

1 **DABs: a new class of inorganic carbon pumps found throughout prokaryotic**
2 **phyla**

3 John J. Desmarais¹, Avi I. Flamholz¹, Cecilia Blikstad¹, Eli J. Dugan¹, Thomas G. Laughlin¹,
4 Luke M. Oltrogge¹, Allen W. Chen³, Kelly Wetmore², Spencer Diamond³, Joy Y. Wang⁴, David F.
5 Savage^{1*}

6

7 ¹Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720

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

10 ³Department of Earth and Planetary Science, University of California, Berkeley, CA 94720

11 ⁴Department of Chemistry, University of California, Berkeley, CA 94720

12 * Corresponding author

13

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

15 **Abstract**

16 Bacterial autotrophs often rely on CO₂ concentrating mechanisms (CCMs) to assimilate carbon.
17 Although many CCM proteins have been identified, including a 200+ MDa protein organelle
18 called the carboxysome, a systematic screen of CCM components has not been carried out.
19 Here, we performed a genome-wide barcoded transposon screen to identify essential and
20 CCM-related genes in the γ-proteobacterium *H. neapolitanus*. Our screen identified a two-
21 gene operon encoding a domain of unknown function (PFAM:PF10070) and a putative
22 cation transporter subunit (PFAM:PF00361) critical for CCM function. Physiological and
23 biochemical assays demonstrated these two proteins, which we name DabA and DabB for
24 “DABs accumulate bicarbonate,” assemble into a heterodimeric complex and function as an
25 energy-coupled inorganic carbon pump. This analysis also reveals that DabA contains a
26 putative zinc-binding site reminiscent of a β-carbonic anhydrase, suggesting a possible
27 mechanism of activity. We further show that DabAB operons are found in diverse bacteria
28 and archaea. Finally, we demonstrate that functional DabAB operons are present in the
29 human pathogens *V. cholera* and *B. anthracis*. Based on these results, we propose that
30 DABs constitute a new class of energized inorganic carbon pump and play a critical role in
31 inorganic carbon metabolism throughout prokaryotic phyla.

32 **Introduction**

33 Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (Rubisco) is the primary
34 carboxylase of the Calvin-Benson-Bassham (CBB) cycle and the major entry point of inorganic
35 carbon (C_i) into the biosphere. Rubisco activity is thus critical to agriculture and a major flux
36 removing anthropogenic CO_2 from the atmosphere. Despite its centrality and abundance,
37 Rubisco is not a fast enzyme¹⁻³. Nor is Rubisco very specific - all known Rubiscos can use
38 molecular oxygen (O_2) as a substrate in place of CO_2 ⁴. The resulting oxygenation reaction is
39 often described as “wasteful” as it fails to incorporate inorganic carbon and produces a product,
40 2-phosphoglycolate, that is not part of the CBB cycle and must be recycled through
41 metabolically-expensive photorespiratory pathways^{5,6}. Many studies support the hypothesis that
42 improvements to Rubisco could improve crop yields, but Rubisco has proven recalcitrant to
43 protein engineering. Indeed, it remains unclear whether or how Rubisco can be improved^{3,7,8}.

44 Organisms that depend on Rubisco for growth often employ supplemental physiological
45 mechanisms to improve its rate and specificity. These mechanisms are collectively termed CO_2
46 concentrating mechanisms (CCMs) because they serve to concentrate CO_2 at the active site of
47 Rubisco so that carboxylation proceeds at its maximum rate and oxygenation is competitively
48 inhibited^{6,9,10}. All cyanobacteria and many chemotrophic proteobacteria have a CCM^{11,12}. The
49 bacterial CCM has garnered particular interest among bioengineers because it is well-
50 understood, composed of only ~20 genes and operates inside single cells¹³. Detailed modeling
51 suggests that transplantation of the bacterial CCM into crops might improve yields^{14,15} and
52 efforts towards transplantation are already underway¹⁶⁻¹⁸.

53 Based on diverse experimental studies, a general model of bacterial CCM function has
54 emerged requiring two major components: active transport of C_i leading to the accumulation of
55 HCO_3^- in the cytosol and organization of RuBisCO with carbonic anhydrase (CA) in the lumen of
56 a 200+ MDa protein organelle known as the carboxysome^{9,19-22}. Energy-coupled C_i pumps
57 ensure that the cytosolic HCO_3^- concentration is high (> 10 mM) and, crucially, out-of-
58 equilibrium with CO_2 ^{19,20,22-25}. CA activity interconverts $CO_2 + H_2O$ with $HCO_3^- + H^+$, and thus,
59 the carboxysomal CA converts a high HCO_3^- concentration into a high carboxysomal CO_2
60 concentration, which promotes faster carboxylation by Rubisco and competitively inhibits
61 oxygenation⁹. Genetic lesions to either component - C_i uptake systems or carboxysomes -
62 disrupt the CCM and mutants require elevated CO_2 for growth²⁶⁻²⁸. This high- CO_2 requiring
63 (HCR) mutant phenotype is commonly used to identify CCM components in screens^{20,26,29,30}.

64 Despite these early screens, a comprehensive list of bacterial CCM components
65 remains unknown, leaving the possibility that additional activities are required for CCM function.

66 Although well-assembled carboxysome structures can be heterologously produced in bacteria
67 and plants^{18,31,32}, the functionality of these carboxysomes in a heterologous CCM has not been
68 demonstrated. Moreover, genetic and bioinformatic studies show that several additional genes
69 are associated with carboxysome function^{33,34}. For example, it was recently demonstrated that
70 carboxysome-associated genes may function as Rubisco chaperones and assembly factors^{35,36}.
71 Moreover, many experimental^{20,37} and modeling studies^{9,21,22} make it clear that energy-coupled
72 C_i uptake systems are required for the CCM to function. Several different C_i pump families,
73 including transporters and facilitated uptakes systems are now known^{13,38}. However, since
74 model carbon-fixing bacteria often express multiple C_i uptake systems and these integral
75 membrane protein systems are difficult to assay biochemically, our mechanistic biochemical
76 understanding of C_i uptake is limited^{38–40}.

77 Here we use a genome-wide barcoded transposon mutagenesis screen (RB-TnSeq) to
78 interrogate the CCM of *Halothiobacillus neapolitanus* (henceforth *Hnea*). *Hnea* is a sulfur
79 oxidizing γ-proteobacterial chemoautotroph and a model system for studying α-
80 carboxysomes^{41,42}. In addition to producing the first catalog of essential genes for a bacterial
81 chemotroph, we leverage our pooled mutant library to comprehensively screen for knockouts
82 that produce an HCR phenotype. This screen identified all known CCM components and
83 confirmed that a two-gene operon containing a large, conserved, poorly-characterized protein
84 (PFAM:PF10070, hereafter DabA) and a member of a large family of cation transporters
85 (PFAM:PF00361, hereafter DabB) is required for CCM function. Recent proteomic analyses
86 and physiological experiments have shown that this operon is involved in C_i transport in
87 proteobacteria^{43,44}. For reasons outlined below, we term this locus the DAB operon for “DABs
88 Accumulate Bicarbonate.”

89 Here we show that the gene products of the DAB operon form a protein complex that is
90 capable of energetically-coupled C_i uptake. Both proteins are necessary for activity in our
91 experiments and treatment with an ionophore abrogates DAB-mediated C_i uptake. Structural
92 homology modeling suggests that DabA contains a domain distantly homologous to a type II β-
93 carbonic anhydrase. Indeed, we demonstrate that DabA binds zinc and depends on two
94 cysteines, one histidine and one aspartic acid residue for activity, in a manner reminiscent of β-
95 CAs^{45,46}. Taken together, these results indicate that DABs are a novel class of C_i pump,
96 energized by coupling to a cation gradient (e.g. H⁺ or Na⁺). Further, these results are consistent
97 with a model of activity dependant on unidirectional hydration of CO₂ to HCO₃⁻ in the cytosol via
98 a CA-like mechanism. Phylogenomic analysis demonstrates that DAB operons are widespread
99 throughout prokaryotes including carbon-fixing bacteria and archaea. Surprisingly, DAB operons

100 are also found in many heterotrophic bacteria. We demonstrate that functional operons are
101 present in the notable pathogens *V. cholera* and *B. anthracis*. We therefore propose that DABs
102 constitute a novel class of C_i uptake pump. Further, the biochemical tractability of these systems
103 facilitates mechanistic analysis and their widespread occurrence merits further investigation.

104

105 **Results**

106 *Transposon mutagenesis and gene essentiality*

107 Gene essentiality was defined by first constructing (Figure 1A) a randomly-barcoded genome-
108 wide pooled knockout library of *Hnea* using conjugation⁴⁷. The donor strain (*E. coli* APA 766)
109 contains a vector with a barcoded Tn5-derived transposon encoding a kanamycin resistance
110 marker. Conjugation was performed under 5% CO₂ so that CCM genes could be knocked out,
111 and the resulting *Hnea* conjugants were selected for growth in the presence of kanamycin at 5%
112 CO₂ to ensure transposon insertion.

113

114 The presence of a unique barcode in each transposon simplifies the use of the library for pooled
115 screens using the ‘barseq’ approach⁴⁷. Transposon insertion sites and associated barcodes
116 must be mapped to the *Hnea* genome in order to perform these screens. Transposon insertions
117 were mapped using standard TnSeq methods⁴⁷ and the library was found to contain ~10⁵
118 insertions, or one insertion for every ≈25 base pairs in the *Hnea* genome. Since the average
119 gene contains ≈35 insertions, genes with no insertions are very likely essential for growth⁴⁸.

120 Following this logic, we used a simple statistical model to identify 551 essential genes and 1787
121 nonessential genes out of 2408 genes in the *Hnea* genome (Methods, Figure 1A-B, File 2). The
122 remaining 70 genes were classified as “ambiguous” due either to their short length or because
123 replicate mapping experiments were discordant (Methods). Genes associated with known
124 essential functions including central carbon metabolism, ribosome production, and DNA
125 replication were categorized as essential (Figures 1C and S1). As the library was generated
126 under 5% CO₂, known CCM genes, including carboxysome components, were not essential for
127 growth (Figure 2).

128

129 *Comprehensive screen for Hnea CCM components*

130 Based on the current model of the bacterial CCM (Figure 2A) knockouts of CCM genes are
131 expected to require high CO₂ for growth^{20,29,30}. CCM gene knockouts should therefore have low
132 fitness in ambient CO₂ concentrations. As our pooled library contains ~70,000 barcodes that
133 map to exactly one position in the *Hnea* genome, we were able to use the barseq method to
134 quantify the fitness defects associated with single gene knockouts for all nonessential *Hnea*
135 genes (Figure 2B). In barseq, a preculture of the library is grown in permissive conditions (5%
136 CO₂) and then back-diluted into two conditions: a reference condition (5% CO₂ again) and a
137 condition of interest (i.e. ambient CO₂). Genomic DNA is extracted from the preculture (called t₀)

138 and both culture outgrowths. Barcodes are then PCR-amplified and sequenced. In this pooled
139 competition assay, the proportional change in barcode abundance is taken to reflect the fitness
140 effect of gene knockouts⁴⁷. A CCM gene knockout should have no fitness defect in 5% CO₂ but
141 a large defect in ambient CO₂. Since the library contains >20 knockouts with unique barcodes
142 per gene (on average), these screens contain multiple internal biological replicates testing the
143 effect of single gene knockouts.

144

145 As expected, knockouts to nearly all carboxysome-associated genes produced large fitness
146 defects in ambient CO₂ (Figures 2B-C). These genes include *cbbLS* - the large and small
147 subunits of the α-carboxysomal Rubisco; *csoS2* - an intrinsically disordered protein required for
148 α-carboxysome assembly⁴⁹; *csoSCA* - the carboxysomal carbonic anhydrase; *csoS4AB* - the
149 pentameric proteins thought to form vertices of the α-carboxysome; and *csoS1CAB* - the
150 hexamers that form the faces of the α-carboxysome shell^{12,42}. Knockouts of *csoS1D*, a shell
151 hexamer with a large central pore^{31,50}, confer a very weak HCR phenotype in this screen and so
152 *csoS1D* did not cross the threshold for being called HCR (Figures 2B-C). The *Hnea* genome
153 also contains a secondary, non-carboxysomal Form II Rubisco that is likely not involved in CCM
154 activity as its disruption confers no fitness defect in ambient CO₂. A number of genes that are
155 not structurally associated with the carboxysome also exhibited HCR phenotypes. These
156 include two LysR transcriptional regulators, a Crp/Fnr type transcriptional regulator, a protein
157 called acRAF that is involved in Rubisco assembly^{35,36}, and two paralogous loci encoding DAB
158 genes (hereafter DAB1 and DAB2, Figure 2B-F).

159

160 *dabA2* and *dabB2* are necessary and sufficient for energy-coupled C_i accumulation in *E. coli*
161 DAB1 is a cluster of 3 genes found in an operon directly downstream of the carboxysome
162 operon (Figure 2C). Though DAB1 is part of a larger 11-gene operon containing several genes
163 associated with Rubisco proteostasis, including acRAF^{35,36} and a cbbOQ-type Rubisco
164 activase⁵¹, we refer to DAB1 as an “operon” for simplicity. DAB2 is a true operon and is not
165 proximal to the carboxysome operon in the *Hnea* genome. These “operons” are unified in that
166 they both display HCR phenotypes and possess similar genes (Figures 2B-D).

167

168 Both operons contain a conserved helical protein of unknown function (PFAM:PF10070) that we
169 term DabA. DabA proteins have no predicted transmembrane helices or signal peptides and
170 thus appear to be large (DabA1: 118.5 kDa, DabA2: 91.7 kDa), soluble, cytoplasmic proteins
171 (Methods, Figure 3A). Both DAB operons also contain a member of the cation transporter family

172 (PFAM:PF00361) that includes H⁺-pumping subunits of respiratory complex I and Mrp Na⁺:H⁺
173 antiporters. This protein, which we call DabB, is smaller than DabA (DabB1: 62.2 kDa, DabB2:
174 59.3 kDa) and is predicted to have 12-13 transmembrane helices (Figure 3A). The complex I
175 subunits in PF00361 are H⁺-pumping proteins and not redox active, i.e. they do not posses iron-
176 sulfur clusters, flavin binding sites, or quinone binding sites. Moreover, DabB proteins form a
177 distinct clade in a phylogenetic tree of PF00361. This clade appears to be as distant from
178 complex I subunits as it is from the Mrp antiporters (Figure S3A). Therefore, homology between
179 DabB and canonical complex I subunits (e.g. NuoL) suggests that DabB is a cation transporter
180 but does not necessarily imply redox activity. Operons of this type were recently demonstrated
181 to be involved C_i transport in proteobacterial chemotrophs^{43,44}.

182
183 In order to facilitate testing for C_i transport abilities, we generated an *E. coli* strain, CAfree, that
184 contains no CA genes (Methods). It was previously shown that deletion of the constitutive CA,
185 *can*, gene produces an HCR phenotype in *E. coli*⁵² that is complemented by expression of
186 cyanobacterial bicarbonate transporters⁵³. However, this approach is limited by the potential of a
187 second, inducible CA, *cynT*, to act as a suppressor of the HCR phenotype. Deleting both of
188 these genes removes this avenue for suppression of the phenotype. Since DAB2 disruption is
189 associated with a larger fitness defect than DAB1 (Figure 2B), we used CAfree to test DAB2 for
190 C_i uptake activity. DAB2 expression enables growth of CAfree in ambient CO₂ while expression
191 of either gene alone is not sufficient (Figures 3B and S4). Uptake assays using radiolabeled ¹⁴C_i
192 demonstrates that DAB2 facilitates transport of C_i levels significantly above that of steady-state
193 (Figure 3C). Moreover, DAB2-associated C_i uptake is strongly inhibited by the ionophore CCCP
194 (white bars in Figure 3C), also indicating that DAB2 is energetically-coupled, either directly or
195 indirectly, to a cation gradient (e.g. H⁺ or Na⁺).

196
197 *DabA2 and DabB2 interact to form a complex*
198 In order to determine if the genetic interaction between *dabA2* and *dabB2* is due to a physical
199 interaction, we attempted to purify the two proteins as a complex. DabA2 was genetically fused
200 to a C-terminal Strep-tag, DabB2 was fused to a C-terminal GFP with 6xHis-tag, and the genes
201 were assayed for co-expression in *E. coli* (Methods). Tandem-affinity purification following
202 detergent solubilization in 1% β-dodecyl-maltoside revealed that DabA2 and DabB2 interact
203 physically to form a complex in *E. coli* (Figure 4A). The complex runs as a single major peak in
204 size exclusion chromatography and has a retention volume consistent with a heterodimer of
205 DabA2 and DabB2 (Figure 4B). We did not observe co-purification of *E. coli* complex I subunits

206 or any other proteins with the DabA-DabB complex (Figure 4A), suggesting that DAB2 operates
207 as an independent complex within the membrane. Relatedly, *DAB2* expression rescues CAfree
208 growth even when complex I is knocked out ($\Delta(nuoA\text{-}nuoN)$) (Figure S5), providing further
209 evidence that DAB function is independent of complex I.

210

211 *pH independence of dabAB rescue suggests that CO₂ is the true substrate*

212 Aqueous CO₂ spontaneously interconverts with the gas phase as well as hydrated C_i species
213 (H₂CO₃, HCO₃⁻, CO₃²⁻). The equilibrium of CO_{2(aq)} and CO_{2(gas)} is not affected by pH, but the
214 conversion from CO₂ to hydrated C_i is highly pH dependant. The result of this interaction is that
215 in low volume, well mixed solutions, the equilibrium concentration of HCO₃⁻ increases 100 fold
216 between pH 5 and 7 without an accompanying change in CO₂ concentration (Figure S6A)⁹.
217 SbtA, a known HCO₃⁻ transporter, rescues CAfree growth at pH 7 but not at pH 5, while DabAB2
218 rescues growth at both pHs (Figure S7). Since DabAB2 rescue is pH-independent in this range,
219 its substrate is likely CO₂ and H₂CO₃, HCO₃⁻, or CO₃²⁻.

220

221 *DabAB2 binds zinc at a predicted CA-like active site in dabA*

222 Structural homology modeling software predicts that the middle of DabA2 has sequence
223 elements related to a β-CA (Figure 3A). Specifically, Phyre2 predictions identify C539 and H524
224 as part of a potential Zn²⁺ binding site distantly homologous to a bacterial type II β-CA (10%
225 coverage of DabA, 90.8% confidence). I-TASSER predicts a Zn²⁺ binding site including the
226 same residues along with an additional cysteine (C351), and aspartic acid (D353). As shown in
227 Figure 4C, these residues could make up the active site of a type II β-CA⁵⁴⁻⁵⁶. We generated
228 individual alanine mutants for each of these putative active site residues (C351A, D353A,
229 H534A and C539A) and tested their ability to rescue CAfree. All mutants failed to rescue CAfree
230 in ambient CO₂ (Figure 4D). We proceeded to assay zinc binding of purified dabAB complex
231 using X-ray fluorescence spectroscopy and found that wild-type dabAB2 and three of the single
232 mutants (C351A, D353A, and H534A) bind zinc (Figure 4E). These single mutants retain three
233 of four zinc-coordinating residues⁴⁶, which could explain why the mutants appear to bind zinc.
234 Indeed, mutational studies of the human CA II show that single mutations to Zn²⁺-binding
235 residues reduce but do not abrogate zinc binding^{45,57}.

236

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

238 We tested whether detergent solubilized, purified DabAB2 displays carbonic anhydrase activity
239 (Figure 4F). CA activity was not detected. DabAB2 was assayed at high protein concentrations

240 (> 650-fold more protein than the positive control) and under CO₂ concentrations that are
241 typically saturating for CAs, but showed as much CA activity as the negative control (Figure 4F).
242 Absence of activity *in vitro* argues either that DabAB2 has extremely low CA activity or, more
243 likely, that DabAB2 must reside in a cell membrane holding a cation gradient to function as an
244 energetically-activated carbonic anhydrase.

245

246 *dabA is widespread in prokaryotes and functional variants are present in human pathogens*

247 Searching the Uniprot database with the DabA pfam (PF10070) and pruning truncated and
248 poorly aligned sequences yielded 878 DabA sequences. DabA sequences were found in a wide
249 variety of prokaryotes including bacteria and archaea (fig. 5A and S8). Represented clades
250 include not only γ -Proteobacteria but also α -Proteobacteria, \square -Proteobacteria, Euryarchaeota,
251 Firmicutes, Planctomycetes, and Bacteroides. Many DabA sequences were found in genomes
252 of heterotrophic organisms that cannot fix CO₂. Additionally the important heterotrophic human
253 pathogens *V. cholera*, *B. anthracis*, and *L. pneumophila* contain apparent DabA homologs. We
254 then wanted to know if these dabA genes were part of DAB operons, we defined a putative DAB
255 operon as a i gene that had an identifiable *dabB* gene present as a direct fusion to *dabA* or
256 within three genes and oriented in the same direction. Inspection of local gene neighborhoods of
257 sequences from the tree revealed that 843 (96%) of *dabA* sequences from the tree are in
258 putative DAB operons.

259

260 We tested whether DAB homologs from heterotrophic pathogens are functional C_i pumps. *V.*
261 *cholera* E7946 El Tor Ogawa and *B. anthracis* Sterne both contain putative DAB operons.
262 These operons were cloned and expressed in CAfree *E. coli*. Expression of either of these DAB
263 operons rescues growth of CAfree in ambient CO₂ (Fig. 5B and S9). Thus, the DAB operons
264 from even non-autotrophic human pathogens contain functional variants.

265

266 **Discussion**

267 Bacterial CCMs exist as two convergently evolved families the α - and β -lineages. Both lineages
268 require two major components: i. energy-coupled uptake of inorganic carbon to concentrate
269 HCO₃⁻ in the cytosol and ii. carboxysome structures that co-localize Rubisco with CA enzymes
270 that convert concentrated HCO₃⁻ into a high concentration of the Rubisco substrate CO₂⁹. While
271 the carboxysome components are well-documented for both α - and β -families, C_i uptake
272 systems of the proteobacterial CCM have only been identified very recently^{43,44}. Moreover,

273 though several laboratories have spent decades studying the bacterial CCM, it remains unclear
274 whether our current “parts list” for α - and β -CCMs is complete.

275

276 Here we undertook an effort to complete the genetic “parts list” of the α -family CCM of the
277 proteobacterial chemotroph *H. neapolitanus*. We generated a genome-wide knockout library
278 containing \approx 35 individual knockouts for every gene in the *Hnea* genome and compiled the first
279 list of essential genes for a chemotroph (Figure 1). Because we generated the library at
280 elevated CO₂ (5%, Figure 1A) we were able to knockout all known CCM components, including
281 genes known to form the α -carboxysome (Figure 2C). We subsequently used this library to
282 screen for genes associated with CCM activity by screening for knockouts with fitness defects
283 specific to ambient CO₂ growth conditions (Figure 2B). As expected, this screen highlighted
284 most known carboxysome components. It also identified several genes whose relationship to
285 the CCM is not fully understood (Figures 2B-F), including several transcriptional regulators, a
286 putative Rubisco chaperone³⁵ and two small operons (DAB1 and DAB2) that are involved in
287 CCM-associated C_i uptake in chemotrophic proteobacteria^{43,44}.

288

289 We showed that the DAB2 operon encodes a two-component protein complex that has C_i
290 uptake activity when heterologously expressed in *E. coli* (Figure 3B-C). This complex is likely a
291 heterodimer as suggested by size-exclusion chromatography (Figure 4B). As C_i uptake is
292 strongly inhibited by the ionophore CCCP (Figure 3C), we suspect that DAB2 activity is
293 energetically-coupled to a cation gradient (Figure 5A). Since DabAB2 shows pH-independent
294 rescue of CAfree *E. coli* (pH 5-7) CO₂ is likely the transported substrate (Figure 4C). This idea is
295 further supported by the fact that DabA has limited homology to a type II β -CA and binds a zinc
296 (Figures 3-4), which could indicate the presence of a CA active site that hydrates transported
297 CO₂. Mutations to the putative zinc-binding residues (C351A, D353A, H534A, and C539A)
298 ablate function in-vivo (Figure 4D). For these reasons, we propose a model of DAB activity
299 wherein CO₂ is passively taken into the cell and then vectorially (unidirectionally) hydrated to
300 HCO₃⁻ by energy-coupled CA activity of DabA.

301

302 Model carbonic anhydrases are not directly coupled to any energy source (e.g., ATP) and so
303 they only accelerate the equilibration of CO₂ and HCO₃⁻^{45,56}. Energy coupled CA activity could
304 produce unidirectional CO₂ hydration, allowing the DAB system to actively accumulate HCO₃⁻ in
305 the cytosol and power the CCM (as diagrammed in Figure 2A). Given the similarity of DabB to

306 other H⁺-pumping proteins, we propose the DAB is coupled to the H⁺ gradient, but our results
307 are equally consistent with other cation gradients, e.g. Na⁺. This mechanism would require tight
308 coupling of cation flow to CO₂ hydration by the CA-like DabA protein, consistent with our
309 observation that purified DabAB2 displays no measurable CA activity. Interestingly, type II β-
310 CAs are the only CAs known to display allosteric regulation^{46,55}. This allosteric control is thought
311 to be mediated by Zn²⁺ capping and uncapping by the active site aspartic acid (D353 in DabA2).
312 A similar mechanism might couple cation movement through DabB to the active site of DabA.

313

314 Cyanobacteria also possess vectorial CA systems called CUPs, which may provide clues to the
315 DAB mechanism^{37,38,58–60}. Indeed, both DAB and CUP systems contain subunits in the Mrp
316 protein family (DabB and NdhD/F are in PF00361) that also contains the H⁺-pumping subunits
317 of complex I. This commonality might suggest a shared mechanism. CO₂ hydration by CupA/B
318 is thought to be coupled to energetically-favorable electron flow because CupA/B proteins
319 appear to associate with the cyanobacterial complex I^{39,61} (Figure S8B). However, the Mrp
320 protein family (PF00361) is very diverse and contains many cation transporters that do not
321 associate with complex I or any other redox-coupled membrane complex^{43,62,63}. Moreover, DabB
322 and NdhD/F sequences are only distantly related to complex I subunits (Figure S3A), the two
323 DAB subunits do not appear to bind *E. coli* complex I (Figure 4A) and DAB2 rescue of CAfree
324 growth does not require complex I (Figure S5). We therefore propose that DAB activity is not
325 coupled to electron flow through complex I but, rather, to a cation gradient across the
326 membrane as described above (Figure 6).

327

328 DabAB2 functions very robustly, as demonstrated by complementation of CAfree *E. coli* (Figure
329 3B) and ¹⁴C uptake measurements (Figure 3C). Indeed, we observed that DabAB2 functions
330 substantially better in *E. coli* than SbtA, a primary C_i transporter of freshwater cyanobacteria^{53,59}
331 (Figure 3C). As *E. coli* and *Hnea* are both proteobacteria, this observation could be due to
332 greater “compatibility” of proteobacterial proteins with *E. coli* expression. It may also be the
333 case, though, that the α-CCM of proteobacteria is more “portable” than the β-CCM of freshwater
334 cyanobacteria. Indeed, α-CCM genes are typically found in a single gene cluster in
335 chemoautotrophs throughout α-, β- and γ-proteobacteria and the α-CCM was clearly horizontally
336 transferred at least once from proteobacteria to marine cyanobacteria⁵⁹. Phylogenomic analysis
337 of DabA homologs reveals they are widespread in prokaryotes and were likely horizontally
338 transferred multiple times (Figure 5B). Since DAB complexes are prevalent among prokaryotes
339 and have superlative activity, DAB-family transporters are an attractive target for protein

340 engineering and heterologous expression in plants and industrial microbes, where elevated
341 intracellular C_i could be technologically useful⁶⁴.

342

343 Finally, we were surprised to find evidence of active DABs outside of known carbon-fixing
344 bacteria (Figure 5). High-confidence DabA homologs are found in many lineages, including
345 notable heterotrophic pathogens including *V. cholerae*, *B. anthracis*, and *L. pneumophila* (Figure
346 5A). Moreover, we showed that DAB operons from *V. cholerae* and *B. anthracis* are active in *E.*
347 *coli*, which leads us to wonder: what do heterotrophic pathogens need C_i uptake systems for?
348 Carbonic anhydrase activity is essential for growth of the heterotrophs *E. coli* and *S. cerevisiae*
349 in ambient CO₂^{52,65}. In the heterotrophic context, CA activity is thought to supply bicarbonate for
350 the biotin-dependent carboxylases of central metabolism, for which HCO₃⁻ is the true
351 substrate^{52,65}. Additionally, bicarbonate levels have been linked to virulence in both *V. cholera*
352 and *B. anthracis*⁶⁶⁻⁶⁸. Perhaps DAB-family C_i uptake systems play roles in the growth or
353 virulence of these important pathogens? We hope that future research will delineate the role of
354 energetically-activated C_i uptake in heterotrophic and pathogenic organisms.

355 **Materials and Methods**

356 *Important strains and reagents*

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

358

359 *Bacterial strains and growth conditions*

360 *E. coli* strain APA766 was used as the conjugation donor to transfer the Tn5 transposon to
361 *Halothiobacillus neapolitanus* C2 (*Hnea*) via conjugation⁴⁷. The *E. coli* double CA deletion strain
362 “CAfree” (BW25113 *Δcan ΔcynT*) was generated by curing the KEIO collection *cynT* knockout
363 (BW25113 *ΔcynT*, KEIO strain JW0330) of kanamycin resistance via pCP20-mediated FLP
364 recombination and subsequent P1 transduction (and curing) of kanamycin resistance from the
365 *can* knockout strain EDCM636 (MG1655 *Δcan*, Yale Coli Genomic Stock Center,^{52,69}). Complex
366 I knock out strains (*Δ(nuoA-nuoN)*) were generated in both the BW25113 and CAfree
367 backgrounds. These strains were generated by lambda red mediated recombination of a Kan^R
368 resistance cassette flanked by frt sites into the *nuo* locus such that the entire operon was
369 removed. This was followed by heat curing of the pSIM5 plasmid⁷⁰ expressing the lambda red
370 recombinase at 42 °C. Lysogeny broth (LB) and LB agar were used as *E. coli* growth media
371 unless otherwise specified. *E. coli* strains were grown at 37 °C in the presence of 0.1 mg/ml
372 carbenicillin, 0.06 mg/ml kanamycin, or 0.025 mg/ml chloramphenicol as appropriate. *Hnea* was
373 grown in DSMZ-68 media at 30 °C and in the presence of 0.03 mg/ml kanamycin when
374 appropriate.

375

376 *Transposon mutagenesis and RB-TnSeq library production*

377 A barcoded library of *Hnea* transposon mutants was generated by adapting the methods of
378 Wetmore *et al.*⁴⁷. Conjugations were performed as follows. *Hnea* and APA766 were cultured
379 and harvested by centrifugation. Both cultures were washed once in 10 mL antibiotic-free
380 growth media per conjugation reaction and resuspended in 100 µl. 5 OD600 units of *Hnea* were
381 mixed with 20 OD600 units of APA766 on a 0.45 µM Millipore MCE membrane filter and
382 cultured overnight at 30 °C in 5% CO₂ on an antibiotic-free LB agar plate containing 0.06 mg/ml
383 diaminopimelic acid. Cells were scraped from the filter into 2 mL DSMZ-68 and collected in a 2
384 mL microcentrifuge tube. Recovered cells were pelleted by centrifugation at 16000 × g for 1
385 minute, washed in 2 mL DSMZ-68, pelleted again at 9000 × g for 1 minute, and resuspended in
386 2 ml DSMZ-68 before 200 µl was plated onto 10 separate DSMZ-68 kanamycin plates (per
387 conjugation). Plates were incubated at 30 °C under 5% CO₂ until colonies formed (~ 7 days).
388 Colonies were counted and scraped into 55 mL DSMZ-68. Two 1.4 OD600 unit samples were

389 taken and used to prepare genomic DNA (Qiagen DNeasy blood and tissue kit). Transposon
390 insertions were amplified from gDNA following protocols in Wetmore *et al.*⁴⁷. Transposons were
391 mapped after Illumina sequencing using software developed in Wetmore *et al.*⁴⁷ 1.6 OD600 unit
392 aliquots were then flash frozen in 50% glycerol for subsequent BarSeq experiments.

393

394 *Essential gene assignment*

395 Following the logic of Wetmore *et al.* and Rubin *et al.*^{47,48}, we categorized genes as essential if
396 we observed significantly fewer transposon insertions than would be expected by chance. If
397 insertion occurred uniformly at random, the number of insertions per gene would be expected to
398 follow a binomial distribution. The probability of observing at most k insertions into a gene of
399 length n is therefore expressed as:

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

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

407

408 *Gene fitness experiments*

409 Fitness experiments were performed according to a modification of the protocol in Wetmore *et*
410 *al.*⁴⁷. A library aliquot was thawed and used to inoculate three 33 mL cultures. Cultures were
411 grown to OD600 ~0.08 in 5% CO₂. At this point, 20 mL were removed and harvested by
412 centrifugation as two t_0 (input) samples. Cultures were back-diluted 1:64 into 128 mL and
413 incubated for 6.5-7.5 doublings under 5% CO₂ or ambient conditions. 50 mL of culture was
414 harvested by centrifugation. gDNA was prepared and barcodes were amplified for fitness
415 determination via Illumina sequencing as described in Wetmore *et al.*⁴⁷.

416

417 *CAfree rescue experiments*

418 Electrocompetent CAfree cells were prepared using standard protocols⁷¹ and transformed with
419 pFE plasmids expressing genes of interest by electroporation. CAfree pre-cultures were grown
420 overnight in 10% CO₂ and diluted into 96 well plates (3 µl cells in 250 µl media). Growth curves

421 were measured by culturing cells in a Tecan M1000 microplate reader under ambient conditions
422 with continuous shaking, and measuring OD600 every 15 minutes. When samples are marked
423 "induced," 200 nM anhydrotetracycline (aTc) was added to the media. Growth yields are
424 calculated as the maximum OD600 achieved after 24 hours of growth and normalized to the
425 yield of a wild type control. CFU experiments were performed by back diluting cultures to
426 OD600 0.2 before performing 10X serial dilutions. 3 μ l of the OD600 0.2 sample and each of the
427 serial dilutions were then spotted on plates with 200 nM aTc and grown overnight in
428 atmosphere. The spot with the highest dilution that yielded more than one colony was counted
429 and a minimum of six replicates were averaged for each strain.

430

431 *Silicone oil centrifugation measurement of inorganic carbon uptake*

432 The silicone oil filtration method was modified from Dobrinski *et al.*⁷² and used to measure
433 uptake of radiolabeled inorganic carbon. Assay tubes were generated using 0.6 ml
434 microcentrifuge tubes containing 20 μ l of dense kill solution (66.7% v/v 1 M glycine pH 10,
435 33.3% v/v triton X-100) covered by 260 μ l of silicone oil (4 parts AR20:3.5 parts AR200).
436 Electrocompetent CAfree cells were prepared using standard protocols and transformed with
437 pFA-based plasmids containing genes of interest by electroporation. CAfree cultures were
438 grown overnight in 10% CO₂, back diluted to an OD600 of 0.1 and allowed to grow to mid-log
439 phase in 10% CO₂ in the presence of 200 nM aTc for induction. Cells were then harvested by
440 centrifugation, washed once in PBS (pH 7.55) and resuspended to OD600 0.6 in PBS + 0.4%
441 glucose. ¹⁴C-labeled sodium bicarbonate (PerkinElmer) was added to a final concentration of
442 4.1 nM and an activity of 0.23 μ Ci. Cells were incubated with ¹⁴C for 4 minutes before
443 centrifugation at 17,000 x g for 4 minutes to separate cells from buffer. Pellets were clipped into
444 scintillation vials containing 5 ml Ultima Gold scintillation fluid and 300 μ l 3M NaOH using
445 microcentrifuge tube clippers or medium dog toenail clippers. Counts were measured on a
446 PerkinElmer scintillation counter. ¹⁴C counts are normalized to 1 OD600 unit of cells added.
447 During inhibition assays, cells were incubated in PBS pH 7.55 with 0.4% glucose + 0.4% DMSO
448 and the inhibitor (100 μ M CCCP) for 10 minutes before assay.

449

450 *Generation of DabA phylogenetic tree*

451 We searched the Uniprot reference proteome database using the Pfam Hidden Markov Model
452 PF10070.9 with a cutoff e-value of 10⁻⁴. Our search recovered 941 candidate DabA proteins.
453 These sequences were aligned using MAFFT and manually pruned to remove fragments and
454 poorly aligning sequences. The remaining 878 candidate DabA sequences were re-aligned

455 MAFFT and an approximate maximum likelihood phylogenetic tree was constructed using
456 FastTree. Taxonomy was assigned to nodes in the tree based on NCBI taxonomy information
457 for the genomes harboring each sequence. Genomic neighborhoods for each gene in the tree
458 were determined using the EFIGNT online server⁷³ and genomes with a *dabB* gene within 3
459 genes of *dabA* and oriented in the same direction were considered to have full DAB operons.
460 *dabAB* fusions were found by visual inspection of genomic neighborhoods from those genomes
461 that did not have separate *dabB* genes located close to *dabA*.

462

463 *Generation of DabB phylogenetic tree*

464 DabB homologs were collected manually by searching MicrobesOnline for close homologs of
465 four PF00361 members in the *Hnea* genome (*dabB1*, *dabB2*, *Hneap_1953*, *Hneap_1130*) and
466 other characterized PF00361 members including *Synechococcus elongatus ndhF1*,
467 *Synechococcus elongatus ndhF3*, and *Synechococcus elongatus ndhF4*. Genes were clustered
468 to 95% similarity and genes with divergent operon structure were removed manually using
469 MicrobesOnline treeview⁷⁴. *nuoL* from *Escherichia coli*, *nqo12* from *Thermus thermophilus*, and
470 *ndhF1/3/4* from *Thermosynechococcus elongatus* BP-1 were added as markers. ClustalOmega
471 was used to construct a multiple sequence alignment and the resulting nearest-neighbor tree
472 was visualized using the Interactive Tree of Life^{75,76}.

473

474 *Protein annotation and structural homology modeling*

475 Secondary structural annotations for DabA and DabB were generated using XtalPred⁷⁷.
476 Structural Homology modeling of DabA was performed using Phyre2 and I-TASSER web
477 servers with default parameters^{78,79}. A list of close DabB homologs was assembled by searching
478 MicrobesOnline for PF00361 members with similar operon structure. A ClustalOmega alignment
479 was used to calculate residue-level conservation of DabB proteins while the MAFFT alignment
480 generated during the creation of the DabA tree was used to calculate residue level conservation
481 of DabA proteins (Figure S3B).

482

483 *Purification of DAB2*

484 Chemically competent BL21-AI *E. coli* were transformed with a pET14b-based vector containing
485 the *dabAB* genes. 1 liter of 2xYT media was inoculated with 20 ml of an overnight culture of
486 BL21-AI *E. coli* in LB+CARB and allowed to grow to mid log at 37 °C. When midlog was
487 reached, cells were induced with 20 ml of 50 mg/ml arabinose and transitioned to 20 °C for
488 overnight growth. Cultures were pelleted and resuspended in 10 ml TBS (50 mM Tris, 150 mM

489 NaCl, pH 7.5) supplemented with 1.2 mM phenylmethylsulfonyl fluoride, 0.075 mg/ml lysozyme
490 and 0.8 ug/ml DNase I per liter of starting culture and then incubated at room temperature on a
491 rocker for 20 minutes. Cells were lysed with four passes through a homogenizer (Avestin).
492 Lysate was clarified at 15,000 x g for 30 minutes. Membranes were pelleted at 140,000 x g for
493 90 minutes. Membrane pellets were resuspended overnight in 25 ml TBS supplemented with 1
494 mM phenylmethylsulfonyl fluoride and 1% β-dodecyl-maltoside (DDM, Anatrace) per liter of
495 culture following⁸⁰. Membranes were then re-pelleted at 140,000 - 200,000 x g for 60 minutes
496 and the supernatant was incubated with Ni-NTA beads (Thermo Fisher) for 90 min at 4 °C. The
497 resin was washed with “Ni buffer” (20 mM Tris + 300 mM NaCl + 0.03% DDM, pH 7.5)
498 supplemented with 30 mM imidazole and eluted with Ni buffer supplemented with 300 mM
499 imidazole. Eluent was then incubated with Strep-Tactin (Millipore) resin for 90 min at 4 °C.
500 Resin was washed with “strep buffer” (TBS + 0.03% DDM) and eluted with strep buffer
501 supplemented with 2.5 mM desthiobiotin. Eluent was concentrated using Vivaspin 6 100 kDa
502 spin concentrators and buffer exchanged into strep buffer by either spin concentration or using
503 Econo-Pac 10DG (Biorad) desalting columns. For analytical purposes, 300 µg of strep-purified
504 protein was injected onto a Superdex 200 Increase 3.2/300 size-exclusion column pre-
505 equilibrated in strep buffer and eluted isocratically in the same buffer.

506

507 *Carbonic anhydrase assays*

508 CA-catalyzed CO₂ hydration of purified DAB2 complex and human carbonic anhydrase (hCA)
509 was measured using the buffer/indicator assay of Khalifah⁸¹ on a KinTek AutoSF-120 stopped-
510 flow spectrophotometer at 25 °C. The buffer/indicator pair used was TAPS/m-cresol purple
511 measured at a wavelength of 578 nm using a pathlength of 0.5 cm. Final buffer concentration
512 after mixing was 50 mM TAPS, pH 8.0 with the ionic strength adjusted to 50 mM with Na₂SO₄,
513 and 50 µM of pH-indicator. Final protein concentration used was: 9.8 µM DAB2 (His-elution) and
514 0.015 µM hCA (positive control; Sigma Aldrich C6624). Saturated solution of CO₂ (32.9 mM)
515 was prepared by bubbling CO₂ gas into milli-Q water at 25 °C. The saturated solution was
516 injected into the stopped-flow using a gas-tight Hamilton syringe, and measurements were
517 performed in a final CO₂ concentration of 16.5 mM. Progression curves were measured in 7
518 replicates.

519

520 *X-ray fluorescence spectroscopy for metal analysis*

521 50-100 µg of protein in 20-200 µl of TBS + 0.03% DDM was precipitated by addition of 4
522 volumes of acetone and incubation at -20 °C for 1 hour. Samples were centrifuged at 21,130 x g

523 for 15 minutes in a benchtop centrifuge and the supernatant was removed. Pellets were stored
524 at 4 °C until analysis. Fluorescence analysis was performed by breaking up the pellet into 5 µl of
525 TBS + 0.03% DDM with a pipette tip. Small pieces of the pellet were looped with a nylon loop
526 and flash frozen at the beamline under a nitrogen stream. The sample was excited with a 14
527 keV X-ray beam and a fluorescence spectrum was collected. Sample emission spectra were
528 then used to identify metals. Metal analysis was performed on wild-type DAB2, Zn-binding
529 mutants C351A, D353A, and H534A, bovine CA (positive control; Sigma Aldrich C7025), and a
530 buffer blank was used as a negative control. A Rubisco crystal containing cobalt salts was also
531 used as a zinc free control. Displayed traces are averages of at least two experiments.
532 Experiments were performed at the Lawrence Berkeley National Laboratory Advanced Light
533 Source Beamline 8.3.1.

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 **Author contributions**

549 J.J.D., A.I.F., and D.F.S. conceived and designed this study, and wrote the final manuscript with
550 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.,
551 and D.F.S. conducted the research or interpreted results. All authors reviewed and approved
552 the final manuscript.

553

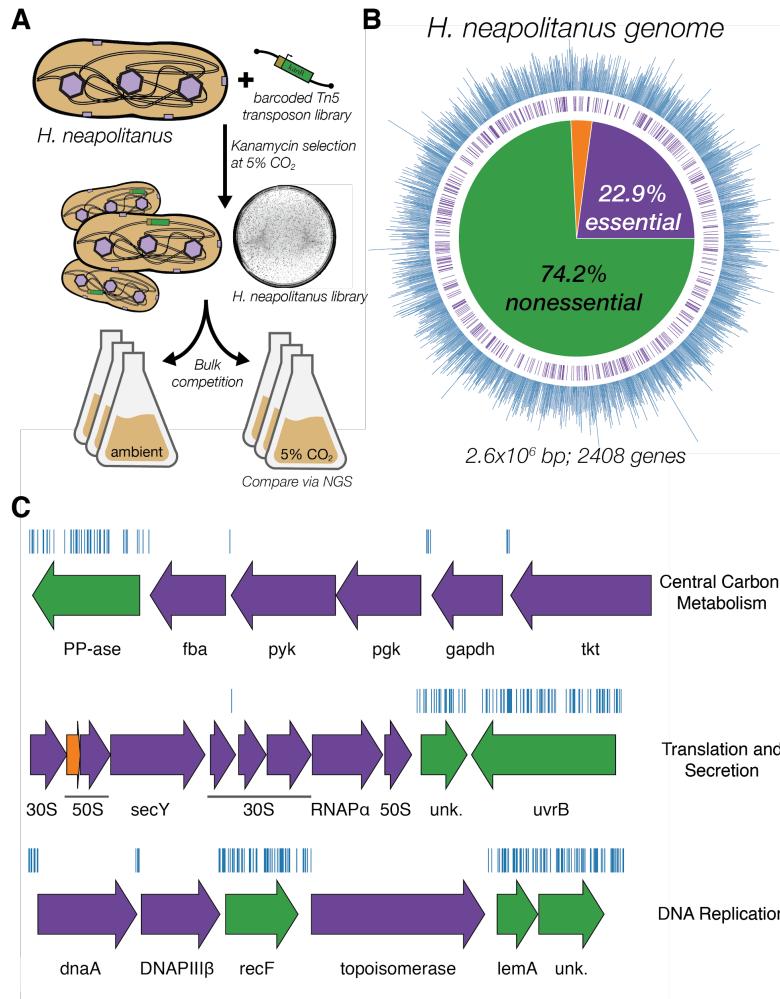
554 **Competing Interests**

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

559

560 **Materials & Correspondence**

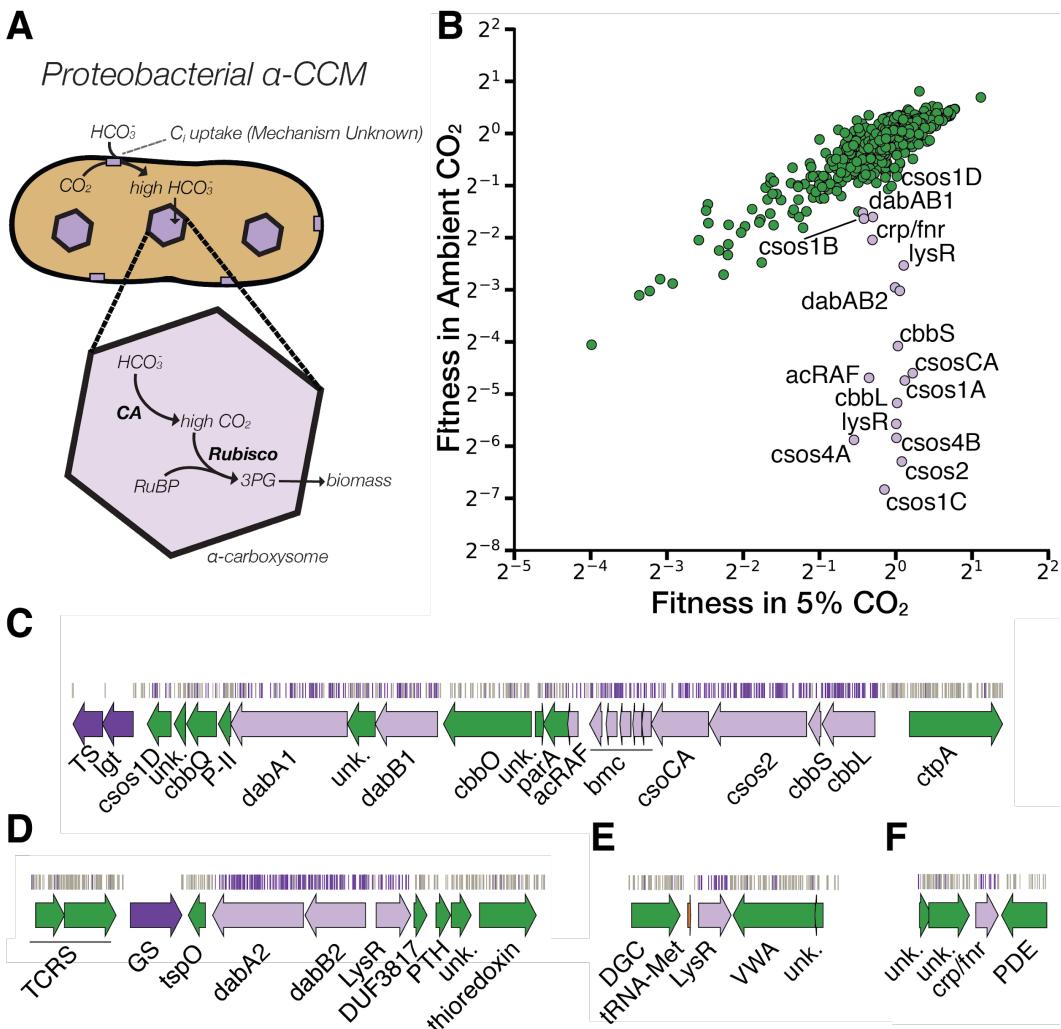
561 Correspondence should be addressed to: savage@berkeley.edu.



562

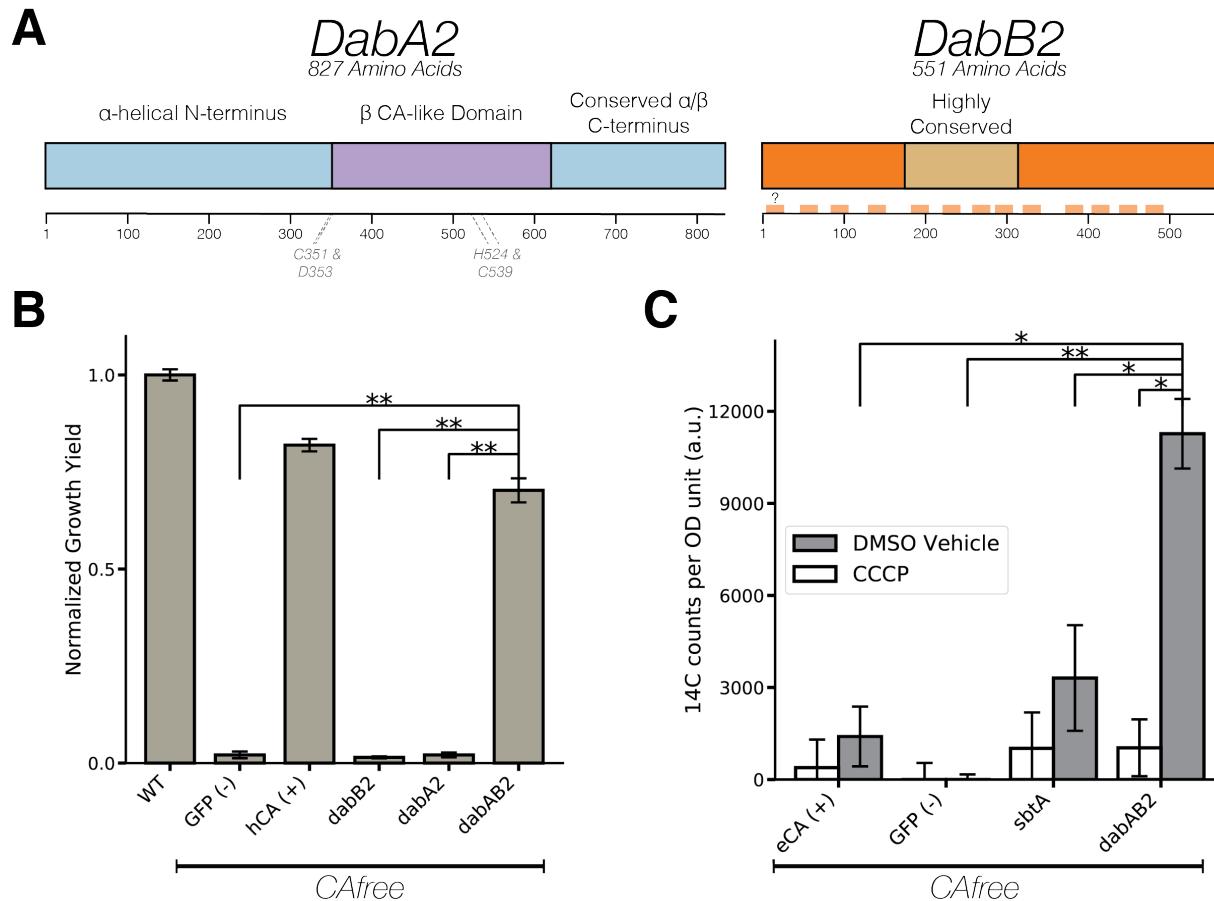
563 **Figure 1. Transposon mutagenesis reveals the essential gene set of a chemoautotrophic**
 564 **organism.** **A.** Schematic depicting the generation and screening of the RB-TnSeq library. Transposons
 565 were inserted into the *Hne*a genome by conjugation with an *E. coli* donor strain. The transposon contains
 566 a random 20 base pair barcode (yellow) and a kanamycin selection marker (green). Selection for colonies
 567 containing insertions was performed in the presence of kanamycin at 5% CO₂ and insertions were
 568 mapped by sequencing as described in the Methods. Subsequent screens were carried out as bulk
 569 competition assays and quantified by BarSeq. **B.** Insertions and essential genes are well-distributed
 570 throughout the *Hne*a genome. The outer track (blue) is a histogram of the number of barcodes that were
 571 mapped to a 1 kb window. The inner track annotates essential genes in purple. The pie chart shows the
 572 percentages of the genome called essential (purple), ambiguous (orange), and nonessential (green). **C.**
 573 Representative essential genes and nonessential genes in the *Hne*a genome. The blue track indicates
 574 the presence of an insertion. Genes in purple were called essential and genes in green are nonessential.
 575 Genes labeled “unk.” are hypothetical proteins. The first genomic locus contains 5 essential genes
 576 involved in glycolysis or the CBB cycle including pyruvate kinase (pyk) and transketolase (tkt). The 8
 577 essential genes in the second locus encoding 30S and 50S subunits of the ribosome, the secY secretory

578 channel, and an RNA polymerase subunit. Essential genes in the third example locus include
 579 topoisomerase and DNA polymerase III β . A full analysis with gene names is in Figure S1.

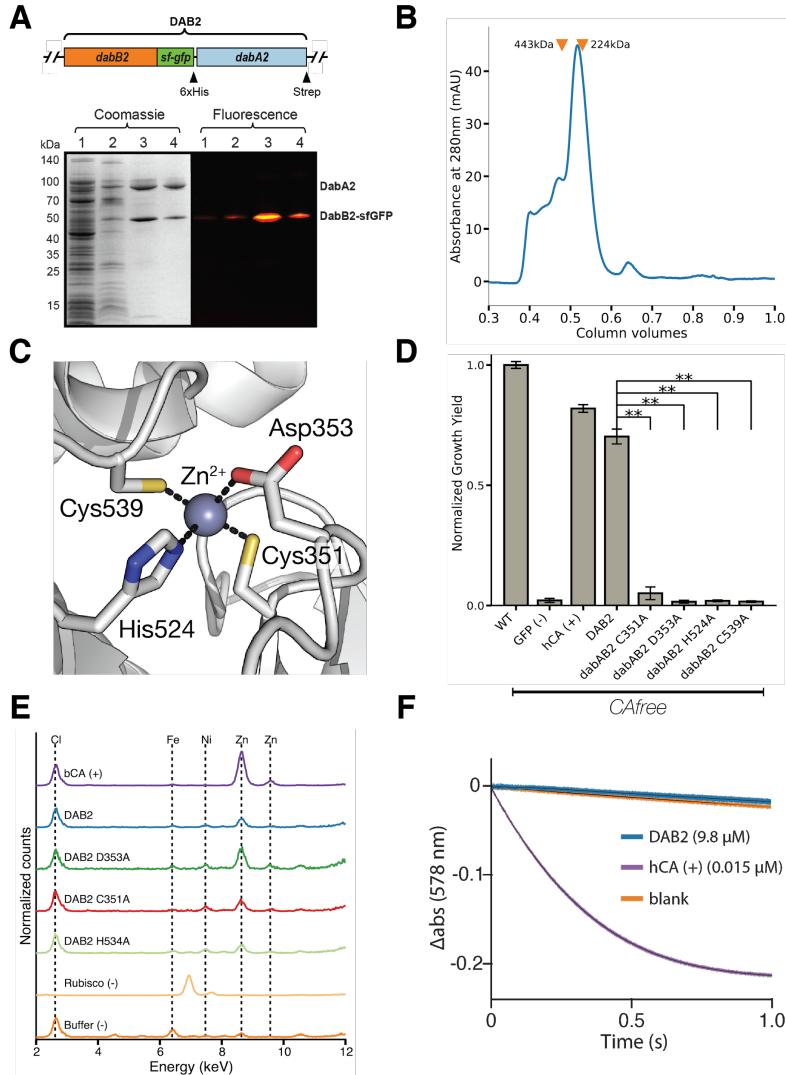


580
 581 **Figure 2. A systematic screen for high CO_2 -requiring mutants identifies genes putatively**
 582 **associated with the CCM. A.** Simplified model of the α -CCM of chemotrophic proteobacteria. Inorganic
 583 carbon is concentrated via an unknown mechanism, producing a high cytosolic HCO_3^- concentration. High
 584 cytosolic HCO_3^- is converted into high carboxysomal CO_2 by CA, which is localized only to the
 585 carboxysome. **B.** Fitness effects of gene knockouts in 5% CO_2 as compared to ambient CO_2 . Data is from
 586 one of two replicates of the BarSeq - the second replicate gives consistent results. When the effect of
 587 single transposon insertions into a gene are mutually consistent, those effects are averaged to produce
 588 the gene-level fitness value plotted⁴⁷. We define HCR mutants as those displaying a twofold fitness defect
 589 in ambient CO_2 relative to 5% CO_2 . HCR genes are colored light purple. Panels **C-F** show regions of the
 590 *Hnea* genome containing genes annotated as HCR in panel A. Essential genes are in dark purple, HCR
 591 genes are in light purple, and other genes are in green. The top tracks show the presence of an insertion
 592 in that location. Insertions are colored grey unless they display a twofold or greater fitness defect in

593 ambient CO₂, in which case they are colored light purple. **C.** The gene cluster containing the
594 carboxysome operon and a second CCM-associated operon. This second operon contains acRAF, a
595 Form IC associated cbbOQ-type Rubisco activase and *dabAB1*. **D.** The DAB2 operon and surrounding
596 genomic context. **E.** The genomic context of a lysR-type transcriptional regulator that shows an HCR
597 phenotype. **F** Genomic context of a crp/fnr-type transcriptional regulator that displays an HCR phenotype.
598 Genes labeled “unk.” are hypothetical proteins. Full gene names are given in Figure S2.



599
600 **Figure 3. The DABs catalyze active transport of C_i and are energized by a cation gradient. A.**
601 Diagrammatic representation of DabA2 and DabB2 based on bioinformatic annotation. DabA2 is an 827
602 amino acid protein with predicted homology to a type II β -CA enzyme. The four predicted active site
603 residues (C351, D353, H524, C539) are marked on the primary amino acid sequence. DabB2 is a 551
604 amino acid protein with 12-13 transmembrane helices. There is a highly conserved region in the middle of
605 its sequence and predicted transmembrane helices are marked in light orange. **B.** DAB2 was tested for
606 ability to rescue growth of CAfree *E. coli* in ambient CO_2 conditions. Expression of the full operon
607 (DabAB2) is required to rescue growth, as does the positive control, human carbonic anhydrase II (hCA)..
608 Error bars represent standard deviations of 4 replicate cultures. **C.** CAfree *E. coli* were tested for C_i
609 uptake using the silicone-oil centrifugation method. Expression of DabAB2 produced a large and
610 statistically significant increase in ^{14}C uptake as compared to all controls. Moreover, treatment with the
611 ionophore CCCP greatly reduces DabAB2-mediated ^{14}C uptake, suggesting that DabAB2 is coupled to a
612 cation gradient. *E. coli* CA (eCA) was used as a control for a non-vectorial CA. *Synechococcus elongatus*
613 PCC 7942 sbtA was used as a known C_i transporter. GFP was used as a vector control. Error bars
614 represent standard deviations of 3 technical replicates. In (B) and (C) “**” denotes that the means are
615 significantly different with Bonferroni corrected $P < 0.05$ according to a two-tailed T-test. “***” denotes $P <$
616 5×10^{-4} .

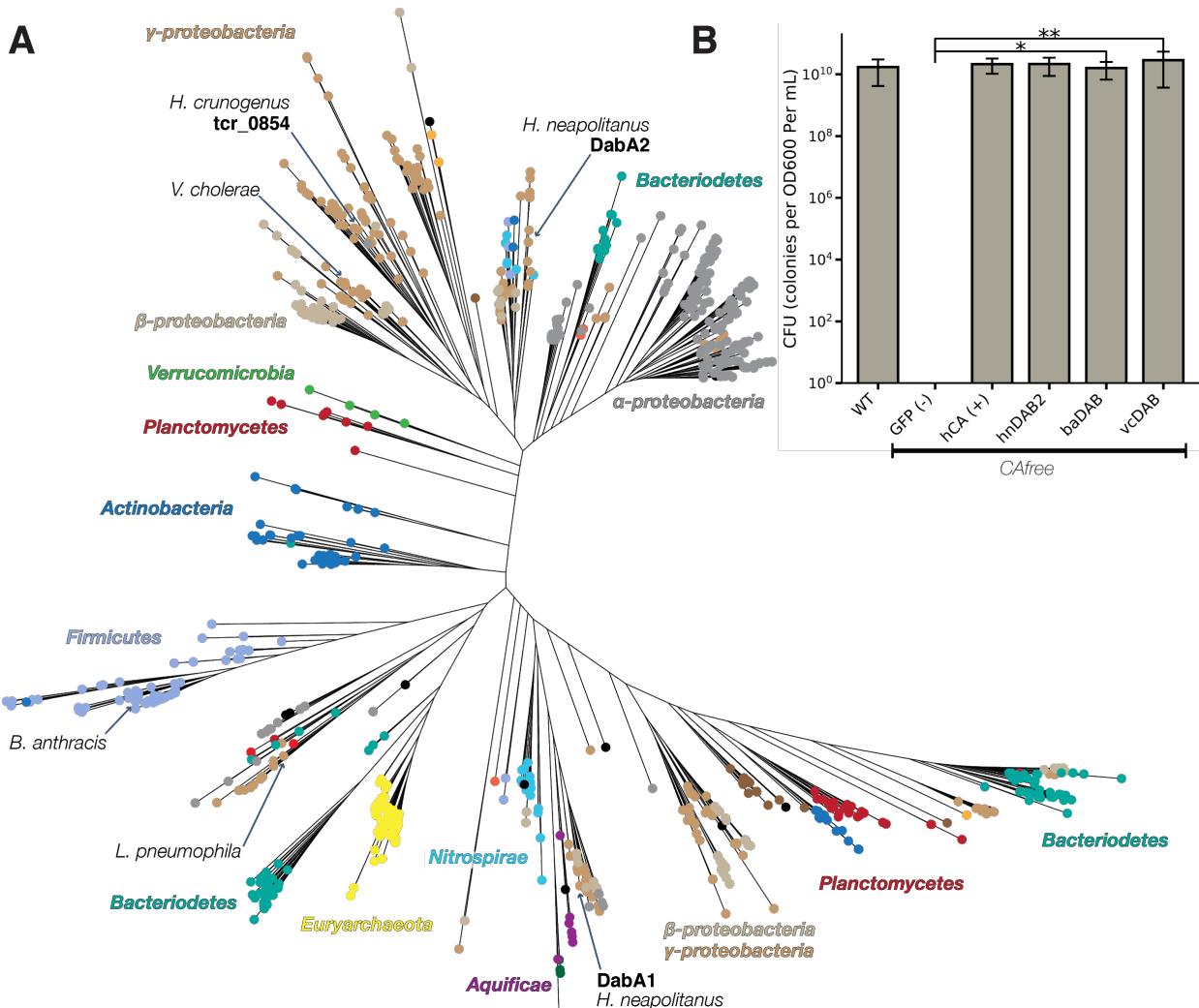


617

618 **Figure 4. DabA contains a β -CA-like active site but is not active outside of the membrane.** **A.**
 619 Purification of DabAB2 complex from *E. coli*. DabA2 was C-terminally tagged to a Strep-tag and DabB2
 620 was C-terminally tagged with sf-GFP and a 6xHis-tag. Purification was monitored using SDS-PAGE
 621 imaged with fluorescence (right view) before coomassie staining (left view). Lane 1: clarified lysate; 2:
 622 solubilized membranes; 3: Ni-NTA resin eluent; 4: strep-tactin resin eluent. DabA2 and DabB2 co-purify
 623 as a single complex without any obvious interactors. **B.** Size-exclusion chromatogram of His/Strep
 624 purified DabAB2 with retention volumes (orange arrows) and molecular weights (kDa) indicated for
 625 standard samples (apo ferritin, 443 kDa; β -amylase, 224 kDa). DabAB2 runs with a mass of \sim 270 kDa,
 626 which is likely an oligomer of DabA and DabB. **C.** Structural model of the DabA2 active site based on a β -
 627 CA of *E. coli* (PDB 1I6P). Typical β -CAs rely on two cysteine and one histidine residues to bind Zn^{2+} . The
 628 aspartic acid coordinates Zn^{2+} but is likely displaced during catalysis⁵⁵. **D.** Alanine mutants of the putative
 629 DabA2 active site residues (C351A, D353A, H524A, C539A) abrogate rescue of CAfree *E. coli*. Error bars
 630 give standard deviations of four replicates. **E.** X-ray fluorescence data indicate that DabAB2 binds zinc like
 631 all known β -CAs. Single mutations to the active site do not abrogate zinc binding. **F.** Purified DabAB2

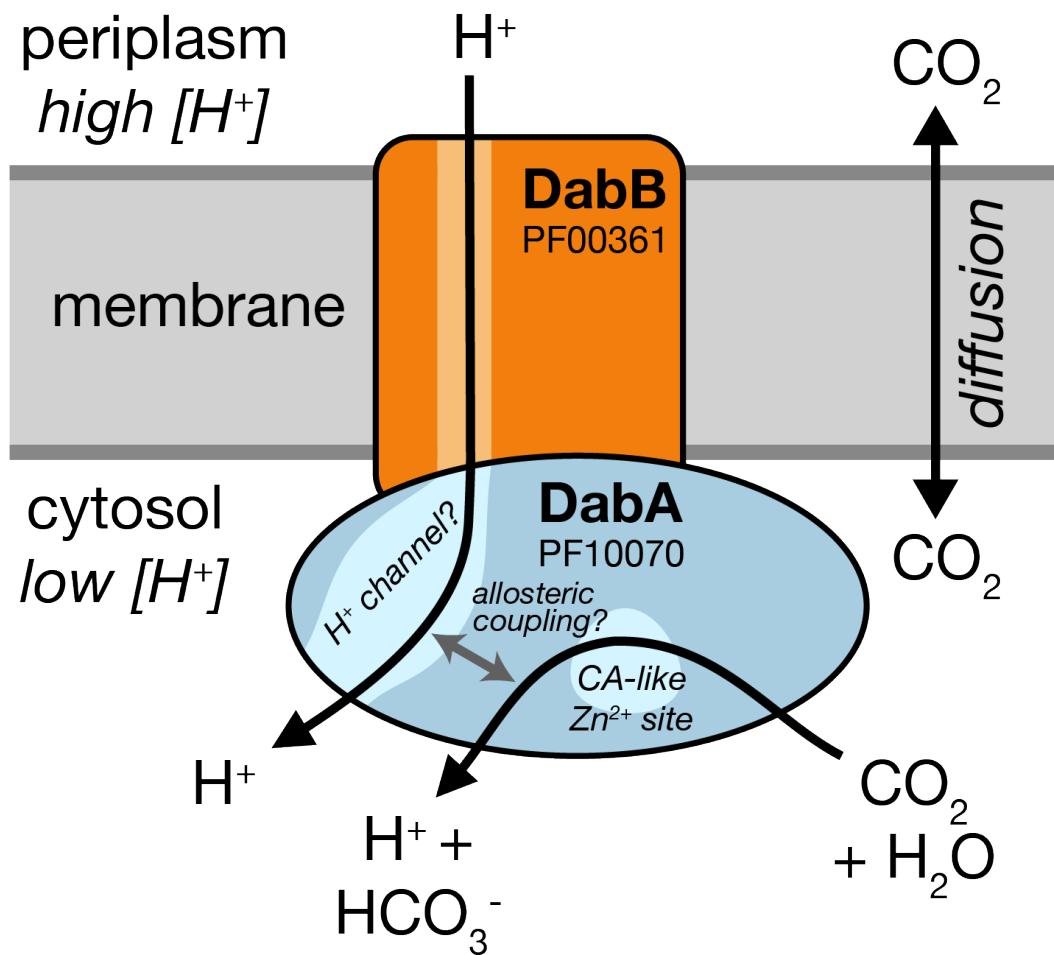
632 does not display any obvious CA activity despite being present in 650-fold excess over the positive
633 control (Human carbonic anhydrase II, hCA) in our assays. In all panels “*” denotes that means differ with
634 bonferroni corrected $P < 0.05$ by a two-tailed T-test, and “**” denotes $P < 5 \times 10^{-4}$.

635



638 **Figure 5. DAB operons are widespread in prokaryotes. A.** Approximate maximum likelihood
639 phylogenetic tree of DabA homologs associated with PF10070.9 (Methods). DabA homologs are found in
640 > 15 prokaryotic clades, including archaea. *Hne* DabA1 and DabA2 represent two different groupings
641 that are commonly found in proteobacteria. Inspecting the tree reveals several likely incidents of
642 horizontal transfer, e.g. between Proteobacteria and Firmicutes, Nitrospirae and Actinobacteria.
643 Moreover, the genomes of several known pathogens contain a high-confidence DabA homolog, including
644 *B. anthracis*, *V. cholerae*, and *L. pneumophila*. Detailed annotations are given in Figure S8. **B.** Functional
645 DABs are found in human pathogens. Colony forming units per OD600 per ml were measured on LB
646 plates with induction in air. DAB operons from *B. anthracis* (baDAB) and *V. cholerae* (vcDAB) rescued
647 growth of CAfree cells. The *Hne* operon DAB2 is abbreviated as hnDAB2. Error bars represent the
648 standard deviation of 6 replicate platings for WT, GFP (-), hCA (+), and hnDAB2. Error bars represent
649 standard deviations of 12 replicate platings for baDAB and vcDAB. “**” denotes that means differ with

650 bonferroni corrected $P < 0.05$ by a two-tailed T-test, and “**” denotes $P < 5 \times 10^{-4}$. CFU plates are shown
651 in Figure S9.
652



653
654 **Figure 6 A model of the unidirectional energy-coupled CA activity of DAB complexes.** We propose
655 that DabAB complexes couple CA activity of DabA to a cation gradient across the cell membrane,
656 producing unidirectional hydration of CO₂ to HCO₃⁻. The cation gradient could be H⁺ or Na⁺. Energy-
657 coupled CA activity is required for the DABs role as a C_i uptake system in the proteobacterial CCM, as
658 discussed in the text. Because it appears that DabAB2 is not active as a purified complex outside of the
659 membrane, it is assumed protein tightly couples the inflow of cations with CO₂ hydration so that there is
660 no “slippage.” Indeed, slippage - i.e., uncoupled CA activity - would be counterproductive for CCM
661 function^{9,19}. Notably, Zn²⁺ binding by the active site aspartic acid of type II β-CAs (D353 in DabA2, Figure
662 4A) is thought to allosterically regulate activity^{46,55}. This Asp-mediated activity switch could, therefore,
663 provide a means for allosteric coupling of a β-CA active site to distal ion transport.

664 **References**

- 665 1. Bar-Even, A. *et al.* The Moderately Efficient Enzyme: Evolutionary and Physicochemical
666 Trends Shaping Enzyme Parameters. *Biochemistry* (2011). doi:10.1021/bi2002289
- 667 2. Bathellier, C., Tcherkez, G., Lorimer, G. H. & Farquhar, G. D. Rubisco is not really so bad.
668 *Plant Cell Environ.* **41**, 705–716 (2018).
- 669 3. Flamholz, A. *et al.* Revisiting tradeoffs in Rubisco kinetic parameters. *bioRxiv* 470021
670 (2018). doi:10.1101/470021
- 671 4. Tcherkez, G. The mechanism of Rubisco-catalysed oxygenation. *Plant Cell Environ.* **39**,
672 983–997 (2016).
- 673 5. Bauwe, H., Hagemann, M. & Fernie, A. R. Photorespiration: players, partners and origin.
674 *Trends Plant Sci.* **15**, 330–336 (2010).
- 675 6. Buchanan, B. B., Gruissem, W. & Jones, R. L. *Biochemistry and Molecular Biology of*
676 *Plants*. (Wiley, 2015).
- 677 7. Tcherkez, G. G. B., Farquhar, G. D. & Andrews, T. J. Despite slow catalysis and confused
678 substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly
679 optimized. *Proc. Natl. Acad. Sci. U. S. A.* (2006). doi:10.1073/pnas.0600605103
- 680 8. Savir, Y., Noor, E., Milo, R. & Tlusty, T. Cross-species analysis traces adaptation of
681 Rubisco toward optimality in a low-dimensional landscape. *Proc. Natl. Acad. Sci. U. S. A.*
682 **107**, 3475–3480 (2010).
- 683 9. Mangan, N. M., Flamholz, A., Hood, R. D., Milo, R. & Savage, D. F. pH determines the
684 energetic efficiency of the cyanobacterial CO₂ concentrating mechanism. *Proc. Natl. Acad.*
685 *Sci. U. S. A.* **113**, E5354–62 (2016).
- 686 10. Raven, J. A., Beardall, J. & Sánchez-Baracaldo, P. The possible evolution, and future, of
687 CO₂-concentrating mechanisms. *J. Exp. Bot.* (2017). doi:10.1093/jxb/erx110
- 688 11. Badger, M. R. & Price, G. D. CO₂ concentrating mechanisms in cyanobacteria: molecular

- 689 components, their diversity and evolution. *J. Exp. Bot.* **54**, 609–622 (2003).
- 690 12. Cannon, G. C. *et al.* Microcompartments in prokaryotes: carboxysomes and related
691 polyhedra. *Appl. Environ. Microbiol.* **67**, 5351–5361 (2001).
- 692 13. Long, B. M., Rae, B. D., Rolland, V., Förster, B. & Price, G. D. Cyanobacterial CO₂-
693 concentrating mechanism components: function and prospects for plant metabolic
694 engineering. *Curr. Opin. Plant Biol.* **31**, 1–8 (2016).
- 695 14. Price, G. D., Badger, M. R. & von Caemmerer, S. The prospect of using cyanobacterial
696 bicarbonate transporters to improve leaf photosynthesis in C₃ crop plants. *Plant Physiol.*
697 **155**, 20–26 (2011).
- 698 15. McGrath, J. M. & Long, S. P. Can the cyanobacterial carbon-concentrating mechanism
699 increase photosynthesis in crop species? A theoretical analysis. *Plant Physiol.* **164**, 2247–
700 2261 (2014).
- 701 16. Lin, M. T., Occhialini, A., Andralojc, P. J., Parry, M. A. J. & Hanson, M. R. A faster Rubisco
702 with potential to increase photosynthesis in crops. *Nature* **513**, 547–550 (2014).
- 703 17. Occhialini, A., Lin, M. T., Andralojc, P. J., Hanson, M. R. & Parry, M. A. J. Transgenic
704 tobacco plants with improved cyanobacterial Rubisco expression but no extra assembly
705 factors grow at near wild-type rates if provided with elevated CO₂. *Plant J.* **85**, 148–160
706 (2016).
- 707 18. Long, B. M. *et al.* Carboxysome encapsulation of the CO₂-fixing enzyme Rubisco in
708 tobacco chloroplasts. *Nat. Commun.* **9**, 3570 (2018).
- 709 19. Price, G. D. & Badger, M. R. Expression of Human Carbonic Anhydrase in the
710 Cyanobacterium *Synechococcus* PCC7942 Creates a High CO₂-Requiring Phenotype
711 Evidence for a Central Role for Carboxysomes in the CO₂ Concentrating Mechanism. *Plant*
712 *Physiol.* **91**, 505–513 (1989).
- 713 20. Price, G. D. & Badger, M. R. Isolation and characterization of high CO₂-requiring-mutants
714 of the cyanobacterium *Synechococcus* PCC7942: two phenotypes that accumulate

- 715 inorganic carbon but are apparently unable to generate CO₂ within the carboxysome. *Plant*
716 *Physiol.* **91**, 514–525 (1989).
- 717 21. Reinhold, L., Kosloff, R. & Kaplan, A. A model for inorganic carbon fluxes and
718 photosynthesis in cyanobacterial carboxysomes. *Can. J. Bot.* **69**, 984–988 (1991).
- 719 22. Hopkinson, B. M., Young, J. N., Tansik, a. L. & Binder, B. J. The Minimal CO₂-
720 Concentrating Mechanism of Prochlorococcus spp. MED4 Is Effective and Efficient. *Plant*
721 *Physiol.* **166**, 2205–2217 (2014).
- 722 23. Whitehead, L., Long, B. M., Price, G. D. & Badger, M. R. Comparing the in Vivo Function of
723 α-Carboxysomes and β-Carboxysomes in Two Model Cyanobacteria. *Plant Physiol.* **165**,
724 398–411 (2014).
- 725 24. Holthuijzen, Y. A., van Dissel-Emiliani, F. F. M., Kuenen, J. G. & Konings, W. N. Energetic
726 aspects of CO₂ uptake in *Thiobacillus neapolitanus*. *Arch. Microbiol.* **147**, 285–290 (1987).
- 727 25. Kaplan, A., Badger, M. R. & Berry, J. A. Photosynthesis and the intracellular inorganic
728 carbon pool in the bluegreen alga *Anabaena variabilis*: Response to external CO₂
729 concentration. *Planta* **149**, 219–226 (1980).
- 730 26. Ogawa, T., Kaneda, T. & Omata, T. A Mutant of *Synechococcus PCC7942* Incapable of
731 Adapting to Low CO₂ Concentration. *Plant Physiol.* **84**, 711–715 (1987).
- 732 27. Dou, Z. *et al.* CO₂ fixation kinetics of *Halothiobacillus neapolitanus* mutant carboxysomes
733 lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO₂. *J. Biol.*
734 *Chem.* **283**, 10377–10384 (2008).
- 735 28. Cai, F. *et al.* The pentameric vertex proteins are necessary for the icosahedral
736 carboxysome shell to function as a CO₂ leakage barrier. *PLoS One* **4**, e7521 (2009).
- 737 29. Marcus, Y., Schwarz, R., Friedberg, D. & Kaplan, A. High CO(2) Requiring Mutant of
738 *Anacystis nidulans* R(2). *Plant Physiol.* **82**, 610–612 (1986).
- 739 30. Mackinder, L. C. M. *et al.* A repeat protein links Rubisco to form the eukaryotic carbon-
740 concentrating organelle. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 5958–5963 (2016).

- 741 31. Bonacci, W. *et al.* Modularity of a carbon-fixing protein organelle. *Proc. Natl. Acad. Sci. U.*
742 *S. A.* **109**, 478–483 (2012).
- 743 32. Fang, Y. *et al.* Engineering and Modulating Functional Cyanobacterial CO₂-Fixing
744 Organelles. *Front. Plant Sci.* **9**, 739 (2018).
- 745 33. Jorda, J., Lopez, D., Wheatley, N. M. & Yeates, T. O. Using comparative genomics to
746 uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci.* **22**, 179–
747 195 (2013).
- 748 34. Axen, S. D., Erbilgin, O. & Kerfeld, C. A. A taxonomy of bacterial microcompartment loci
749 constructed by a novel scoring method. *PLoS Comput. Biol.* **10**, e1003898 (2014).
- 750 35. Wheatley, N. M., Sundberg, C. D., Gidaniyan, S. D., Cascio, D. & Yeates, T. O. Structure
751 and Identification of a Pterin Dehydratase-like Protein as a Ribulose-bisphosphate
752 Carboxylase/Oxygenase (RuBisCO) Assembly Factor in the α-Carboxysome. *J. Biol.*
753 *Chem.* **289**, 7973–7981 (2014).
- 754 36. Aigner, H. *et al.* Plant RuBisCo assembly in *E. coli* with five chloroplast chaperones
755 including BSD2. *Science* **358**, 1272–1278 (2017).
- 756 37. Shibata, M., Ohkawa, H., Katoh, H., Shimoyama, M. & Ogawa, T. Two CO₂ uptake
757 systems in cyanobacteria: four systems for inorganic carbon acquisition in *Synechocystis*
758 sp. strain PCC6803. *Funct. Plant Biol.* **29**, 123–129 (2002).
- 759 38. Price, G. D. Inorganic carbon transporters of the cyanobacterial CO₂ concentrating
760 mechanism. *Photosynth. Res.* **109**, 47–57 (2011).
- 761 39. Battchikova, N., Eisenhut, M. & Aro, E. M. Cyanobacterial NDH-1 complexes: Novel
762 insights and remaining puzzles. *Biochimica et Biophysica Acta - Bioenergetics* **1807**, 935–
763 944 (2011).
- 764 40. Artier, J., Holland, S. C., Miller, N. T., Zhang, M. & Burnap, R. L. Synthetic DNA system for
765 structure-function studies of the high affinity CO₂ uptake NDH-13 protein complex in
766 cyanobacteria. *Biochim. Biophys. Acta Bioenerg.* (2018). doi:10.1016/j.bbabi.2018.06.015

- 767 41. Robertson, L. A. & Kuenen, J. G. The Genus *Thiobacillus*. in *The Prokaryotes: Volume 5: Proteobacteria: Alpha and Beta Subclasses* (eds. Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H. & Stackebrandt, E.) 812–827 (Springer New York, 2006).
- 768 42. Heinhorst, S., Cannon, G. C. & Shively, J. M. Carboxysomes and Carboxysome-like Inclusions. in *Complex Intracellular Structures in Prokaryotes* (ed. Shively, J. M.) 141–165 (Springer Berlin Heidelberg, 2006).
- 769 43. Mangiapia, M. *et al.* Proteomic and mutant analysis of the CO₂ concentrating mechanism of hydrothermal vent chemolithoautotroph *Thiomicrospira crunogena*. *J. Bacteriol.* (2017). doi:10.1128/JB.00871-16
- 770 44. Scott, K. M. *et al.* Genomes of ubiquitous marine and hypersaline *Hydrogenovibrio*, *Thiomicrorhabdus* and *Thiomicrospira* spp. encode a diversity of mechanisms to sustain chemolithoautotrophy in heterogeneous environments. *Environ. Microbiol.* **20**, 2686–2708 (2018).
- 771 45. Krishnamurthy, V. M. *et al.* Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein- ligand binding. *Chem. Rev.* **108**, 946–1051 (2008).
- 772 46. Rowlett, R. S. Structure and catalytic mechanism of the β-carbonic anhydrases. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1804**, 362–373 (2010).
- 773 47. Wetmore, K. M. *et al.* Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *MBio* **6**, e00306–15 (2015).
- 774 48. Rubin, B. E. *et al.* The essential gene set of a photosynthetic organism. *Proceedings of the National Academy of Sciences* 201519220 (2015).
- 775 49. Chaijaraspong, T. *et al.* Programmed Ribosomal Frameshifting Mediates Expression of the α-Carboxysome. *J. Mol. Biol.* **428**, 153–164 (2016).
- 776 50. Roberts, E. W., Cai, F., Kerfeld, C. A., Cannon, G. C. & Heinhorst, S. Isolation and characterization of the *Prochlorococcus* carboxysome reveal the presence of the novel shell protein CsoS1D. *J. Bacteriol.* **194**, 787–795 (2012).

- 793 51. Mueller-Cajar, O. The Diverse AAA+ Machines that Repair Inhibited Rubisco Active Sites.
- 794 *Front Mol Biosci* **4**, 31 (2017).
- 795 52. Merlin, C. & Masters, M. Why is carbonic anhydrase essential to Escherichia coli? *J.*
- 796 *Bacteriol.* **185**, (2003).
- 797 53. Du, J., Förster, B., Rourke, L., Howitt, S. M. & Price, G. D. Characterisation of
- 798 Cyanobacterial Bicarbonate Transporters in E. coli Shows that SbtA Homologs Are
- 799 Functional in This Heterologous Expression System. *PLoS One* **9**, e115905 (2014).
- 800 54. Cronk, J. D., Endrizzi, J. a., Cronk, M. R., O'neill, J. W. & Zhang, K. Y. Crystal structure of
- 801 E. coli beta-carbonic anhydrase, an enzyme with an unusual pH-dependent activity. *Protein*
- 802 *Sci.* **10**, 911–922 (2001).
- 803 55. Cronk, J. D. *et al.* Identification of a novel noncatalytic bicarbonate binding site in
- 804 eubacterial beta-carbonic anhydrase. *Biochemistry* **45**, 4351–4361 (2006).
- 805 56. Supuran, C. T. Structure and function of carbonic anhydrases. *Biochem. J* **473**, 2023–2032
- 806 (2016).
- 807 57. Ippolito, J. A. *et al.* Structure of His94-->Asp carbonic anhydrase II in a new crystalline form
- 808 reveals a partially occupied zinc binding site. *Protein Eng.* **8**, 975–980 (1995).
- 809 58. Shibata, M. *et al.* Distinct constitutive and low-CO₂-induced CO₂ uptake systems in
- 810 cyanobacteria: genes involved and their phylogenetic relationship with homologous genes
- 811 in other organisms. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 11789–11794 (2001).
- 812 59. Rae, B. D., Long, B. M., Badger, M. R. & Price, G. D. Functions, compositions, and
- 813 evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate
- 814 CO₂ fixation in cyanobacteria and some proteobacteria. *Microbiol. Mol. Biol. Rev.* **77**, 357–
- 815 379 (2013).
- 816 60. Maeda, S.-I., Badger, M. R. & Price, G. D. Novel gene products associated with NdhD3/D4-
- 817 containing NDH-1 complexes are involved in photosynthetic CO₂ hydration in the
- 818 cyanobacterium, *Synechococcus* sp. PCC7942. *Mol. Microbiol.* **43**, 425–435 (2002).

- 819 61. Birungi, M. *et al.* Possibilities of subunit localization with fluorescent protein tags and
820 electron microscopy exemplified by a cyanobacterial NDH-1 study. *Biochimica et*
821 *Biophysica Acta - Bioenergetics* **1797**, 1681–1686 (2010).
- 822 62. Krulwich, T. A., Hicks, D. B. & Ito, M. Cation/proton antiporter complements of bacteria: why
823 so large and diverse? *Mol. Microbiol.* **74**, 257–260 (2009).
- 824 63. Marreiros, B. C., Batista, A. P., Duarte, A. M. S. & Pereira, M. M. A missing link between
825 complex I and group 4 membrane-bound [NiFe] hydrogenases. *Biochim. Biophys. Acta*
826 **1827**, 198–209 (2013).
- 827 64. Antonovsky, N. *et al.* Sugar Synthesis from CO₂ in Escherichia coli. *Cell* **166**, 115–125
828 (2016).
- 829 65. Aguilera, J., Van Dijken, J. P., De Winde, J. H. & Pronk, J. T. Carbonic anhydrase
830 (Nce103p): an essential biosynthetic enzyme for growth of *Saccharomyces cerevisiae* at
831 atmospheric carbon dioxide pressure. *Biochem. J.* **391**, 311–316 (2005).
- 832 66. Sirard, J. C., Mock, M. & Fouet, A. The three *Bacillus anthracis* toxin genes are
833 coordinately regulated by bicarbonate and temperature. *J. Bacteriol.* **176**, 5188–5192
834 (1994).
- 835 67. Bongiorni, C. *et al.* Dual promoters control expression of the *Bacillus anthracis* virulence
836 factor AtxA. *J. Bacteriol.* **190**, 6483–6492 (2008).
- 837 68. Abuaita, B. H. & Withey, J. H. Bicarbonate Induces *Vibrio cholerae* virulence gene
838 expression by enhancing ToxT activity. *Infect. Immun.* **77**, 4111–4120 (2009).
- 839 69. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout
840 mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
- 841 70. Datta, S., Costantino, N. & Court, D. L. A set of recombineering plasmids for gram-negative
842 bacteria. *Gene* **379**, 109–115 (2006).
- 843 71. Oakes, B. L., Nadler, D. C. & Savage, D. F. Chapter Twenty-Three - Protein Engineering of
844 Cas9 for Enhanced Function. in *Methods in Enzymology* (ed. Sontheimer, J. A. D. A. E. J.)

- 845 **546**, 491–511 (Academic Press, 2014).
- 846 72. Dobrinski, K. P., Longo, D. L. & Scott, K. M. The Carbon-Concentrating Mechanism of the
847 Hydrothermal Vent Chemolithoautotroph *Thiomicrospira crunogena*. *J. Bacteriol.* **187**,
848 5761–5766 (2005).
- 849 73. Zallot, R., Oberg, N. O. & Gerlt, J. A. ‘Democratized’ genomic enzymology web tools for
850 functional assignment. *Curr. Opin. Chem. Biol.* **47**, 77–85 (2018).
- 851 74. Dehal, P. S. *et al.* MicrobesOnline: an integrated portal for comparative and functional
852 genomics. *Nucleic Acids Res.* **38**, D396–400 (2010).
- 853 75. Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein
854 sequences. *Protein Sci.* **27**, 135–145 (2018).
- 855 76. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and
856 annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–5 (2016).
- 857 77. Slabinski, L. *et al.* XtalPred: a web server for prediction of protein crystallizability.
858 *Bioinformatics* **23**, 3403–3405 (2007).
- 859 78. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web
860 portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
- 861 79. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein
862 structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).
- 863 80. Newby, Z. E. R. *et al.* A general protocol for the crystallization of membrane proteins for X-
864 ray structural investigation. *Nat. Protoc.* **4**, 619–637 (2009).
- 865 81. Khalifah, R. G. The Carbon Dioxide Hydration Activity of Carbonic Anhydrase. *J. Biol.*
866 *Chem.* **246**, 2561–2573 (1971).

867