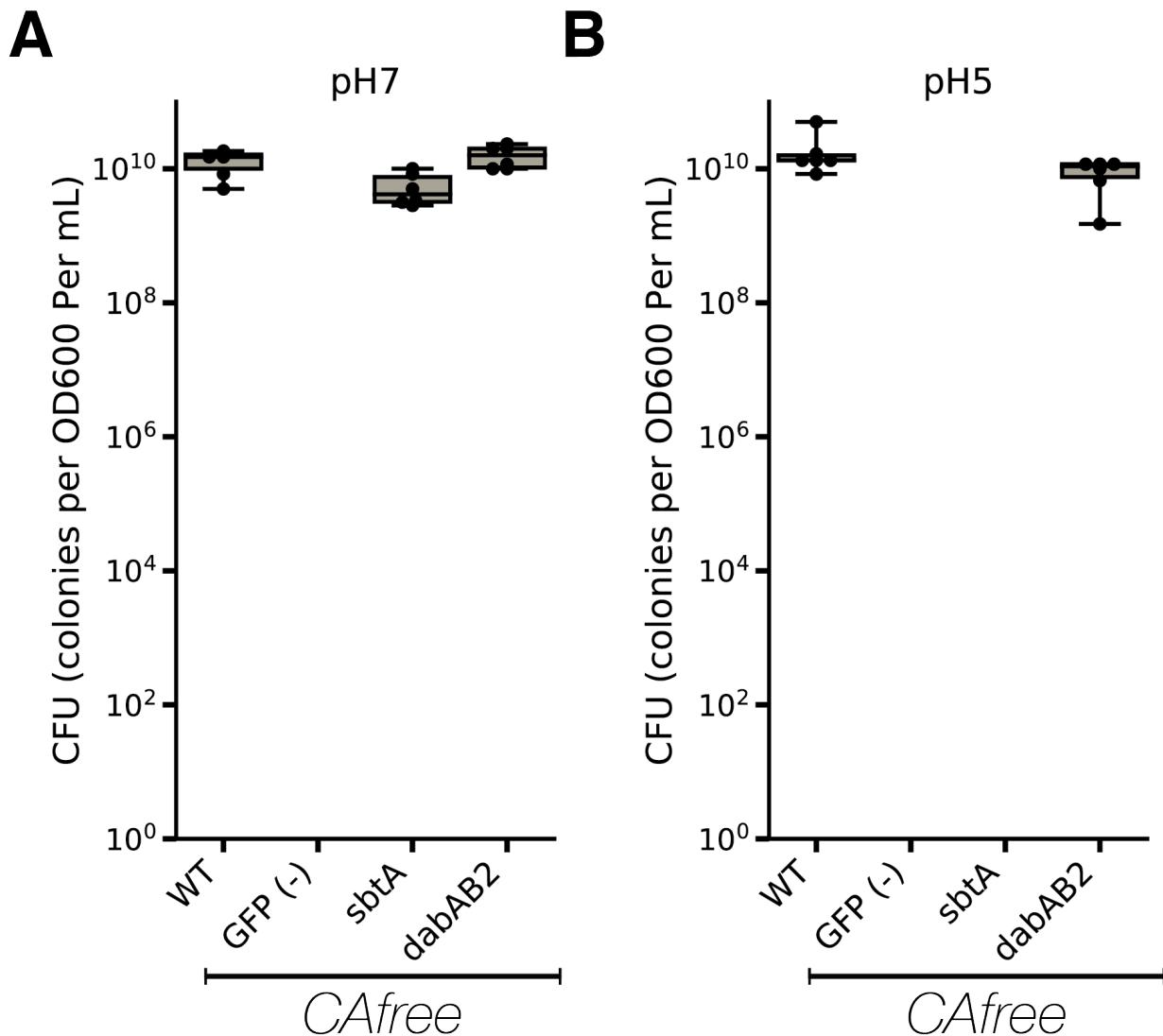
953 **Figure S6. DAB2 function is not dependant on complex 1.** A. DAB2 is still able to rescue growth of  
954 CAfree cells in the absence of Complex I ( $\Delta(nuoA-nuoN)$ ). dabAB2 rescues better than GFP ( $t=15.7$ ,  
955  $p=2.37 \times 10^{-8}$ ). Error bars represent standard deviation of six replicate cultures. "n.s." denotes means do  
956 not differ significantly, "\*" denotes that means differ with bonferroni corrected  $p < 0.05$  by a two-tailed t-  
957 test, and "\*\*\*" denotes  $p < 5 \times 10^{-4}$ . B. These growth curves were used to generate the growth yield values

958 in Figure S6A. Mean OD<sub>600</sub> is plotted +/- standard error of six replicate cultures. All strains are Complex I  
 959 knockout strains. DAB2 is still able to rescue growth of CAfree cells in the absence of Complex I.  
 960



961  
 962 **Figure S7. Comparison of models of vectorial CA activity for DABs and the Cyanobacterial CUP**  
 963 **systems.** A. Equilibrium concentrations of dissolved inorganic carbon as a function of pH. In this plot we  
 964 assume the growth medium is in Henry's law equilibrium with present-day atmosphere (400 PPM  $CO_2$ ) at  
 965 25 °C giving a soluble  $CO_2$  concentration of roughly 15  $\mu M$ . The equilibrium concentrations of hydrated  $C_i$   
 966 species ( $H_2CO_3$ ,  $HCO_3^-$ ,  $CO_3^{2-}$ ) is determined by the pH. As such, the organisms will "see" a  $C_i$  species in  
 967 very different ratios depending on the environmental pH. In a oceanic pH near 8,  $HCO_3^-$  dominates the  $C_i$   
 968 pool.  $HCO_3^-$  is also the dominant constituent of the  $C_i$  pool in freshwater, but less so (by a factor of ~10  
 969 since freshwater and oceanic environments differ by about 1 pH unit). In acid conditions (pH < 6.1)  $CO_2$   
 970 will be the dominant constituent of the  $C_i$  pool. The pH of our *Hnea* culture media ranges from 6.8 (when  
 971 freshly made) to ~3.5 when cells reach stationary phase (*Hnea* make  $H_2SO_4$  as a product of their sulfur  
 972 oxidizing metabolism). As such we expect that *Hnea* regularly experiences environments wherein it is

973 advantageous to pump CO<sub>2</sub> and not HCO<sub>3</sub><sup>-</sup>. **B.** CupA/B proteins are CA-like subunits of a class of  
974 cyanobacterial Ci uptake systems. Cup-type systems are believed to couple electron transfer to vectorial  
975 CA activity and, potentially, outward-directed proton pumping. This model is based on the observation  
976 that Cup systems displace the two distal H<sup>+</sup>-pumping subunits of the cyanobacterial complex I and  
977 replace them with related subunits that bind CupA/B (illustrated in green as NdhD/F). **C.** As our data are  
978 consistent with DAB2 functioning as a standalone complex (i.e. DabAB do not appear to bind or require  
979 the *E. coli* complex I), we propose a different model for DAB function where energy for unidirectional  
980 hydration of CO<sub>2</sub> is drawn from the movement of cations along their electrochemical gradient (right panel  
981 above). **D.** An alternative model for DAB activity is that DabA is localized to the periplasm and DabB is  
982 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  
983 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  
984 is not preferred because no secretion signals were observed in the DabA sequence. Moreover, the  
985 *Acidimicrobium ferrooxidans* genome contains an apparent DabA:DabB fusion protein. The predicted  
986 architecture the fusion would place DabA in the cytoplasm.

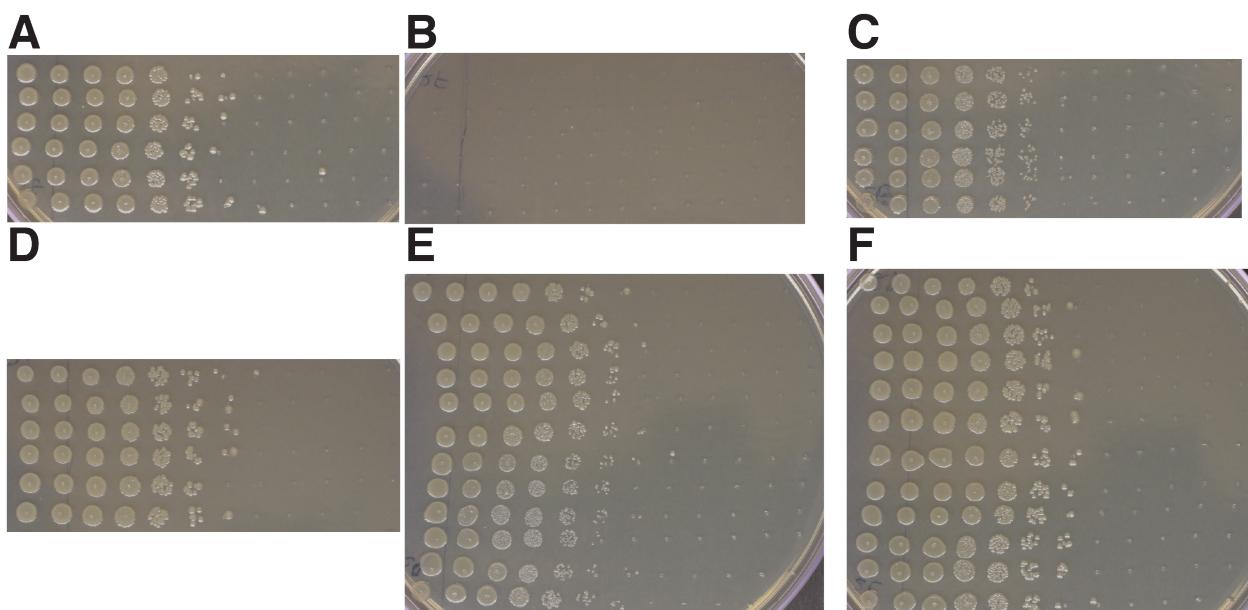


987  
988  
989  
990  
991  
992

**Figure S8. pH independence of *dabAB2* rescue of CAfree** Colony forming units per OD600 per ml were measured on LB plates with induction in air at both pH 7 (A.) and 5 (B.). *dabAB2* rescued growth at both pH7 and pH 5, *sbtA* only rescued growth at pH 7. Whiskers represent the range of the data, the box represents the interquartile range, and the middle line represents the median. Data is from 6 replicate platings of all conditions.



994 **Figure S9. Fully annotated approximate maximum likelihood phylogenetic trees of DabA.** **A.** A  
995 phylogenetic tree emphasizing the clades containing high-confidence DabA homologs. DabA homologs  
996 are found in > 15 prokaryotic clades, including some archaea. *Hnea* DabA1 and DabA2 represent two  
997 different groupings that are commonly found in proteobacteria. The tcr\_0854 gene of *H. crunogenus* is  
998 more closely related to DabA2 than DabA1. Inspecting the tree reveals several likely incidents of  
999 horizontal transfer, e.g. between proteobacteria and Firmicutes, Nitrospirae and Actinobacteria.  
1000 Moreover, the genomes of several known pathogens contain a high-confidence DabA homolog, including  
1001 *B. anthracis*, *L. pneumophila*, *V. cholerae*. **B.** Association of various Rubisco isoforms with DabA  
1002 homologs. Many organisms that have DabA also have a Rubisco. However, there are numerous  
1003 examples of DabA homologs that are found in genomes with no Rubisco (denoted by leaves with no  
1004 colored marking), suggesting that this uptake system might play a role in heterotrophic metabolism. DabA  
1005 is most-frequently associated with Form I Rubiscos (red and purple leaves in panel B), which is sensible  
1006 because all known bacterial CCMs involve a Form I Rubisco exclusively. Some DabA-bearing genomes  
1007 have only a Form II Rubisco (blue) and the Euryarchaeota genomes have that DabA have a Form III  
1008 Rubisco (green) or none at all. For both panels, scale bars indicate one substitution per site.  
1009



1010  
1011 **Figure S10. Plates used for determining CFU counts for Figure 5B.** **A.** Wt positive control. **B.** CAfree  
1012 sfGFP negative control does not rescue. **C.** CAfree hCA positive control rescues growth. **D.** CAfree DAB2  
1013 rescues growth. **E.** baDAB from *Bacillus anthracis* rescues growth of CAfree. **F.** vcDAB from *Vibrio*  
1014 *cholera* rescues growth of CAfree. Panels **A-D** represent 6 technical replicates of the plating. Panels **E**  
1015 and **F** represent 6 technical replicates each of 2 biological replicates. In all panels, the first spot  
1016 represents 3 ul of an OD 0.2 culture grown at 10% CO<sub>2</sub> each subsequent spot is 3 ul of a 1:10 dilution of  
1017 the previous spot.  
1018