

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

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

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

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

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

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

\* Corresponding author

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

## **Abstract**

Bacterial autotrophs often rely on CO<sub>2</sub> concentrating mechanisms (CCMs) to assimilate carbon. Although many CCM proteins have been identified, including a 200+ MDa protein organelle called the carboxysome, a systematic screen of CCM components has not been carried out. Here, we performed a genome-wide barcoded transposon screen to identify essential and CCM-related genes in the γ-proteobacterium *H. neapolitanus*. Our screen identified a two-gene operon encoding a domain of unknown function (PFAM:PF10070) and a putative cation transporter subunit (PFAM:PF00361) critical for CCM function. Physiological and biochemical assays demonstrated these two proteins, which we name DabA and DabB for “DABs accumulate bicarbonate,” assemble into a heterodimeric complex and function as an energy-coupled inorganic carbon pump. This analysis also reveals that DabA contains a putative zinc-binding site reminiscent of a β-carbonic anhydrase, suggesting a possible mechanism of activity. We further show that DabAB operons are found in diverse bacteria and archaea. Finally, we demonstrate that functional DabAB operons are present in the human pathogens *V. cholera* and *B. anthracis*. Based on these results, we propose that DABs constitute a new class of energized inorganic carbon pump and play a critical role in inorganic carbon metabolism throughout prokaryotic phyla.

## Introduction

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

Organisms that depend on Rubisco for growth often employ supplemental physiological mechanisms to improve its rate and specificity. These mechanisms are collectively termed  $CO_2$  concentrating mechanisms (CCMs) because they serve to concentrate  $CO_2$  at the active site of Rubisco so that carboxylation proceeds at its maximum rate and oxygenation is competitively inhibited<sup>6,9,10</sup>. All cyanobacteria and many chemotrophic proteobacteria have a CCM<sup>11,12</sup>. The bacterial CCM has garnered particular interest among bioengineers because it is well-understood, composed of only ~20 genes and operates inside single cells<sup>13</sup>. Detailed modeling suggests that transplantation of the bacterial CCM into crops might improve yields<sup>14,15</sup> and efforts towards transplantation are already underway<sup>16–18</sup>.

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

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

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

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

Here we show that the gene products of the DAB operon form a protein complex that is capable of energetically-coupled C<sub>i</sub> uptake. Both proteins are necessary for activity in our experiments and treatment with an ionophore abrogates DAB-mediated C<sub>i</sub> uptake. Structural homology modeling suggests that DabA contains a domain distantly homologous to a type II β-carbonic anhydrase. Indeed, we demonstrate that DabA binds zinc and depends on two cysteines, one histidine and one aspartic acid residue for activity, in a manner reminiscent of β-CAs<sup>45,46</sup>. Taken together, these results indicate that DABs are a novel class of C<sub>i</sub> pump, energized by coupling to a cation gradient (e.g. H<sup>+</sup> or Na<sup>+</sup>). Further, these results are consistent with a model of activity dependant on unidirectional hydration of CO<sub>2</sub> to HCO<sub>3</sub><sup>−</sup> in the cytosol via a CA-like mechanism. Phylogenomic analysis demonstrates that DAB operons are widespread throughout prokaryotes including carbon-fixing bacteria and archaea. Surprisingly, DAB operons

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

## Results

### *Transposon mutagenesis and gene essentiality*

Gene essentiality was defined by first constructing (Figure 1A) a randomly-barcoded genome-wide pooled knockout library of *Hnea* using conjugation<sup>47</sup>. The donor strain (*E. coli* APA 766) contains a vector with a barcoded Tn5-derived transposon encoding a kanamycin resistance marker. Conjugation was performed under 5% CO<sub>2</sub> so that CCM genes could be knocked out, and the resulting *Hnea* conjugants were selected for growth in the presence of kanamycin at 5% CO<sub>2</sub> to ensure transposon insertion.

The presence of a unique barcode in each transposon simplifies the use of the library for pooled screens using the ‘barseq’ approach<sup>47</sup>. Transposon insertion sites and associated barcodes must be mapped to the *Hnea* genome in order to perform these screens. Transposon insertions were mapped using standard TnSeq methods<sup>47</sup> and the library was found to contain ~10<sup>5</sup> insertions, or one insertion for every ≈25 base pairs in the *Hnea* genome. Since the average gene contains ≈35 insertions, genes with no insertions are very likely essential for growth<sup>48</sup>.

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

### *Comprehensive screen for *Hnea* CCM components*

Based on the current model of the bacterial CCM (Figure 2A) knockouts of CCM genes are expected to require high CO<sub>2</sub> for growth<sup>20,29,30</sup>. CCM gene knockouts should therefore have low fitness in ambient CO<sub>2</sub> concentrations. As our pooled library contains ~70,000 barcodes that map to exactly one position in the *Hnea* genome, we were able to use the barseq method to quantify the fitness defects associated with single gene knockouts for all nonessential *Hnea* genes (Figure 2B). In barseq, a preculture of the library is grown in permissive conditions (5% CO<sub>2</sub>) and then back-diluted into two conditions: a reference condition (5% CO<sub>2</sub> again) and a condition of interest (i.e. ambient CO<sub>2</sub>). Genomic DNA is extracted from the preculture (called t<sub>0</sub>)

and both culture outgrowths. Barcodes are then PCR-amplified and sequenced. In this pooled competition assay, the proportional change in barcode abundance is taken to reflect the fitness effect of gene knockouts<sup>47</sup>. A CCM gene knockout should have no fitness defect in 5% CO<sub>2</sub> but a large defect in ambient CO<sub>2</sub>. Since the library contains >20 knockouts with unique barcodes per gene (on average), these screens contain multiple internal biological replicates testing the effect of single gene knockouts.

As expected, knockouts to nearly all carboxysome-associated genes produced large fitness defects in ambient CO<sub>2</sub> (Figures 2B-C). These genes include *cbbLS* - the large and small subunits of the α-carboxysomal Rubisco; *csoS2* - an intrinsically disordered protein required for α-carboxysome assembly<sup>49</sup>; *csoSCA* - the carboxysomal carbonic anhydrase; *csoS4AB* - the pentameric proteins thought to form vertices of the α-carboxysome; and *csoS1CAB* - the hexamers that form the faces of the α-carboxysome shell<sup>12,42</sup>. Knockouts of *csoS1D*, a shell hexamer with a large central pore<sup>31,50</sup>, confer a very weak HCR phenotype in this screen and so *csoS1D* did not cross the threshold for being called HCR (Figures 2B-C). The *Hnea* genome also contains a secondary, non-carboxysomal Form II Rubisco that is likely not involved in CCM activity as its disruption confers no fitness defect in ambient CO<sub>2</sub>. A number of genes that are not structurally associated with the carboxysome also exhibited HCR phenotypes. These include two LysR transcriptional regulators, a Crp/Fnr type transcriptional regulator, a protein called acRAF that is involved in Rubisco assembly<sup>35,36</sup>, and two paralogous loci encoding DAB genes (hereafter DAB1 and DAB2, Figure 2B-F).

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

DAB1 is a cluster of 3 genes found in an operon directly downstream of the carboxysome operon (Figure 2C). Though DAB1 is part of a larger 11-gene operon containing several genes associated with Rubisco proteostasis, including acRAF<sup>35,36</sup> and a cbbOQ-type Rubisco activase<sup>51</sup>, we refer to DAB1 as an “operon” for simplicity. DAB2 is a true operon and is not proximal to the carboxysome operon in the *Hnea* genome. These “operons” are unified in that they both display HCR phenotypes and possess similar genes (Figures 2B-D).

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

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

In order to facilitate testing for C<sub>i</sub> transport abilities, we generated an *E. coli* strain, CAfree, that contains no CA genes (Methods). It was previously shown that deletion of the constitutive CA, *can*, gene produces an HCR phenotype in *E. coli*<sup>52</sup> that is complemented by expression of cyanobacterial bicarbonate transporters<sup>53</sup>. However, this approach is limited by the potential of a second, inducible CA, *cynT*, to act as a suppressor of the HCR phenotype. Deleting both of these genes removes this avenue for suppression of the phenotype. Since DAB2 disruption is associated with a larger fitness defect than DAB1 (Figure 2B), we used CAfree to test DAB2 for C<sub>i</sub> uptake activity. DAB2 expression enables growth of CAfree in ambient CO<sub>2</sub> while expression of either gene alone is not sufficient (Figures 3B and S4). Uptake assays using radiolabeled <sup>14</sup>C<sub>i</sub> demonstrates that DAB2 facilitates transport of C<sub>i</sub> levels significantly above that of steady-state (Figure 3C). Moreover, DAB2-associated C<sub>i</sub> uptake is strongly inhibited by the ionophore CCCP (white bars in Figure 3C), also indicating that DAB2 is energetically-coupled, either directly or indirectly, to a cation gradient (e.g. H<sup>+</sup> or Na<sup>+</sup>).

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

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

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

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

Aqueous CO<sub>2</sub> spontaneously interconverts with the gas phase as well as hydrated C<sub>i</sub> species (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(aq)</sub> and CO<sub>2(gas)</sub> is not affected by pH, but the conversion from CO<sub>2</sub> to hydrated C<sub>i</sub> is highly pH dependant. The result of this interaction is that in low volume, well mixed solutions, the equilibrium concentration of HCO<sub>3</sub><sup>-</sup> increases 100 fold between pH 5 and 7 without an accompanying change in CO<sub>2</sub> concentration (Figure S6A)<sup>9</sup>. SbtA, a known HCO<sub>3</sub><sup>-</sup> transporter, rescues CAfree growth at pH 7 but not at pH 5, while DabAB2 rescues growth at both pHs (Figure S7). Since DabAB2 rescue is pH-independent in this range, its substrate is likely CO<sub>2</sub> and H<sub>2</sub>CO<sub>3</sub>, HCO<sub>3</sub><sup>-</sup>, or CO<sub>3</sub><sup>-2</sup>.

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

Structural homology modeling software predicts that the middle of DabA2 has sequence elements related to a β-CA (Figure 3A). Specifically, Phyre2 predictions identify C539 and H524 as part of a potential Zn<sup>2+</sup> binding site distantly homologous to a bacterial type II β-CA (10% coverage of DabA, 90.8% confidence). I-TASSER predicts a Zn<sup>2+</sup> binding site including the same residues along with an additional cysteine (C351), and aspartic acid (D353). As shown in Figure 4C, these residues could make up the active site of a type II β-CA<sup>54-56</sup>. We generated individual alanine mutants for each of these putative active site residues (C351A, D353A, H534A and C539A) and tested their ability to rescue CAfree. All mutants failed to rescue CAfree in ambient CO<sub>2</sub> (Figure 4D). We proceeded to assay zinc binding of purified dabAB complex using X-ray fluorescence spectroscopy and found that wild-type dabAB2 and three of the single mutants (C351A, D353A, and H534A) bind zinc (Figure 4E). These single mutants retain three of four zinc-coordinating residues<sup>46</sup>, which could explain why the mutants appear to bind zinc. Indeed, mutational studies of the human CA II show that single mutations to Zn<sup>2+</sup>-binding residues reduce but do not abrogate zinc binding<sup>45,57</sup>.

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

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

(> 650-fold more protein than the positive control) and under CO<sub>2</sub> concentrations that are typically saturating for CAs, but showed as much CA activity as the negative control (Figure 4F). Absence of activity *in vitro* argues either that DabAB2 has extremely low CA activity or, more likely, that DabAB2 must reside in a cell membrane holding a cation gradient to function as an energetically-activated carbonic anhydrase.

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

Searching the Uniprot database with the DabA pfam (PF10070) and pruning truncated and poorly aligned sequences yielded 878 DabA sequences. DabA sequences were found in a wide variety of prokaryotes including bacteria and archaea (fig. 5A and S8). Represented clades include not only  $\gamma$ -Proteobacteria but also  $\alpha$ -Proteobacteria,  $\beta$ -Proteobacteria, Euryarchaeota, Firmicutes, Planctomycetes, and Bacteroides. Many DabA sequences were found in genomes of heterotrophic organisms that cannot fix CO<sub>2</sub>. Additionally the important heterotrophic human pathogens *V. cholera*, *B. anthracis*, and *L. pneumophila* contain apparent DabA homologs. We then wanted to know if these dabA genes were part of DAB operons, we defined a putative DAB operon as a i gene that had an identifiable *dabB* gene present as a direct fusion to *dabA* or within three genes and oriented in the same direction. Inspection of local gene neighborhoods of sequences from the tree revealed that 843 (96%) of *dabA* sequences from the tree are in putative DAB operons.

We tested whether DAB homologs from heterotrophic pathogens are functional C<sub>i</sub> pumps. *V. cholera* E7946 El Tor Ogawa and *B. anthracis* Sterne both contain putative DAB operons. These operons were cloned and expressed in CAfree *E. coli*. Expression of either of these DAB operons rescues growth of CAfree in ambient CO<sub>2</sub> (Fig. 5B and S9). Thus, the DAB operons from even non-autotrophic human pathogens contain functional variants.

## Discussion

Bacterial CCMs exist as two convergently evolved families the  $\alpha$ - and  $\beta$ -lineages. Both lineages require two major components: i. energy-coupled uptake of inorganic carbon to concentrate HCO<sub>3</sub><sup>-</sup> in the cytosol and ii. carboxysome structures that co-localize Rubisco with CA enzymes that convert concentrated HCO<sub>3</sub><sup>-</sup> into a high concentration of the Rubisco substrate CO<sub>2</sub><sup>9</sup>. While the carboxysome components are well-documented for both  $\alpha$ - and  $\beta$ -families, C<sub>i</sub> uptake systems of the proteobacterial CCM have only been identified very recently<sup>43,44</sup>. Moreover,

though several laboratories have spent decades studying the bacterial CCM, it remains unclear whether our current “parts list” for  $\alpha$ - and  $\beta$ -CCMs is complete.

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

We showed that the DAB2 operon encodes a two-component protein complex that has C<sub>i</sub> uptake activity when heterologously expressed in *E. coli* (Figure 3B-C). This complex is likely a heterodimer as suggested by size-exclusion chromatography (Figure 4B). As C<sub>i</sub> uptake is strongly inhibited by the ionophore CCCP (Figure 3C), we suspect that DAB2 activity is energetically-coupled to a cation gradient (Figure 5A). Since DabAB2 shows pH-independent rescue of CAfree *E. coli* (pH 5-7) CO<sub>2</sub> is likely the transported substrate (Figure 4C). This idea is further supported by the fact that DabA has limited homology to a type II  $\beta$ -CA and binds a zinc (Figures 3-4), which could indicate the presence of a CA active site that hydrates transported CO<sub>2</sub>. Mutations to the putative zinc-binding residues (C351A, D353A, H534A, and C539A) ablate function in-vivo (Figure 4D). For these reasons, we propose a model of DAB activity wherein CO<sub>2</sub> is passively taken into the cell and then vectorially (unidirectionally) hydrated to HCO<sub>3</sub><sup>-</sup> by energy-coupled CA activity of DabA.

Model carbonic anhydrases are not directly coupled to any energy source (e.g., ATP) and so they only accelerate the equilibration of CO<sub>2</sub> and HCO<sub>3</sub><sup>-</sup><sup>45,56</sup>. Energy coupled CA activity could produce unidirectional CO<sub>2</sub> hydration, allowing the DAB system to actively accumulate HCO<sub>3</sub><sup>-</sup> in the cytosol and power the CCM (as diagrammed in Figure 2A). Given the similarity of DabB to

other H<sup>+</sup>-pumping proteins, we propose the DAB is coupled to the H<sup>+</sup> gradient, but our results are equally consistent with other cation gradients, e.g. Na<sup>+</sup>. This mechanism would require tight coupling of cation flow to CO<sub>2</sub> hydration by the CA-like DabA protein, consistent with our observation that purified DabAB2 displays no measurable CA activity. Interestingly, type II β-CAs are the only CAs known to display allosteric regulation<sup>46,55</sup>. This allosteric control is thought to be mediated by Zn<sup>2+</sup> capping and uncapping by the active site aspartic acid (D353 in DabA2). A similar mechanism might couple cation movement through DabB to the active site of DabA.

Cyanobacteria also possess vectorial CA systems called CUPs, which may provide clues to the DAB mechanism<sup>37,38,58–60</sup>. Indeed, both DAB and CUP systems contain subunits in the Mrp protein family (DabB and NdhD/F are in PF00361) that also contains the H<sup>+</sup>-pumping subunits of complex I. This commonality might suggest a shared mechanism. CO<sub>2</sub> hydration by CupA/B is thought to be coupled to energetically-favorable electron flow because CupA/B proteins appear to associate with the cyanobacterial complex I<sup>39,61</sup> (Figure S8B). However, the Mrp protein family (PF00361) is very diverse and contains many cation transporters that do not associate with complex I or any other redox-coupled membrane complex<sup>43,62,63</sup>. Moreover, DabB and NdhD/F sequences are only distantly related to complex I subunits (Figure S3A), the two DAB subunits do not appear to bind *E. coli* complex I (Figure 4A) and DAB2 rescue of CAfree growth does not require complex I (Figure S5). We therefore propose that DAB activity is not coupled to electron flow through complex I but, rather, to a cation gradient across the membrane as described above (Figure 6).

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

engineering and heterologous expression in plants and industrial microbes, where elevated intracellular C<sub>i</sub> could be technologically useful<sup>64</sup>.

Finally, we were surprised to find evidence of active DABs outside of known carbon-fixing bacteria (Figure 5). High-confidence DabA homologs are found in many lineages, including notable heterotrophic pathogens including *V. cholerae*, *B. anthracis*, and *L. pneumophila* (Figure 5A). Moreover, we showed that DAB operons from *V. cholerae* and *B. anthracis* are active in *E. coli*, which leads us to wonder: what do heterotrophic pathogens need C<sub>i</sub> uptake systems for? Carbonic anhydrase activity is essential for growth of the heterotrophs *E. coli* and *S. cerevisiae* in ambient CO<sub>2</sub><sup>52,65</sup>. In the heterotrophic context, CA activity is thought to supply bicarbonate for the biotin-dependent carboxylases of central metabolism, for which HCO<sub>3</sub><sup>-</sup> is the true substrate<sup>52,65</sup>. Additionally, bicarbonate levels have been linked to virulence in both *V. cholera* and *B. anthracis*<sup>66-68</sup>. Perhaps DAB-family C<sub>i</sub> uptake systems play roles in the growth or virulence of these important pathogens? We hope that future research will delineate the role of energetically-activated C<sub>i</sub> uptake in heterotrophic and pathogenic organisms.

## Materials and Methods

### Important strains and reagents

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

### Bacterial strains and growth conditions

*E. coli* strain APA766 was used as the conjugation donor to transfer the Tn5 transposon to *Halothiobacillus neapolitanus* C2 (*Hnea*) via conjugation<sup>47</sup>. The *E. coli* double CA deletion strain “CAfree” (BW25113  $\Delta can \Delta cynT$ ) was generated by curing the KEIO collection *cynT* knockout (BW25113  $\Delta cynT$ , KEIO strain JW0330) of kanamycin resistance via pCP20-mediated FLP recombination and subsequent P1 transduction (and curing) of kanamycin resistance from the *can* knockout strain EDCM636 (MG1655  $\Delta can$ , Yale Coli Genomic Stock Center,<sup>52,69</sup>). Complex I knock out strains ( $\Delta(nuoA-nuoN)$ ) were generated in both the BW25113 and CAfree backgrounds. These strains were generated by lambda red mediated recombination of a Kan<sup>R</sup> resistance cassette flanked by frt sites into the *nuo* locus such that the entire operon was removed. This was followed by heat curing of the pSIM5 plasmid<sup>70</sup> expressing the lambda red recombinase at 42 °C. Lysogeny broth (LB) and LB agar were used as *E. coli* growth media unless otherwise specified. *E. coli* strains were grown at 37 °C in the presence of 0.1 mg/ml carbenicillin, 0.06 mg/ml kanamycin, or 0.025 mg/ml chloramphenicol as appropriate. *Hnea* was grown in DSMZ-68 media at 30 °C and in the presence of 0.03 mg/ml kanamycin when appropriate.

### Transposon mutagenesis and RB-TnSeq library production

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

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

#### *Essential gene assignment*

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

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

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

#### *Gene fitness experiments*

Fitness experiments were performed according to a modification of the protocol in Wetmore *et al.*<sup>47</sup>. A library aliquot was thawed and used to inoculate three 33 mL cultures. Cultures were grown to OD600 ~0.08 in 5% CO<sub>2</sub>. At this point, 20 mL were removed and harvested by centrifugation as two t<sub>0</sub> (input) samples. Cultures were back-diluted 1:64 into 128 mL and incubated for 6.5-7.5 doublings under 5% CO<sub>2</sub> or ambient conditions. 50 mL of culture was harvested by centrifugation. gDNA was prepared and barcodes were amplified for fitness determination via Illumina sequencing as described in Wetmore *et al.*<sup>47</sup>.

#### *CAfree rescue experiments*

Electrocompetent CAfree cells were prepared using standard protocols<sup>71</sup> and transformed with pFE plasmids expressing genes of interest by electroporation. CAfree pre-cultures were grown overnight in 10% CO<sub>2</sub> and diluted into 96 well plates (3 µl cells in 250 µl media). Growth curves

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

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

The silicone oil filtration method was modified from Dobrinski *et al.*<sup>72</sup> and used to measure uptake of radiolabeled inorganic carbon. Assay tubes were generated using 0.6 ml microcentrifuge tubes containing 20  $\mu$ l of dense kill solution (66.7% v/v 1 M glycine pH 10, 33.3% v/v triton X-100) covered by 260  $\mu$ l of silicone oil (4 parts AR20:3.5 parts AR200). Electrocompetent CAfree cells were prepared using standard protocols and transformed with pFA-based plasmids containing genes of interest by electroporation. CAfree cultures were grown overnight in 10% CO<sub>2</sub>, back diluted to an OD600 of 0.1 and allowed to grow to mid-log phase in 10% CO<sub>2</sub> in the presence of 200 nM aTc for induction. Cells were then harvested by centrifugation, washed once in PBS (pH 7.55) and resuspended to OD600 0.6 in PBS + 0.4% glucose. <sup>14</sup>C-labeled sodium bicarbonate (PerkinElmer) was added to a final concentration of 4.1 nM and an activity of 0.23  $\mu$ Ci. Cells were incubated with <sup>14</sup>C for 4 minutes before centrifugation at 17,000 x g for 4 minutes to separate cells from buffer. Pellets were clipped into scintillation vials containing 5 ml Ultima Gold scintillation fluid and 300  $\mu$ l 3M NaOH using microcentrifuge tube clippers or medium dog toenail clippers. Counts were measured on a PerkinElmer scintillation counter. <sup>14</sup>C counts are normalized to 1 OD600 unit of cells added. During inhibition assays, cells were incubated in PBS pH 7.55 with 0.4% glucose + 0.4% DMSO and the inhibitor (100  $\mu$ M CCCP) for 10 minutes before assay.

#### *Generation of DabA phylogenetic tree*

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

MAFFT and an approximate maximum likelihood phylogenetic tree was constructed using FastTree. Taxonomy was assigned to nodes in the tree based on NCBI taxonomy information for the genomes harboring each sequence. Genomic neighborhoods for each gene in the tree were determined using the EFIGNT online server<sup>73</sup> and genomes with a *dabB* gene within 3 genes of *dabA* and oriented in the same direction were considered to have full DAB operons. *dabAB* fusions were found by visual inspection of genomic neighborhoods from those genomes that did not have separate *dabB* genes located close to *dabA*.

#### *Generation of DabB phylogenetic tree*

DabB homologs were collected manually by searching MicrobesOnline for close homologs of four PF00361 members in the *Hnea* genome (*dabB1*, *dabB2*, *Hneap\_1953*, *Hneap\_1130*) and other characterized PF00361 members including *Synechococcus elongatus ndhF1*, *Synechococcus elongatus ndhF3*, and *Synechococcus elongatus ndhF4*. Genes were clustered to 95% similarity and genes with divergent operon structure were removed manually using MicrobesOnline treeview<sup>74</sup>. *nuoL* from *Escherichia coli*, *nqo12* from *Thermus thermophilus*, and *ndhF1/3/4* from *Thermosynechococcus elongatus* BP-1 were added as markers. ClustalOmega was used to construct a multiple sequence alignment and the resulting nearest-neighbor tree was visualized using the Interactive Tree of Life<sup>75,76</sup>.

#### *Protein annotation and structural homology modeling*

Secondary structural annotations for DabA and DabB were generated using XtalPred<sup>77</sup>. Structural Homology modeling of DabA was performed using Phyre2 and I-TASSER web servers with default parameters<sup>78,79</sup>. A list of close DabB homologs was assembled by searching MicrobesOnline for PF00361 members with similar operon structure. A ClustalOmega alignment was used to calculate residue-level conservation of DabB proteins while the MAFFT alignment generated during the creation of the DabA tree was used to calculate residue level conservation of DabA proteins (Figure S3B).

#### *Purification of DAB2*

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

NaCl, pH 7.5) supplemented with 1.2 mM phenylmethylsulfonyl fluoride, 0.075 mg/ml lysozyme and 0.8 ug/ml DNase I per liter of starting culture and then incubated at room temperature on a rocker for 20 minutes. Cells were lysed with four passes through a homogenizer (Avestin). Lysate was clarified at 15,000 x g for 30 minutes. Membranes were pelleted at 140,000 x g for 90 minutes. Membrane pellets were resuspended overnight in 25 ml TBS supplemented with 1 mM phenylmethylsulfonyl fluoride and 1% β-dodecyl-maltoside (DDM, Anatrace) per liter of culture following<sup>80</sup>. Membranes were then re-pelleted at 140,000 - 200,000 x g for 60 minutes and the supernatant was incubated with Ni-NTA beads (Thermo Fisher) for 90 min at 4 °C. The resin was washed with “Ni buffer” (20 mM Tris + 300 mM NaCl + 0.03% DDM, pH 7.5) supplemented with 30 mM imidazole and eluted with Ni buffer supplemented with 300 mM imidazole. Eluent was then incubated with Strep-Tactin (Millipore) resin for 90 min at 4 °C. Resin was washed with “strep buffer” (TBS + 0.03% DDM) and eluted with strep buffer supplemented with 2.5 mM desthiobiotin. Eluent was concentrated using Vivaspin 6 100 kDa spin concentrators and buffer exchanged into strep buffer by either spin concentration or using Econo-Pac 10DG (Biorad) desalting columns. For analytical purposes, 300 µg of strep-purified protein was injected onto a Superdex 200 Increase 3.2/300 size-exclusion column pre-equilibrated in strep buffer and eluted isocratically in the same buffer.

#### *Carbonic anhydrase assays*

CA-catalyzed CO<sub>2</sub> hydration of purified DAB2 complex and human carbonic anhydrase (hCA) was measured using the buffer/indicator assay of Khalifah<sup>81</sup> on a KinTek AutoSF-120 stopped-flow spectrophotometer at 25 °C. The buffer/indicator pair used was TAPS/m-cresol purple measured at a wavelength of 578 nm using a pathlength of 0.5 cm. Final buffer concentration 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>, and 50 µM of pH-indicator. Final protein concentration used was: 9.8 µM DAB2 (His-elution) and 0.015 µM hCA (positive control; Sigma Aldrich C6624). Saturated solution of CO<sub>2</sub> (32.9 mM) was prepared by bubbling CO<sub>2</sub> gas into milli-Q water at 25 °C. The saturated solution was injected into the stopped-flow using a gas-tight Hamilton syringe, and measurements were performed in a final CO<sub>2</sub> concentration of 16.5 mM. Progression curves were measured in 7 replicates.

#### *X-ray fluorescence spectroscopy for metal analysis*

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

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

## **Acknowledgements**

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

## **Data availability**

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

## **Code availability**

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

## **Author contributions**

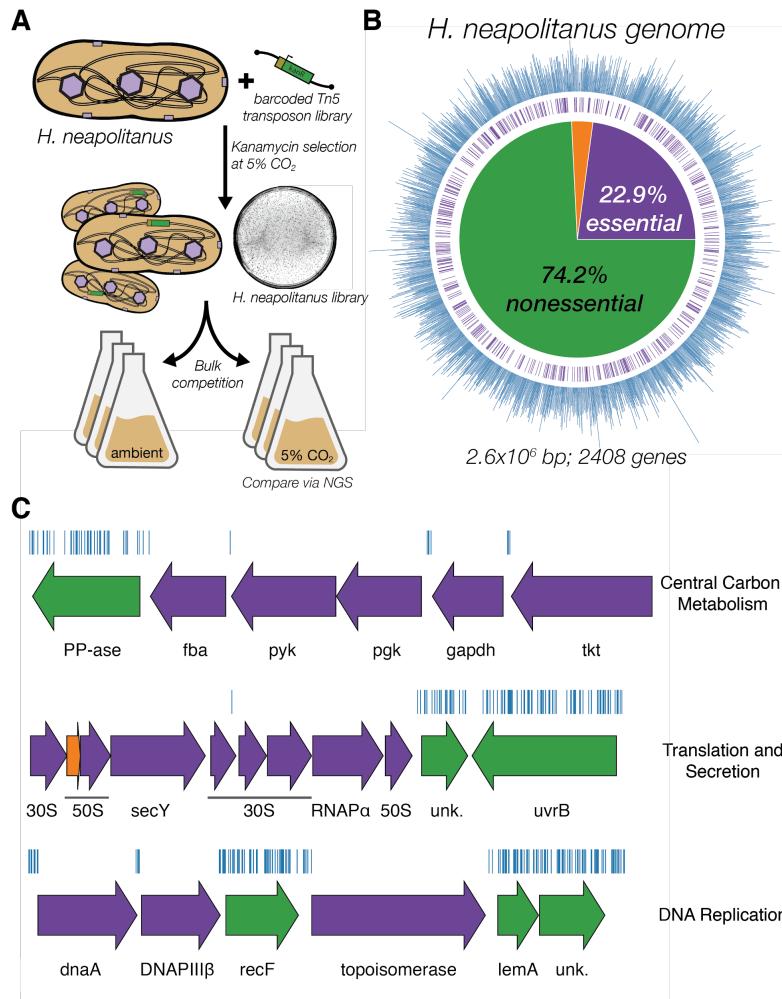
J.J.D., A.I.F., and D.F.S. conceived and designed this study, and wrote the final manuscript with input from all authors; J.J.D., A.I.F., C.B., E.J.D., T.G.L., L.M.O., A.W.C., S.D., K.W., J.Y.W., and D.F.S. conducted the research or interpreted results. All authors reviewed and approved the final manuscript.

## **Competing Interests**

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

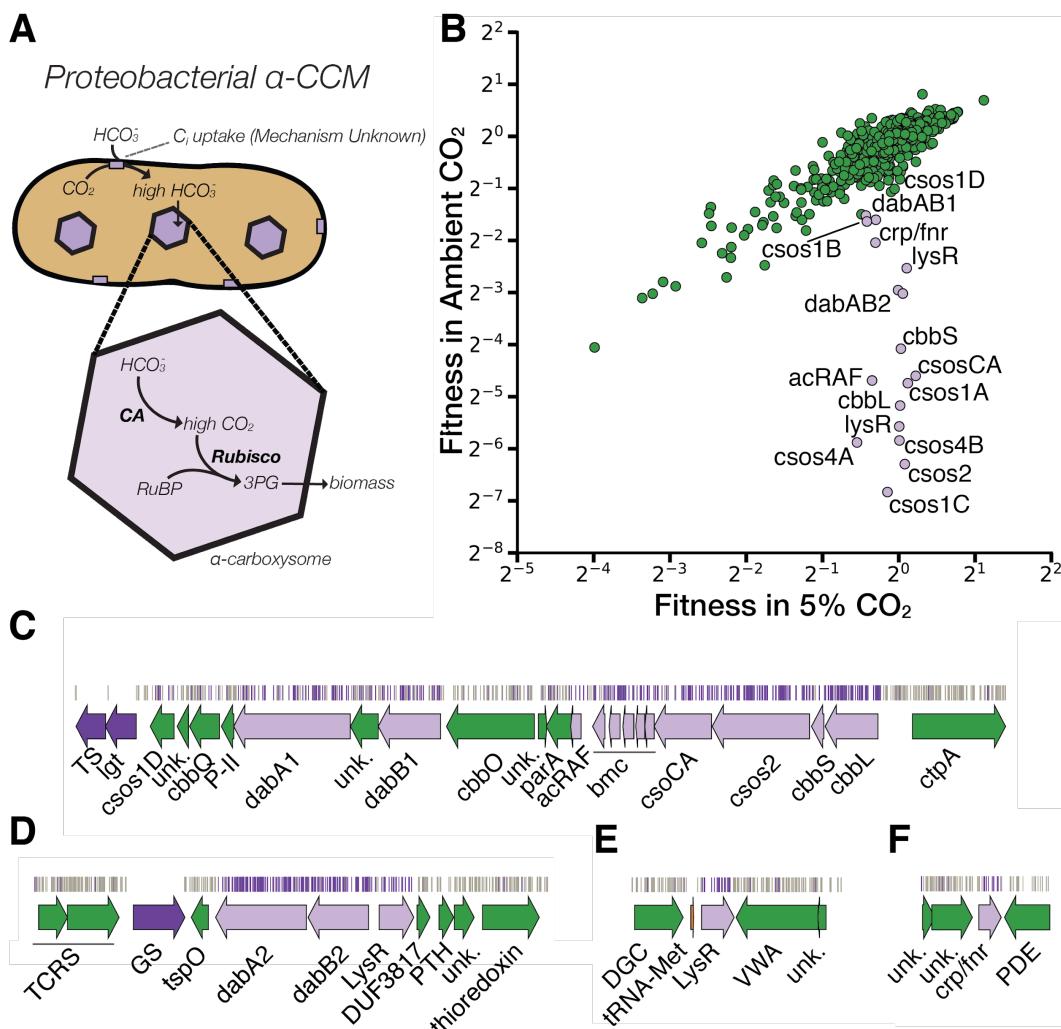
**Materials & Correspondence**

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



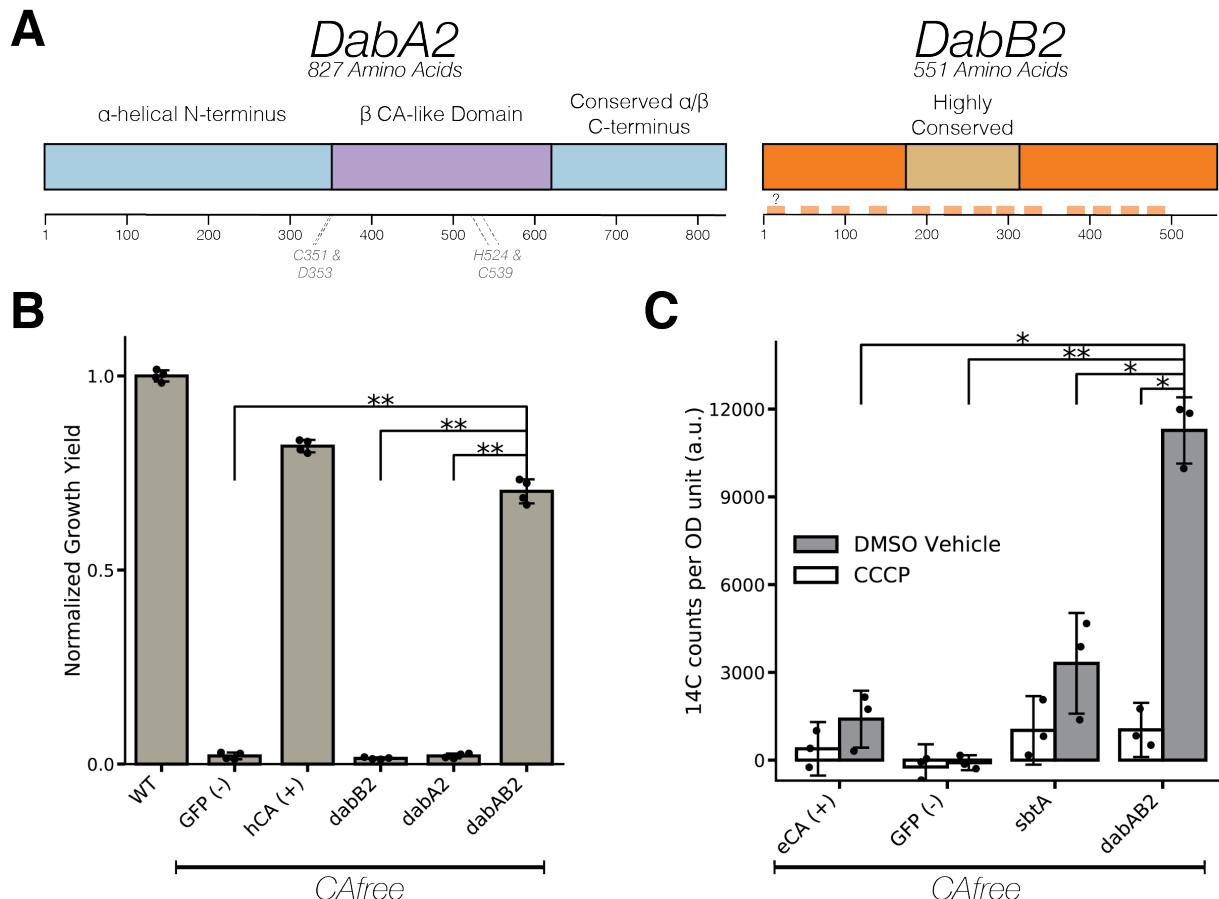
**Figure 1. Transposon mutagenesis reveals the essential gene set of a chemoautotrophic organism.** **A.** Schematic depicting the generation and screening of the RB-TnSeq library. Transposons were inserted into the *Hne*a genome by conjugation with an *E. coli* donor strain. The transposon contains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

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



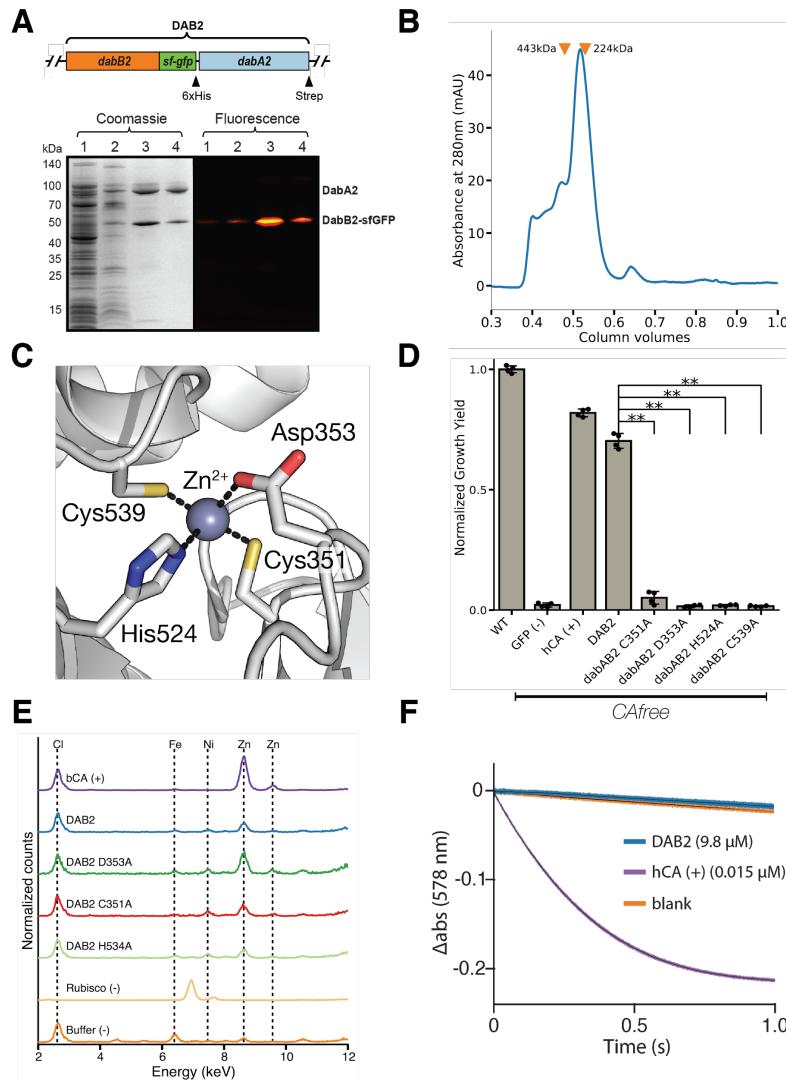
**Figure 2. A systematic screen for high  $\text{CO}_2$ -requiring mutants identifies genes putatively associated with the CCM.** **A.** Simplified model of the  $\alpha$ -CCM of chemotrophic proteobacteria. Inorganic carbon is concentrated via an unknown mechanism, producing a high cytosolic  $\text{HCO}_3^-$  concentration. High cytosolic  $\text{HCO}_3^-$  is converted into high carboxysomal  $\text{CO}_2$  by CA, which is localized only to the carboxysome. **B.** Fitness effects of gene knockouts in 5%  $\text{CO}_2$  as compared to ambient  $\text{CO}_2$ . Data is from one of two replicates of the BarSeq - the second replicate gives consistent results. When the effect of single transposon insertions into a gene are mutually consistent, those effects are averaged to produce the gene-level fitness value plotted<sup>47</sup>. We define HCR mutants as those displaying a twofold fitness defect in ambient  $\text{CO}_2$  relative to 5%  $\text{CO}_2$ . HCR genes are colored light purple. Panels **C-F** show regions of the *Hnea* genome containing genes annotated as HCR in panel A. Essential genes are in dark purple, HCR genes are in light purple, and other genes are in green. The top tracks show the presence of an insertion in that location. Insertions are colored grey unless they display a twofold or greater fitness defect in

ambient CO<sub>2</sub>, in which case they are colored light purple. **C.** The gene cluster containing the carboxysome operon and a second CCM-associated operon. This second operon contains acRAF, a Form IC associated cbbOQ-type Rubisco activase and *dabAB1*. **D.** The DAB2 operon and surrounding genomic context. **E.** The genomic context of a lysR-type transcriptional regulator that shows an HCR phenotype. **F** Genomic context of a crp/fnr-type transcriptional regulator that displays an HCR phenotype. Genes labeled “unk.” are hypothetical proteins. Full gene names are given in Figure S2.



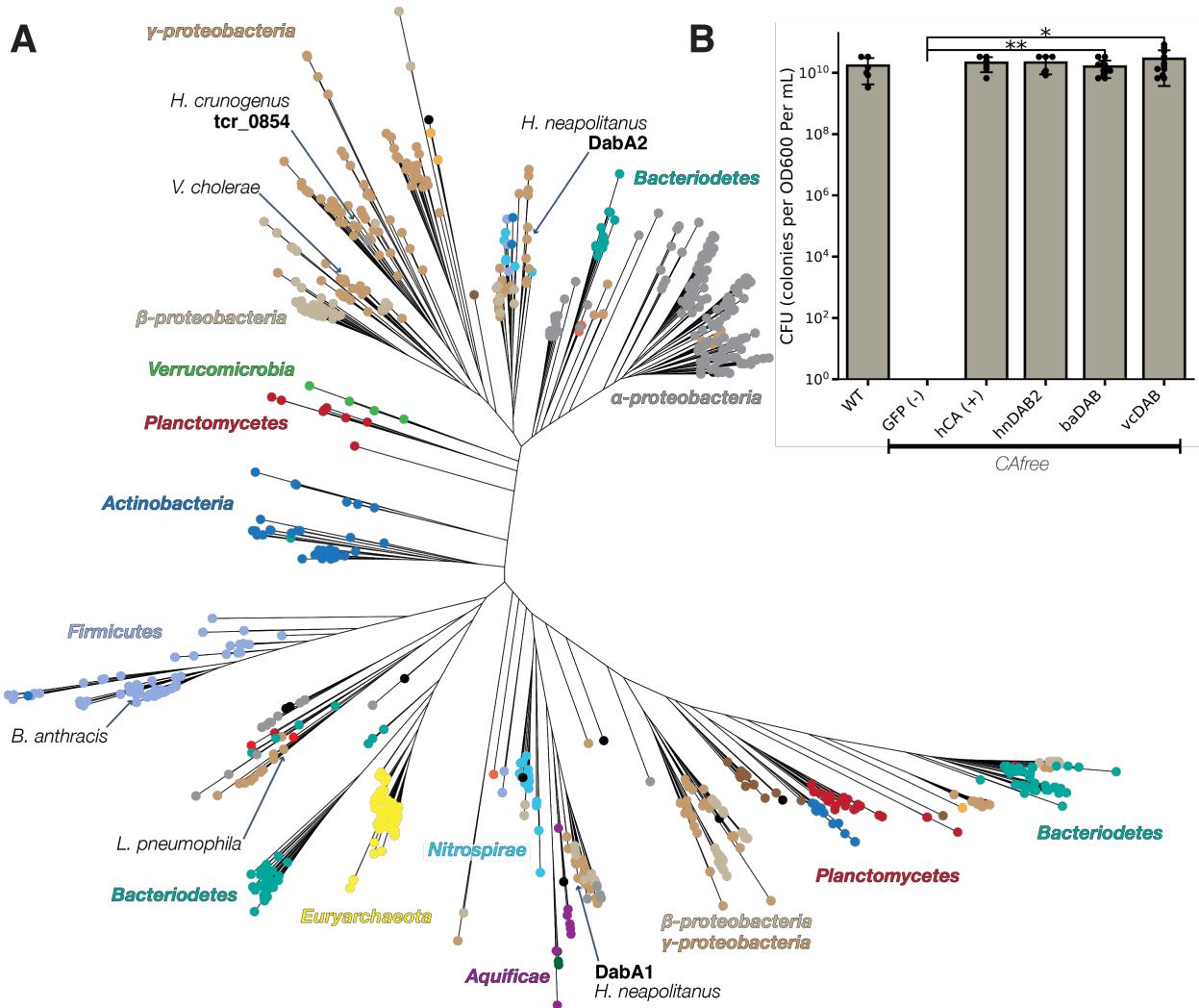
**Figure 3. The DABs catalyze active transport of  $\text{C}_i$  and are energized by a cation gradient. A.** Diagrammatic representation of DabA2 and DabB2 based on bioinformatic annotation. DabA2 is an 827 amino acid protein with predicted homology to a type II  $\beta$ -CA enzyme. The four predicted active site residues (C351, D353, H524, C539) are marked on the primary amino acid sequence. DabB2 is a 551 amino acid protein with 12-13 transmembrane helices. There is a highly conserved region in the middle of its sequence and predicted transmembrane helices are marked in light orange. **B.** DAB2 was tested for ability to rescue growth of CAfree *E. coli* in ambient  $\text{CO}_2$  conditions. Expression of the full operon (DabAB2) rescues growth, as does the positive control, and human carbonic anhydrase II (hCA). Error bars represent standard deviations of 4 replicate cultures. **C.** CAfree *E. coli* were tested for  $\text{C}_i$  uptake using the silicone-oil centrifugation method. Expression of DabAB2 produced a large and statistically significant increase in  $^{14}\text{C}$  uptake as compared to all controls. Moreover, treatment with the ionophore CCCP greatly reduces DabAB2-mediated  $^{14}\text{C}$  uptake, suggesting that DabAB2 is coupled to a cation gradient. *E. coli* CA (eCA) was used as a control for a non-vectorial CA. *Synechococcus elongatus* PCC 7942 sbtA was used as a known  $\text{C}_i$  transporter. GFP was used as a vector control. Error bars represent standard deviations of 3 technical replicates. In (B) and (C) “\*\*” denotes that the means are significantly different with Bonferroni corrected  $P < 0.05$  according to a two-tailed T-test. “\*\*\*” denotes  $P < 5 \times 10^{-4}$ . In panel B, dabAB2 has a larger rescue than GFP ( $t=42.6$ , corrected  $p = 3.37 \times 10^{-8}$ ), dabA2 ( $t=43.4$ ,

corrected  $p = 3.02 \times 10^{-8}$ ), and dabB2 ( $t=44.5$ , corrected  $p = 2.6 \times 10^{-8}$ ). In panel C, dabAB2 expressing cells treated with DMSO have greater uptake than dabAB2 expressing cells treated with CCCP ( $t=13.6$ , corrected  $p = 6.81 \times 10^{-4}$ ), sbtA expressing cells treated with DMSO ( $t=6.71$ , corrected  $p = 1.03 \times 10^{-2}$ ), GFP expressing cells treated with DMSO ( $t=17.1$ , corrected  $p = 2.76 \times 10^{-4}$ ), or eCA expressing cells treated with DMSO ( $t=11.5$ , corrected  $p = 1.31 \times 10^{-3}$ ).



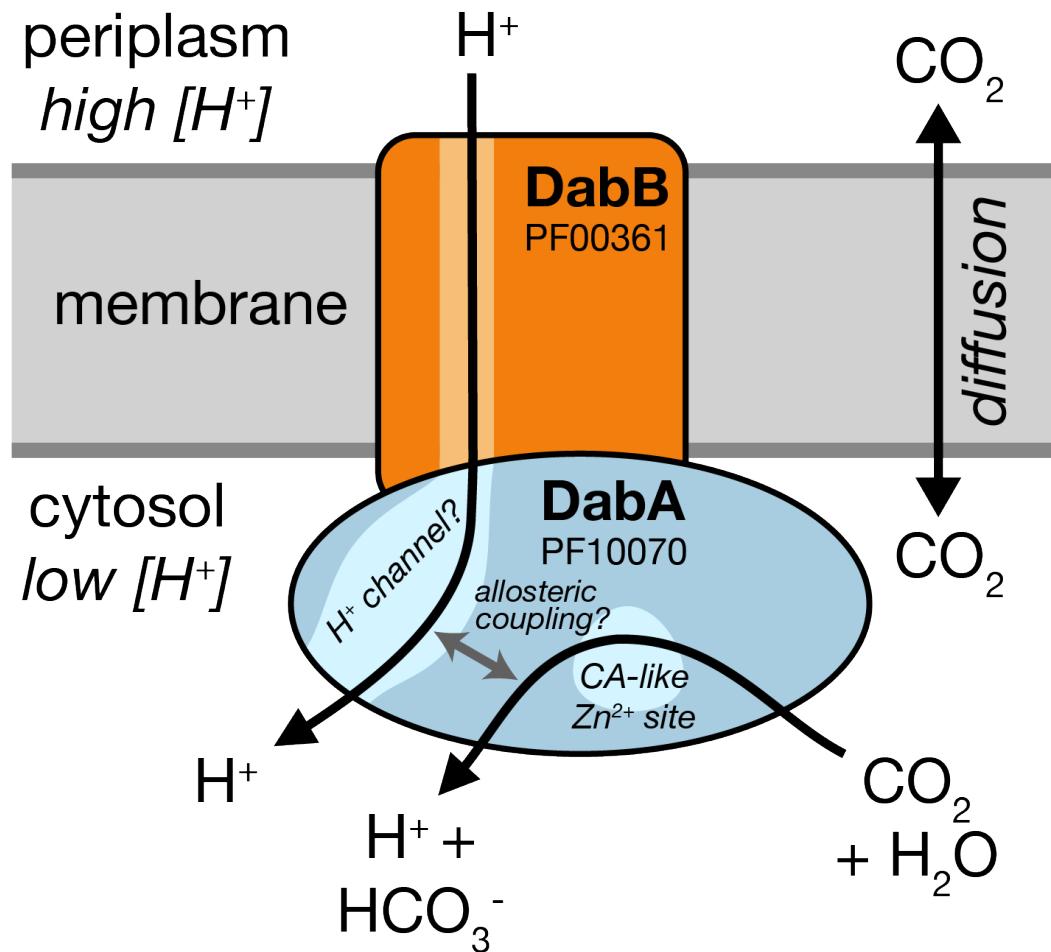
**Figure 4. DabA contains a  $\beta$ -CA-like active site but is not active outside of the membrane.** **A.** Purification of DabAB2 complex from *E. coli*. DabA2 was C-terminally tagged to a Strep-tag and DabB2 was C-terminally tagged with sf-GFP and a 6xHis-tag. Purification was monitored using SDS-PAGE imaged with fluorescence (right view) before coomassie staining (left view). Lane 1: clarified lysate; 2: solubilized membranes; 3: Ni-NTA resin eluent; 4: strep-tactin resin eluent. DabA2 and DabB2 co-purify as a single complex without any obvious interactors. **B.** Size-exclusion chromatogram of His/Strep purified DabAB2 with retention volumes (orange arrows) and molecular weights (kDa) indicated for standard samples (apo ferritin, 443 kDa;  $\beta$ -amylase, 224 kDa). DabAB2 runs with a mass of ~270 kDa, which is likely an oligomer of DabA and DabB. **C.** Structural model of the DabA2 active site based on a  $\beta$ -

CA of *E. coli* (PDB 1I6P). Typical  $\beta$ -CAs rely on two cysteine and one histidine residues to bind  $Zn^{2+}$ . The aspartic acid coordinates  $Zn^{2+}$  but is likely displaced during catalysis<sup>55</sup>. **D.** Alanine mutants of the putative DabA2 active site residues (C351A, t=54.3, p=1.05\*10<sup>-8</sup>; D353A, t=144, p=3.06\*10<sup>-11</sup>; H524A, t=44, p=3.68\*10<sup>-8</sup>; C539A, t=44.3, p=3.54\*10<sup>-8</sup>;) abrogate rescue of CAfree *E. coli* compared to wild-type dabAB2. Error bars give standard deviations of four replicates. “\*\*” denotes that means differ with bonferroni corrected P < 0.05 by a two-tailed T-test, and “\*\*\*” denotes P < 5X10<sup>-4</sup>. **E.** X-ray fluorescence data indicate that DabAB2 binds zinc like all known  $\beta$ -CAs. Single mutations to the active site do not abrogate zinc binding. **F.** Purified DabAB2 does not display any obvious CA activity despite being present in 650-fold excess over the positive control (Human carbonic anhydrase II, hCA) in our assays.



**Figure 5. DAB operons are widespread in prokaryotes. A.** Approximate maximum likelihood phylogenetic tree of DabA homologs associated with PF10070.9 (Methods). DabA homologs are found in > 15 prokaryotic clades, including archaea. *Hne*a DabA1 and DabA2 represent two different groupings that are commonly found in proteobacteria. Inspecting the tree reveals several likely incidents of horizontal transfer, e.g. between Proteobacteria and Firmicutes, Nitrospirae and Actinobacteria. Moreover, the genomes of several known pathogens contain a high-confidence DabA homolog, including *B. anthracis*, *V. cholerae*, and *L. pneumophila*. Detailed annotations are given in Figure S8. **B.** Functional DABs are found in human pathogens. Colony forming units per OD<sub>600</sub> per ml were measured on LB plates with induction in air. DAB operons from *B. anthracis* (baDAB,  $t=5.98$ ,  $p=1.84 \times 10^{-4}$ ) and *V. cholerae* (vcDAB,  $t=3.97$ ,  $p=4.37 \times 10^{-3}$ ) rescued growth of CAfree cells. The *Hne*a operon DAB2 is abbreviated as hnDAB2. Error bars represent the standard deviation of 6 replicate platings for WT, GFP (-), hCA (+), and hnDAB2. Error bars represent standard deviations of 12 replicate platings for baDAB and vcDAB. “\*\*”

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



**Figure 6 A model of the unidirectional energy-coupled CA activity of DAB complexes.** We propose that DabAB complexes couple CA activity of DabA to a cation gradient across the cell membrane, producing unidirectional hydration of  $CO_2$  to  $HCO_3^-$ . The cation gradient could be  $H^+$  or  $Na^+$ . Energy-coupled CA activity is required for the DABs role as a  $C_i$  uptake system in the proteobacterial CCM, as discussed in the text. Because it appears that DabAB2 is not active as a purified complex outside of the membrane, it is assumed protein tightly couples the inflow of cations with  $CO_2$  hydration so that there is no “slippage.” Indeed, slippage - i.e., uncoupled CA activity - would be counterproductive for CCM function<sup>9,19</sup>. Notably,  $Zn^{2+}$  binding by the active site aspartic acid of type II  $\beta$ -CAs (D353 in DabA2, Figure 4A) is thought to allosterically regulate activity<sup>46,55</sup>. This Asp-mediated activity switch could, therefore, provide a means for allosteric coupling of a  $\beta$ -CA active site to distal ion transport.

## References

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

- components, their diversity and evolution. *J. Exp. Bot.* **54**, 609–622 (2003).
12. Cannon, G. C. *et al.* Microcompartments in prokaryotes: carboxysomes and related polyhedra. *Appl. Environ. Microbiol.* **67**, 5351–5361 (2001).
  13. Long, B. M., Rae, B. D., Rolland, V., Förster, B. & Price, G. D. Cyanobacterial CO<sub>2</sub>-concentrating mechanism components: function and prospects for plant metabolic engineering. *Curr. Opin. Plant Biol.* **31**, 1–8 (2016).
  14. Price, G. D., Badger, M. R. & von Caemmerer, S. The prospect of using cyanobacterial bicarbonate transporters to improve leaf photosynthesis in C3 crop plants. *Plant Physiol.* **155**, 20–26 (2011).
  15. McGrath, J. M. & Long, S. P. Can the cyanobacterial carbon-concentrating mechanism increase photosynthesis in crop species? A theoretical analysis. *Plant Physiol.* **164**, 2247–2261 (2014).
  16. Lin, M. T., Occhialini, A., Andralojc, P. J., Parry, M. A. J. & Hanson, M. R. A faster Rubisco with potential to increase photosynthesis in crops. *Nature* **513**, 547–550 (2014).
  17. Occhialini, A., Lin, M. T., Andralojc, P. J., Hanson, M. R. & Parry, M. A. J. Transgenic tobacco plants with improved cyanobacterial Rubisco expression but no extra assembly factors grow at near wild-type rates if provided with elevated CO<sub>2</sub>. *Plant J.* **85**, 148–160 (2016).
  18. Long, B. M. *et al.* Carboxysome encapsulation of the CO<sub>2</sub>-fixing enzyme Rubisco in tobacco chloroplasts. *Nat. Commun.* **9**, 3570 (2018).
  19. Price, G. D. & Badger, M. R. Expression of Human Carbonic Anhydrase in the Cyanobacterium *Synechococcus* PCC7942 Creates a High CO<sub>2</sub>-Requiring Phenotype Evidence for a Central Role for Carboxysomes in the CO<sub>2</sub> Concentrating Mechanism. *Plant Physiol.* **91**, 505–513 (1989).
  20. Price, G. D. & Badger, M. R. Isolation and characterization of high CO<sub>2</sub>-requiring-mutants of the cyanobacterium *Synechococcus* PCC7942: two phenotypes that accumulate

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

31. Bonacci, W. *et al.* Modularity of a carbon-fixing protein organelle. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 478–483 (2012).
32. Fang, Y. *et al.* Engineering and Modulating Functional Cyanobacterial CO<sub>2</sub>-Fixing Organelles. *Front. Plant Sci.* **9**, 739 (2018).
33. Jorda, J., Lopez, D., Wheatley, N. M. & Yeates, T. O. Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci.* **22**, 179–195 (2013).
34. Axen, S. D., Erbilgin, O. & Kerfeld, C. A. A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput. Biol.* **10**, e1003898 (2014).
35. Wheatley, N. M., Sundberg, C. D., Gidaniyan, S. D., Cascio, D. & Yeates, T. O. Structure and Identification of a Pterin Dehydratase-like Protein as a Ribulose-bisphosphate Carboxylase/Oxygenase (RuBisCO) Assembly Factor in the  $\alpha$ -Carboxysome. *J. Biol. Chem.* **289**, 7973–7981 (2014).
36. Aigner, H. *et al.* Plant RuBisCo assembly in *E. coli* with five chloroplast chaperones including BSD2. *Science* **358**, 1272–1278 (2017).
37. Shibata, M., Ohkawa, H., Katoh, H., Shimoyama, M. & Ogawa, T. Two CO<sub>2</sub> uptake systems in cyanobacteria: four systems for inorganic carbon acquisition in *Synechocystis* sp. strain PCC6803. *Funct. Plant Biol.* **29**, 123–129 (2002).
38. Price, G. D. Inorganic carbon transporters of the cyanobacterial CO<sub>2</sub> concentrating mechanism. *Photosynth. Res.* **109**, 47–57 (2011).
39. Battchikova, N., Eisenhut, M. & Aro, E. M. Cyanobacterial NDH-1 complexes: Novel insights and remaining puzzles. *Biochimica et Biophysica Acta - Bioenergetics* **1807**, 935–944 (2011).
40. Artier, J., Holland, S. C., Miller, N. T., Zhang, M. & Burnap, R. L. Synthetic DNA system for structure-function studies of the high affinity CO<sub>2</sub> uptake NDH-13 protein complex in cyanobacteria. *Biochim. Biophys. Acta Bioenerg.* (2018). doi:10.1016/j.bbabi.2018.06.015

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).
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).
43. Mangiapia, M. *et al.* Proteomic and mutant analysis of the CO<sub>2</sub> concentrating mechanism of hydrothermal vent chemolithoautotroph *Thiomicrospira crunogena*. *J. Bacteriol.* (2017). doi:10.1128/JB.00871-16
44. Scott, K. M. *et al.* Genomes of ubiquitous marine and hypersaline *Hydrogenovibrio*, *Thiomicrospira rhabdus* and *Thiomicrospira* spp. encode a diversity of mechanisms to sustain chemolithoautotrophy in heterogeneous environments. *Environ. Microbiol.* **20**, 2686–2708 (2018).
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).
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).
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).
48. Rubin, B. E. *et al.* The essential gene set of a photosynthetic organism. *Proceedings of the National Academy of Sciences* 201519220 (2015).
49. Chaijarasphong, T. *et al.* Programmed Ribosomal Frameshifting Mediates Expression of the α-Carboxysome. *J. Mol. Biol.* **428**, 153–164 (2016).
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).

51. Mueller-Cajar, O. The Diverse AAA+ Machines that Repair Inhibited Rubisco Active Sites. *Front Mol Biosci* **4**, 31 (2017).
52. Merlin, C. & Masters, M. Why is carbonic anhydrase essential to Escherichia coli? *J. Bacteriol.* **185**, (2003).
53. Du, J., Förster, B., Rourke, L., Howitt, S. M. & Price, G. D. Characterisation of Cyanobacterial Bicarbonate Transporters in E. coli Shows that SbtA Homologs Are Functional in This Heterologous Expression System. *PLoS One* **9**, e115905 (2014).
54. Cronk, J. D., Endrizzi, J. a., Cronk, M. R., O'Neill, J. W. & Zhang, K. Y. Crystal structure of E. coli beta-carbonic anhydrase, an enzyme with an unusual pH-dependent activity. *Protein Sci.* **10**, 911–922 (2001).
55. Cronk, J. D. *et al.* Identification of a novel noncatalytic bicarbonate binding site in eubacterial beta-carbonic anhydrase. *Biochemistry* **45**, 4351–4361 (2006).
56. Supuran, C. T. Structure and function of carbonic anhydrases. *Biochem. J* **473**, 2023–2032 (2016).
57. Ippolito, J. A. *et al.* Structure of His94-->Asp carbonic anhydrase II in a new crystalline form reveals a partially occupied zinc binding site. *Protein Eng.* **8**, 975–980 (1995).
58. Shibata, M. *et al.* Distinct constitutive and low-CO<sub>2</sub>-induced CO<sub>2</sub> uptake systems in cyanobacteria: genes involved and their phylogenetic relationship with homologous genes in other organisms. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 11789–11794 (2001).
59. Rae, B. D., Long, B. M., Badger, M. R. & Price, G. D. Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO<sub>2</sub> fixation in cyanobacteria and some proteobacteria. *Microbiol. Mol. Biol. Rev.* **77**, 357–379 (2013).
60. Maeda, S.-I., Badger, M. R. & Price, G. D. Novel gene products associated with NdhD3/D4-containing NDH-1 complexes are involved in photosynthetic CO<sub>2</sub> hydration in the cyanobacterium, Synechococcus sp. PCC7942. *Mol. Microbiol.* **43**, 425–435 (2002).

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

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