DABs are inorganic carbon pumps found throughout prokaryotic phyla

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

Bacterial autotrophs often rely on CO_2 concentrating mechanisms (CCMs) to assimilate carbon. Although many CCM proteins have been identified, a systematic screen of the components of CCMs is lacking. Here, we performed a genome-wide barcoded transposon screen to identify essential and CCM-related genes in the γ-proteobacterium *Halothiobacillus neapolitanus*. Screening revealed that the CCM comprises at least 17 and probably no more than 25 genes, most of which are encoded in 3 operons. Two of these operons (*DAB1* and *DAB2*) contain a two-gene locus that encodes a domain of unknown function (Pfam: PF10070) and a putative cation transporter (Pfam: PF00361). Physiological and biochemical assays demonstrated that these proteins—which we name DabA and DabB, for DABs accumulate bicarbonate—assemble into a heterodimeric complex, which contains a putative β-carbonic anhydrase-like active site and functions as an energy-coupled inorganic carbon (C_i) pump. Interestingly, *DAB* operons are found in a diverse range of bacteria and archaea. We demonstrate that functional DABs are present in the human pathogens *Bacillus anthracis* and *Vibrio cholerae*. On the basis of these results, we propose that DABs constitute a class of energized C_i pumps and play a critical role in the metabolism of C_i throughout prokaryotic phyla.

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) is the primary carboxylase of the Calvin–Benson–Bassham (CBB) cycle and the major entry point of C_i into the biosphere. The activity of Rubisco is critical for agriculture and a major contributor to the removal of anthropogenic CO_2 from the atmosphere. Despite its centrality and abundance, Rubisco is not a fast enzyme[1,2], nor is Rubisco specific—all known Rubiscos can use molecular oxygen (O_2) as a substrate[3] in place of CO_2. Oxygenation does not fix carbon and produces a product that must be recycled through metabolically expensive photorespiratory pathways[4]. Many studies support the hypothesis that improvements to Rubisco could improve crop yields, but Rubisco has proved recalcitrant to protein engineering. It remains unclear whether or how Rubisco can be improved[2,5,6].

Organisms that depend on Rubisco for growth often use CCMs that concentrate CO_2 near Rubisco so that carboxylation can proceed at high rate and specificity[7,8]. All cyanobacteria and many chemolithoautotrophic proteobacteria have a CCM[9]. The bacterial CCM has garnered particular interest among bioengineers because it is well understood; it is thought to consist of relatively few genes and operates inside single cells[10]. Detailed modelling suggests that the transplantation of the bacterial CCM into crops might improve yields[11,12], and efforts towards transplantation are already underway[13].

A diverse range of previous experimental studies showed that bacterial CCM requires two components to function—active transport of C_i that drives the accumulation of HCO_3^- in the cytosol and organization of Rubisco with carbonic anhydrase (CA) in the lumen of a protein organelle called the carboxysome[7,14,15]. Energy-coupled C_i pumps maintain a high cytosolic HCO_3^- concentration (>10 mM) and, crucially, keep HCO_3^- out of equilibrium with CO_2

(refs. [14,16,17]). CA activity interconverts $HCO_3^- + H^+$ with $CO_2 + H_2O$ and, therefore, the carboxysomal CA converts a high cytosolic HCO_3^- concentration into a high carboxysomal CO_2 concentration, promoting faster carboxylation by Rubisco and competitively inhibiting oxygenation[7]. Genetic mutations to either component—C_i-uptake systems or carboxysomes—disrupt the CCM, and mutants have growth defects unless CO_2 is supplemented[18,19]. This high-CO_2-requiring (HCR) mutant phenotype is commonly used to identify CCM components in screens[18,19].

A comprehensive list of bacterial CCM components remains unknown despite previous screens, leaving the possibility that further activities are required for CCM function. Although well-assembled carboxysome structures can be produced in bacteria and plants[13,20], the functionality of these carboxysomes in a heterologous CCM has not been demonstrated. Bioinformatics studies show that several non-carboxysomal genes are associated with carboxysome operons[21,22]. Experimental[14,23] and modelling studies[7,15] also make it clear that energy-coupled C_i-uptake systems are required for the function of CCMs. Several different C_i-pump families—including transporters and facilitated uptake systems—have been described in cyanobacterial lineages, but mechanistic understanding of the uptake of C_i is limited[24].

We used a genome-wide barcoded transposon mutagenesis screen (RB-TnSeq) to study the CCM of *H. neapolitanus*. *H. neapolitanus* is a sulfur-oxidizing γ-proteobacterial chemolithoautotroph and a model system for studying α-carboxysomes[25–27]. Previous physiological measurements suggest that *H. neapolitanus* possesses an energized C_i-uptake system, but the molecular identity of this activity is unknown[17]. In addition to producing the essential gene set for a bacterial chemolithoautotroph, we leveraged our pooled mutant library to comprehensively screen for knockouts that

[1]Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA. [2]Department of Chemistry, University of California, Berkeley, CA, USA. [3]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [4]Department of Earth and Planetary Science, University of California, Berkeley, CA, USA. *e-mail: savage@berkeley.edu

produce an HCR phenotype. This screen identified all of the known CCM components and confirmed that a two-gene operon containing a large conserved poorly characterized protein (Pfam: PF10070, hereafter, DabA) and a cation transporter (Pfam: PF00361, hereafter DabB) is required for the function of CCMs. Scott et al. recently identified and validated homologues of these genes as a C_i-import system in hydrothermal vent chemolithoautotrophs[28–30]; on the basis of this work and results further described below, we propose naming this locus the *DAB* operon, for DABs accumulate bicarbonate.

We show that the products of the *DAB* operon form a protein complex that is capable of energetically coupled uptake of C_i. Both proteins are necessary for activity, and treatment with an ionophore abrogates DAB-mediated uptake of C_i. Structural homology modelling suggests that DabA contains a domain that is distantly homologous to a type II β-CA. Indeed, DabA binds to zinc, probably in a manner similar to β-CAs. These results are consistent with a model in which activity is dependent on unidirectional hydration of CO_2 to HCO_3^- in the cytosol through a CA-like mechanism and energized by coupling to a cation gradient. 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. cholerae* and *B. anthracis*. We therefore propose that DABs constitute a class of C_i-uptake pump, the biochemical tractability of which facilitates mechanistic analyses and the widespread occurrence of which merits further investigation.

Results

Transposon mutagenesis and gene essentiality. We constructed a randomly barcoded genome-wide pooled knockout library of *H. neapolitanus* by conjugation (Fig. 1a). The conjugated vector contained a barcoded Tn5-derived transposon encoding a kanamycin resistance marker. The library was produced in 5% CO_2, enabling isolation of CCM gene knockouts.

Transposon barcodes simplify the use of the library for pooled screens by allowing the abundances of individual transposon mutants to be tracked through barcode sequencing (BarSeq)[31] (see Methods). Transposon insertion sites and cognate barcodes were mapped using standard TnSeq methods[31] (see Methods). The library was found to contain ~10^5 insertions, or one insertion for every ~25 base pairs in the *H. neapolitanus* genome. As the average gene contains ~35 insertions, genes with no insertions are highly probable to be essential for growth. A simple statistical model identified 551 essential genes and 1,787 non-essential genes out of 2,408 genes in the *H. neapolitanus* genome (see Methods; Fig. 1a,b, Supplementary File 2). The remaining 70 genes were classified as ambiguous because of their short length or because replicate mapping experiments were discordant (see Methods). Genes associated with known crucial functions—including central carbon metabolism, ribosome production and DNA replication—were found to be essential (Fig. 1c, Supplementary Fig. 1). Importantly, known CCM genes, including carboxysome components, were not essential for growth at 5% CO_2 (Fig. 2).

Comprehensive screen for *H. neapolitanus* CCM components. On the basis of the current model of the bacterial CCM (Fig. 2a), knockouts of CCM genes are expected to have reduced fitness in atmospheric CO_2 conditions[18,19]. As our pooled library contains around 70,000 barcodes that map to exactly one position in the *H. neapolitanus* genome, we were able to use BarSeq to quantify the fitness effects of single-gene knockouts for all non-essential *H. neapolitanus* genes in a pooled competition experiment[31] (see Methods; Fig. 2b). As the library contains roughly 20 uniquely mapped knockouts per gene, this screen contains multiple internal biological replicates, testing the effect of gene knockouts. Mutants

in a particular gene were designated as HCR if the average effect of a knockout in that gene was a twofold (or greater) growth defect in ambient CO_2 compared with their growth at 5% CO_2 in two replicate experiments.

As expected, knockouts of carboxysome genes consistently produced large and specific fitness defects[27] in ambient CO_2 (Fig. 2b,c). These genes include *cbbLS*, the large and small subunits of the α-carboxysomal Rubisco[9]; *csoS2*, an intrinsically disordered protein required for α-carboxysome assembly[32]; *csoSCA*, the carboxysomal CA[9]; *csoS4AB*, the pentameric proteins thought to form vertices of the α-carboxysome[33]; and *csoS1CAB*, the hexamers that form the faces of the α-carboxysome shell[9,25]. Knockouts of *csoS1D*—a shell hexamer with a large central pore[20,34]—showed a phenotype that was too weak to be considered as HCR (Fig. 2b,c). The *H. neapolitanus* genome also contains a non-carboxysomal form II Rubisco that is probably not involved in CCM activity because its disruption does not confer a fitness defect. A number of genes that are not associated with the carboxysome structure also exhibited HCR phenotypes. These genes include two LysR transcriptional regulators, a Crp/Fnr type transcriptional regulator, a protein named acRAF that is involved in the assembly of Rubisco[35,36] and two paralogous loci (hereafter, *DAB1* and *DAB2*; Fig. 2b–f) that encode *dab*A and *dab*B.

Composition of the *DAB* operons. The *DAB1* locus consists of a cluster of 2 genes found in an operon directly downstream of the carboxysome operon (Fig. 2c). Although *DAB1* is part of an 11-gene operon that contains several genes associated with Rubisco proteostasis—including acRAF[35,36] and a cbbOQ-type Rubisco activase[37]—we refer to *DAB1* as an operon for simplicity. *DAB2* is a true operon and is not proximal to the carboxysome operon in the *H. neapolitanus* genome. These operons are unified in that they both display HCR phenotypes and possess similar genes (Fig. 2b–d).

The *DAB1* and *DAB2* operons contain a gene that encodes a conserved protein of unknown function (Pfam: PF10070), which we term DabA. DabAs have no predicted transmembrane helices or signal peptides and appear to be large (DabA1, 118.5 kDa; DabA2, 91.7 kDa), soluble, cytoplasmic proteins (see Methods; Fig. 3a). The *DAB1* and *DAB2* operons also contain a member of the cation transporter family (Pfam: PF00361) that includes H^+-pumping subunits of respiratory complex I and Mrp Na^+:H^+ antiporters[38]. This protein—which we call DabB (DabB1, 62.2 kDa; DabB2, 59.3 kDa)—is predicted to have 12–13 transmembrane helices (Fig. 3a). The complex I subunits in PF00361 are H^+-pumping proteins and do not contain redox active groups, such as iron-sulfur clusters or quinone-binding sites. Phylogenetic analysis suggests that DabB proteins form a clade among PF00361 members (Supplementary Fig. 4a) that are distinct from complex I subunits. Therefore, homology between DabB and complex I subunits (for example, NuoL) suggests cation transport but does not imply redox activity. Importantly, operons of this type were recently demonstrated to be capable of C_i uptake in the hydrothermal vent chemolithoautotroph *Hydrogenovibrio crunogenus*[28–30].

***dabA2* and *dabB2* are necessary and sufficient for energy-coupled accumulation of C_i in *Escherichia coli*.** To facilitate analysis of C_i-transport activity, we generated an *E. coli* strain—CAfree—that has knockouts of both of the genes that encode CAs (see Methods). Previous studies showed that deletion of the constitutive CA gene, *can*, produces an HCR phenotype in *E. coli*[39] that is complemented by the expression of cyanobacterial bicarbonate transporters[40]. The deletion of both CA genes replicates this phenotype and greatly reduces the likelihood of escape mutants. As disruption of *DAB2* is associated with a greater fitness defect than disruption of *DAB1* (Fig. 2b), we used CAfree to test DAB2 for C_i-uptake activity. Expression of both *dabA2 and dabB2* enabled the growth of CAfree in ambient CO_2, whereas expression of *dabA2* or *dabB2*
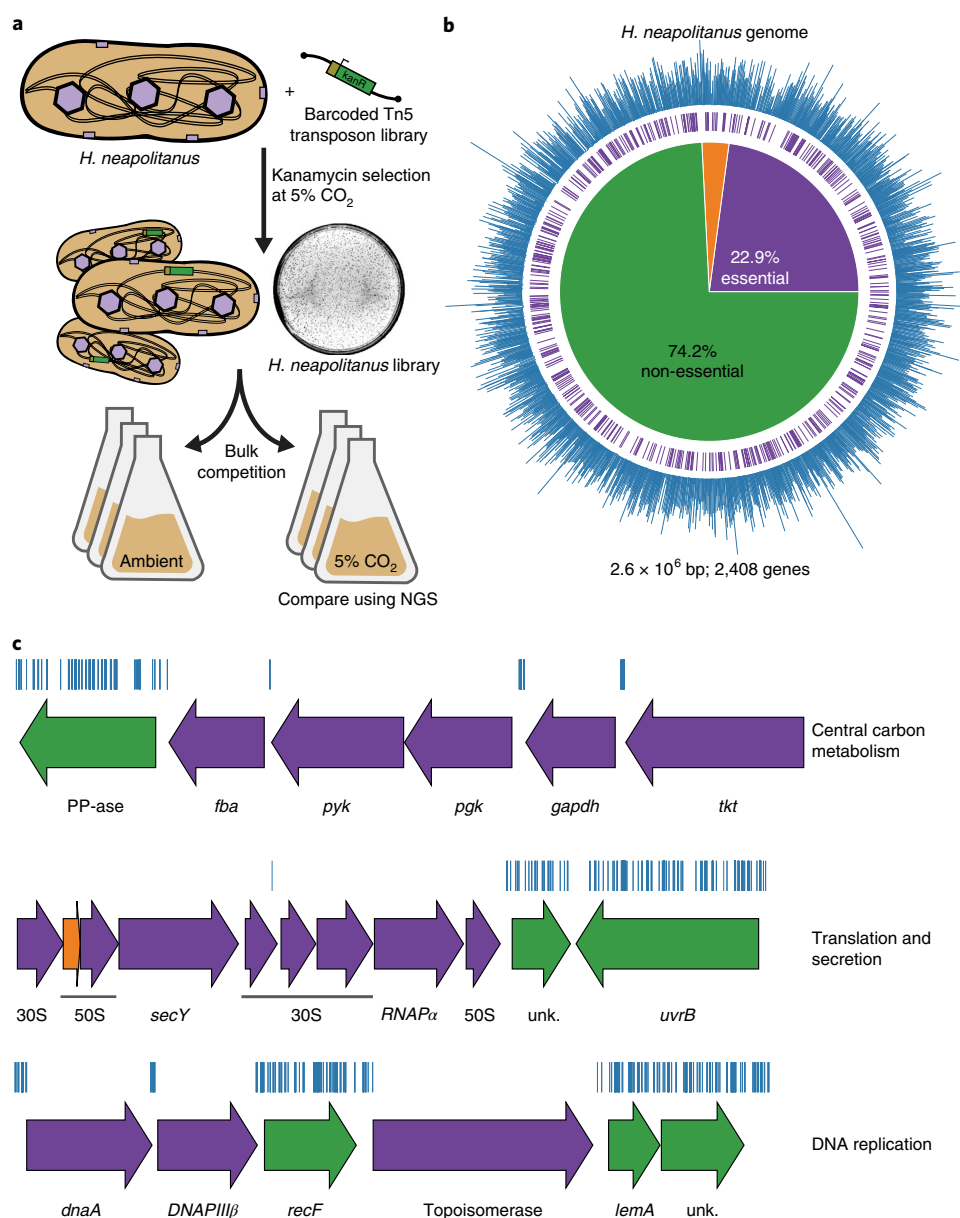
Fig. 1 | Transposon mutagenesis reveals the essential gene set of a chemolithoautotrophic organism. a, Schematic of the generation and screening of the RB-TnSeq library. Transposons were inserted into the *H. neapolitanus* genome by conjugation with an *E. coli* donor strain. The transposon contains a random 20-base-pair barcode (yellow) and a kanamycin selection marker (kan[R]; green). Selection for colonies containing insertions was performed in the presence of kanamycin at 5% CO_2, and insertions were mapped by sequencing as described in the Methods. Subsequent screens were carried out as bulk competition assays and quantified using BarSeq. **b**, Insertions and essential genes are well distributed throughout the *H. neapolitanus* 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 that were found to be essential (purple), ambiguous (orange) and non-essential (green). **c**, Representative essential genes and non-essential genes in the *H. neapolitanus* genome. The blue track indicates the presence of an insertion. Genes indicated in purple were found to be essential and genes indicated in green were found to be non-essential. Genes labelled as unk. encode hypothetical proteins. The first genomic locus contains five essential genes that are involved in glycolysis or the CBB cycle, including pyruvate kinase (*pyk*) and transketolase (*tkt*). The first locus also contains the non-essential gene exopolyphosphatase (PP-ase). The eight essential genes in the second locus encode 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 β. A full analysis with gene names is in Supplementary Fig. 1 and essentiality information for every gene can be found in Supplementary File 2.

alone was not sufficient to enable the growth of CAfree (Fig. 3b, Supplementary Fig. 5). $^{14}C_i$-uptake assays demonstrated that DAB2 facilitates the import of extracellular C_i to levels above that of the appropriate control (Fig. 3c). Moreover, DAB2-associated uptake of C_i is strongly inhibited by the ionophore Carbonyl cyanide m-chlorophenyl hydrazone (CCCP) (Fig. 3c), indicating that DAB2 is energetically coupled—either directly or indirectly—to a cation gradient (for example, H^+ or Na^+). This is consistent with the previous observations that the uptake of C_i in *H. neapolitanus* is powered by a membrane gradient[17].
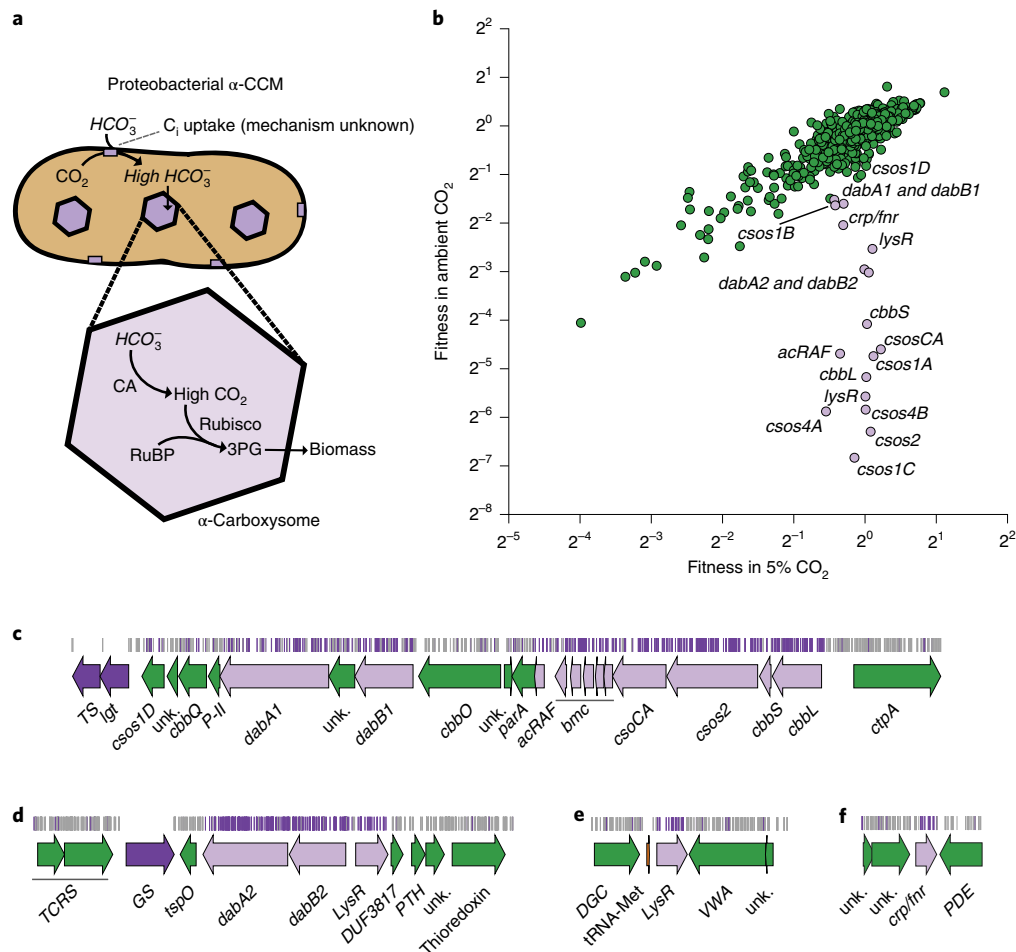
Fig. 2 | A systematic screen for HCR mutants identifies genes putatively associated with the CCM. a, Simplified model of the α-CCM of chemolithoautotrophic proteobacteria. C_i is concentrated by an unknown mechanism, generating a high cytosolic concentration of HCO_3^-. This cytosolic HCO_3^- is converted into a high concentration of carboxysomal CO_2 by CA, which is localized exclusively in the carboxysome. **b**, Fitness effects of gene knockouts in 5% CO_2 compared with their fitness at ambient CO_2. Data are from one of two replicates of BarSeq. The effects of single transposon insertions into a gene were averaged to produce the gene-level fitness values that are plotted. We define HCR mutants as those that display a twofold fitness defect in ambient levels of CO_2 relative to their fitness in 5% CO_2 in both replicates. HCR genes are coloured light purple. Data from both replicates and the associated s.e. are shown in Supplementary Fig. 2 and Supplementary File 3. **c–f**, Regions of the *H. neapolitanus* genome containing genes annotated as HCR in **b**. Essential genes are indicated by dark purple, HCR genes are indicated by light purple and other genes are indicated by green. The top tracks show the presence of an insertion in that location. Insertions are coloured grey unless they display a twofold or greater fitness defect in ambient CO_2, in which case they are coloured 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 *dabA1* and *dabB1*. **d**, The *DAB2* operon and surrounding genomic context. **e**, The genomic context of a lysR-type transcriptional regulator that shows an HCR phenotype. **f**, The genomic context of a Crp/Fnr-type transcriptional regulator that displays an HCR phenotype. Full gene names are provided in Supplementary Fig. 3. Accession numbers and GenInfo Identifiers for selected genes are provided in Supplementary Table 1.

DabA2 and DabB2 interact to form a complex. To determine whether the genetic interaction between *dabA2* and *dabB2* reflects a physical interaction, we attempted to co-purify the two proteins. We fused DabA2 to a C-terminal Strep-tag, and fused DabB2 to a C-terminal green fluorescent protein (GFP) with a 6×His-tag, and the genes were co-expressed in *E. coli* (see Methods). Tandem-affinity purification after detergent solubilization revealed that DabA2 and DabB2 form a complex in *E. coli* (Fig. 4a). The complex appeared as a single major peak when analysed using size-exclusion chromatography and has a retention volume that is consistent with a heterodimer of DabA2 and DabB2 (Fig. 4b). We did not observe co-purification of any *E. coli* proteins, suggesting that DAB2 operates as an independent complex within the membrane (Fig. 4a). Expression of *DAB2* also rescued CAfree growth even when complex I was knocked out (Δ(*nuoA-nuoN*); Supplementary Fig. 6),

providing further evidence that the function of DAB2 is independent of complex I.

pH independence of DAB2 rescue suggests that CO_2 is probably the true substrate. Aqueous CO_2 spontaneously interconverts with the gas phase as well as hydrated C_i species (H_2CO_3, HCO_3^- and CO_3^{2-}). The equilibrium of $CO_{2(aq)}$ and $CO_{2(gas)}$ is not affected by pH, but the conversion from CO_2 to hydrated C_i is pH dependent. Thus, the equilibrium concentration of HCO_3^- increases 100-fold between pH 5 and pH 7 without an accompanying change in CO_2 concentration[7] (Supplementary Fig. 7a). Expression of SbtA—a known HCO_3^- transporter—rescues the growth of CAfree at pH 7 but not at pH 5, whereas the expression of DabA2 and DabB2 rescues growth at both pH 5 and 7 (Supplementary Fig. 8). As the rescue by expression of DabA2 and DabB2 is pH-independent in
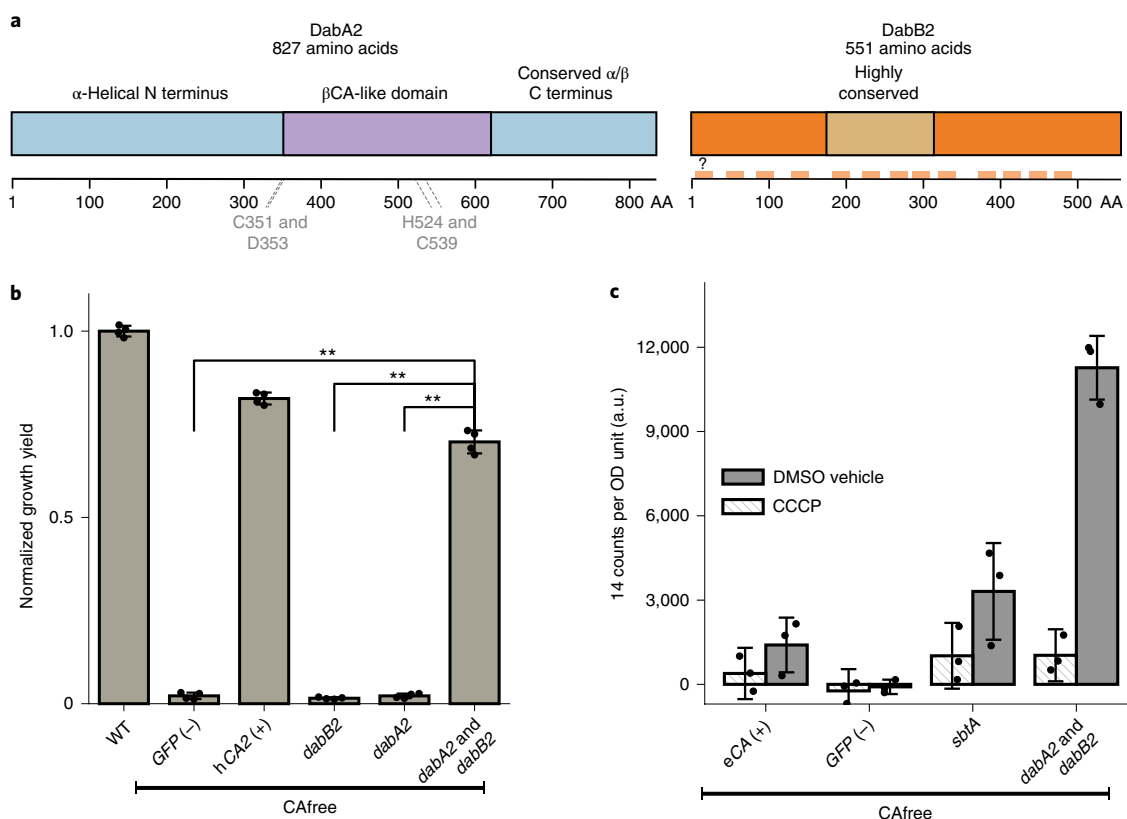
Fig. 3 | DABs catalyse active transport of C_i and are energized by a cation gradient. a, Schematic of DabA2 and DabB2 on the basis of bioinformatic annotation. The four predicted active site residues (C351, D353, H524 and C539) are marked on the primary amino acid sequence. Amino acid numbers are marked below each gene and predicted transmembrane helices are marked in light orange. AA, amino acids. **b**, The ability of DAB2 to rescue growth of CAfree *E. coli* in ambient CO_2 conditions. Expression of the full *DAB2* operon (*dabA2* and *dabB2*) rescued growth, as did the positive control human carbonic anhydrase 2 (h*CA2*). Expression of *dabA2* and *dabB2* showed a larger rescue than *GFP* ($t = 42.6$, corrected $P = 3.4 \times 10^{-8}$), *dabA2* alone ($t = 43.4$, corrected $P = 3 \times 10^{-8}$) and *dabB2* alone ($t = 44.5$, corrected $P = 2.6 \times 10^{-8}$). WT, wild type. **Bonferroni-corrected $P < 5 \times 10^{-4}$ according to a two-tailed *t*-test. Data are mean ± s.d. of four biological replicates. Consistent results were seen in two independent experiments. **c**, CAfree *E. coli* were tested for the uptake of C_i using the silicone-oil centrifugation method. Expression of *dabA2* and *dabB2* generated a large increase in ^{14}C uptake compared with all of the controls. Moreover, treatment with the ionophore CCCP greatly reduced the uptake of ^{14}C mediated by DabA2 and DabB2, suggesting that DabA2 and DabB2 are coupled to a cation gradient. *E. coli* CA (e*CA*) was used as a control for a non-vectorial CA. *S. elongatus* strain PCC 7942 *sbtA* was used as a known C_i transporter. GFP was used as a vector control. DMSO, dimethyl sulfoxide. Data are mean ± s.d. of three technical replicates. Consistent results were seen with three independent experiments.

this range, its substrate is probably CO_2 and not H_2CO_3, HCO_3^- or CO_3^{2-}. This is consistent with previous observations that CO_2 is the probable substrate of C_i uptake in *H. neapolitanus*[17].

Requirement of putative Zn-binding residues for the function of DABs. Structural-homology modelling software predicted that the middle of DabA2 contains sequence elements that are related to a β-CA (Fig. 3a). Predictions using Phyre2 identified C539 and H524 as part of a potential Zn^{2+}-binding site that is distantly homologous to a bacterial type II β-CA (10% coverage of DabA, 90.8% confidence). I-TASSER predicted a Zn^{2+} binding site that includes the same residues, C539 and H524, along with an additional cysteine (C351) and aspartic acid (D353). As shown in Fig. 4c, these residues could make up the active site of a type II β-CA[41]. We generated individual alanine mutants for each of these putative active site residues (C351A, D353A, H524A and C539A) and tested them in CAfree. All of the mutants failed to rescue the growth of CAfree in ambient CO_2 (Fig. 4d). We proceeded to analyse the ability of purified DabA2–DabB2 complex to bind to zinc using X-ray fluorescence spectroscopy and found that wild-type DabA2–DabB2 and three of the single mutants (C351A, D353A and H524A) bind to zinc (Fig. 4e). Single mutants retain three of the four zinc-coordinating residues[41], which could

explain why the mutants bind to zinc. Indeed, mutational studies of the human CA II show that single mutations to Zn^{2+}-binding residues reduce—but do not abrogate—zinc binding[42].

Purified DabA2–DabB2 complex does not have conspicuous CA activity. The assay of detergent-solubilized purified DabA2–DabB2 did not show CA activity compared with the controls (Fig. 4f). DabA2–DabB2 was assayed at high protein concentrations (>650-fold more protein than the positive control) and under CO_2 concentrations that are typically saturating for CAs but did not display any CA activity (Fig. 4f). The absence of activity in vitro indicates that either DabA2–DabB2 has extremely low CA activity or, perhaps, that DabA2–DabB2 must be localized in a cell membrane that maintains a cation gradient to function as an energetically activated CA.

DAB operons are widespread in prokaryotes and present in human pathogens. A query of the UniProt database with the DabA Pfam (PF10070) yielded 878 putative *dabA* sequences. We found *dabA* sequences in a wide variety of prokaryotes—including bacteria and archaea (Fig. 5a, Supplementary Fig. 9)—as is consistent with previous work[28]. Represented clades include not only

Proteobacteria but also Euryarchaeota, Firmicutes, Planctomycetes and Bacterioides. However, we were surprised to observe that many *dabA* sequences were found in the genomes of organisms that cannot fix CO_2, including the heterotrophic human pathogens *V. cholerae*, *B. anthracis* and *Legionella pneumophila* (Fig. 5a). Notably, 843 (96%) of the identified *dabA* sequences were either within three genes of, or fused to, a *dabB*.

Finally, we assayed whether the DAB homologues from heterotrophic pathogens are functional C_i pumps. Operons from *V. cholerae* E7946 El Tor Ogawa and *B. anthracis* Sterne were cloned and expressed in CAfree. Both *DAB* operons rescued CAfree growth in ambient CO_2 (Fig. 5b, Supplementary Fig. 10). Thus, *DAB* operons from heterotrophic human pathogens are functional.

Discussion

We generated a knockout library containing ~35 individual knockouts for every gene in the genome of the proteobacterial chemolithoautotroph *H. neapolitanus*. Using these data, we compiled the essential gene set of a chemolithoautotroph (Fig. 1) and were able to confidently identify 551 essential genes and 1,787 non-essential genes. Mapping essential genes will provide insight into the metabolism and growth physiology of sulfur-oxidizing chemolithoautotrophs.

In addition to mapping essential genes, this library was used to measure conditional phenotypes of non-essential genes. These mutants were isolated in high concentrations of CO_2, enabling the disruption of all of the known components of the bacterial CCM (Fig. 2). The resulting genome-wide knockout library was used to perform a comprehensive screen for bacterial CCM genes. This screen highlighted a small number of genes (17) as having the HCR phenotype associated with the CCM (Fig. 2b–f), nearly all of these genes are known to be associated with the α-carboxysome. Although it is possible that genetic redundancy, conditional phenotypes or impairment at only sub-ambient CO_2 permits some genes to escape notice, these data suggest that the proteobacterial CCM is composed of less than 30 functionally distinct components. Moreover, none of the genes identified have unexpected functions, suggesting that current models of bacterial CCMs incorporate all of the necessary functions.

Our screen identified 3 transcriptional regulators as well as 3 distinct CCM operons (Fig. 2b–f). Identification of transcriptional regulators with HCR phenotypes (Fig. 2d–f) may inform the study of CCM regulation. The first operon contains nearly all of the known components of the α-carboxysome, all of which confer HCR phenotypes on disruption (Fig. 2c). The second operon is adjacent to the carboxysome operon and contains 11 genes. Only 3 of these genes— the Rubisco chaperone a*cRAF*, *dabA1* and *dabB1*—displayed HCR

phenotypes (Fig. 2c). The remaining 8 genes—including *cbbOQ*, *csos1D*, *p-II* and a *parA* homologue—had no associated phenotype but might nonetheless have roles in the CCM (Fig. 2c). The third operon contains 2 genes, *dabA2* and *dabB2*, both with HCR phenotypes (Fig. 2d).

A previous physiological study suggested that the uptake of C_i in *H. neapolitanus* is coupled to the electrochemical potential of the membrane and uses CO_2 as a substrate, but the protein(s) responsible for this activity were unknown[17]. DAB1 and DAB2 are homologous to C_i pumps from hydrothermal vent chemolithoautotrophs that were recently discovered by Scott et al.[28,30], and our screen suggests that DAB1 and DAB2 are probably the C_i pumps in *H. neapolitanus*. These observations raise many mechanistic questions as to how DABs function; we therefore sought to establish a biochemical system to investigate DAB structure and function.

We showed that the *DAB2* operon encodes a two-component protein complex that has C_i-uptake activity when heterologously expressed in *E. coli* (Figs. 3b,c and 4a). This complex is probably a heterodimer, as suggested by size-exclusion chromatography (Fig. 4b). As the uptake of C_i is inhibited by the ionophore CCCP (Fig. 3c), we suspect that DAB2 activity is energetically coupled to a cation gradient (Fig. 5a). Moreover, as DabA2–DabB2 shows pH-independent rescue of CAfree *E. coli*, CO_2 is probably the transported substrate (Fig. 4c). This is further supported by the fact that DabA has distant homology to a type II β-CA and binds to zinc (Figs. 3 and 4), suggesting that a CA active site hydrates transported CO_2. Finally, mutations to the putative zinc-binding residues in DabA2–DabB2 (C351A, D353A, H524A and C539A) ablate function in vivo (Fig. 4d). We therefore propose a model of DAB activity in which CO_2 is passively taken into the cell and then unidirectionally hydrated to HCO_3^- by energy-coupled CA activity of DabA (Fig. 6).

Most well-studied CAs are not coupled to any energy source (for example, ATP or cation gradient). Rather, they equilibrate CO_2 and HCO_3^- (ref. [42]). However, energy-coupled CA activity could favour the hydration of CO_2, enabling the DAB system to actively accumulate HCO_3^- in the cytosol and power the CCM (Fig. 2a). Given the similarity of DabB to H^+-pumping proteins, we propose that DABs use the H^+ gradient, although our results are equally consistent with other cation gradients, such as Na^+. This mechanism would require tight coupling of cation flow to CA activity by DabA, consistent with our observation that purified DabA2–DabB2 displays no measurable CA activity. Interestingly, type II β-CAs are the only CAs that are known to display allosteric regulation[43]. Allosteric control is thought to be mediated by Zn^{2+} capping and uncapping by the active-site aspartic acid residue (D353 in DabA2)[43]. A similar mechanism might couple cation flow through DabB to the active site of DabA.

Fig. 4 | DabA contains a β-CA-like active site but is not active outside of the membrane. a, Purification of DabA2–DabB2 complex from *E. coli*. DabA2 was C-terminally tagged with a Strep-tag and DabB2 was C-terminally tagged with super folder (sf)-GFP and a 6xHis-tag. Purification was monitored using SDS–PAGE and imaged using fluorescence (right) before Coomassie (left) staining. Lane 1, clarified lysate; lane 2, solubilized membranes; lane 3, Ni-NTA resin eluent; lane 4, Strep-Tactin resin eluent. DabA2 and DabB2 co-purify as a single complex without any obvious interactors. Similar results were observed after three independent purifications. **b**, Size-exclusion chromatogram of His/Strep-purified DabA2–DabB2, indicating retention volumes (orange arrows) and molecular weights (kDa) for standard samples (apoferritin, 443 kDa; β-amylase, 224 kDa). DabA2–DabB2 runs with an effective mass of around 270 kDa, which probably reflects an oligomer of DabA2 and DabB2. Similar results were observed after three independent purifications. mAU, milli-absorbance units. **c**, Structural model of the DabA2 active site based on a β-CA from *E. coli* (PDB 1I6P). Typical β-CAs rely on two cysteine residues and one histidine residue to bind to Zn^{2+}. The aspartic acid residue coordinates Zn^{2+} but is probably displaced during catalysis[43]. **d**, Alanine mutants of the putative DabA2 active site residues abrogate the rescue of CAfree *E. coli* compared with wild-type DAB2 (C351A, $t=54.3$, $P=1.1\times10^{-8}$; D353A, $t=144$, $P=3.1\times10^{-11}$; H524A, $t=44$, $P=3.7\times10^{-8}$; C539A, $t=44.3$, $P=3.5\times10^{-8}$; all P values listed here are Bonferroni corrected). **Bonferroni-corrected $P<5\times10^{-4}$ according to a two-tailed Student's t-test. Data are mean ± s.d. of four biological replicate cultures. **e**, X-ray fluorescence data indicate that DabA2–DabB2 binds to zinc like all known β-CAs. Single mutations in the active site do not abrogate the ability of DabA2–DabB2 to bind to zinc. Curves are from representative samples. Technical replicate traces were concordant (WT, 9; D353A, 5; H524A, 4; C351A, 3; Rubisco, 2; buffer blank (Tris-buffered saline containing 0.03% DDM), 4, bCA, 3). Replicate traces for DAB2 H524A include samples from two independent purifications. bCA, bovine carbonic anhydrase. buffer. **f**, Purified DabA2–DabB2 does not display any obvious CA activity despite being present in 650-fold excess over the positive control (hCA2) in our assays. Data are mean ± s.e.m. of seven experimental traces. Similar results were observed in two independent purifications.

Cyanobacteria possess vectorial CAs called CUPs that may provide clues to the mechanism of DAB[24,44,45]. 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[+]-pumping subunits of complex I. This commonality might suggest similar mechanisms. The hydration of CO_2 by CUPs is thought to be coupled

to energetically favourable electron flow because CUPs associate with complex I[46] (Supplementary Fig. 9b). However, the Mrp protein family (PF00361) is highly diverse and contains many cation transporters that do not associate with complex I or any other redox-coupled complex[38]. Furthermore, DabB sequences are only distantly related to complex I and CUP subunits (Supplementary
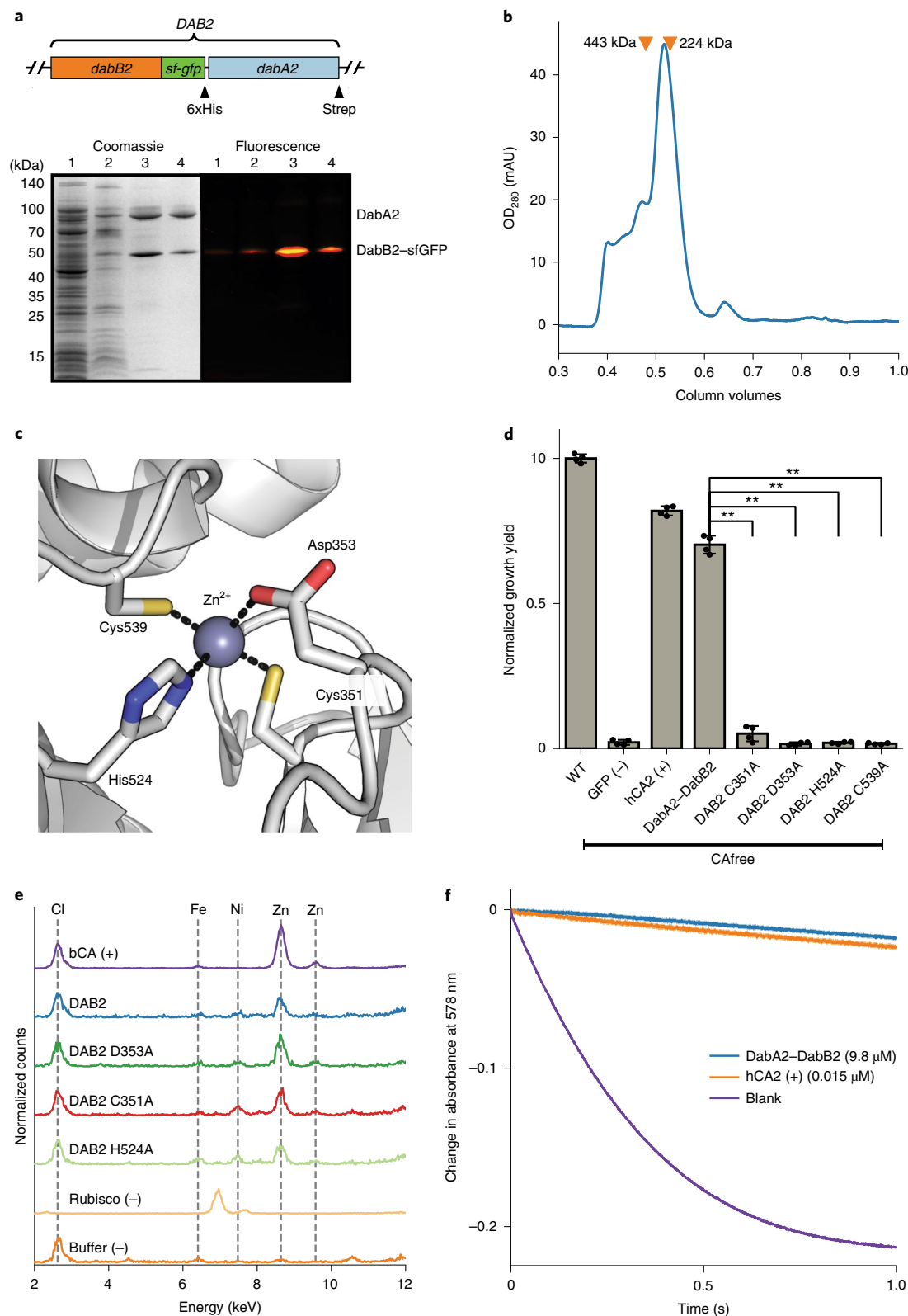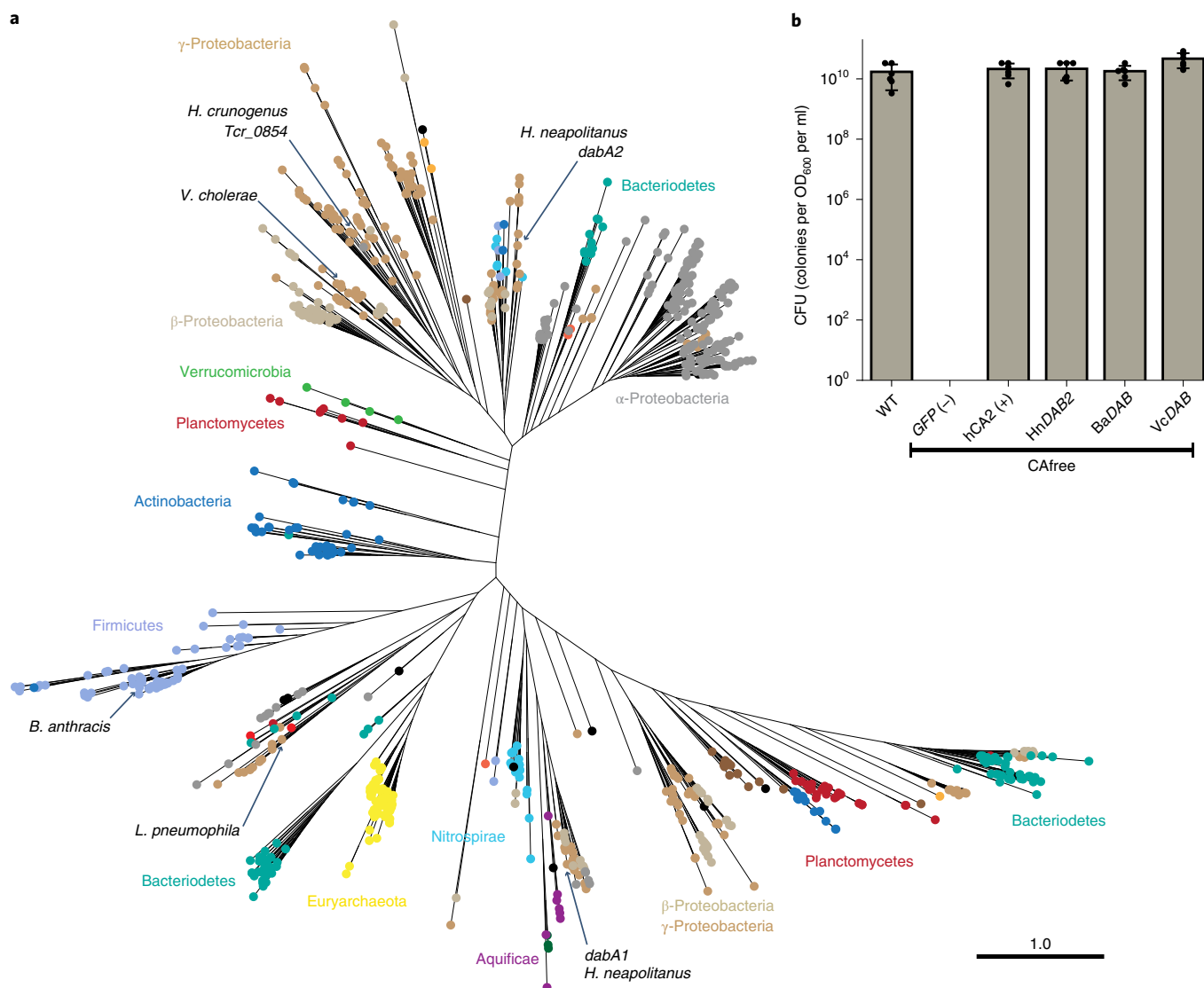
Fig. 5 | DAB operons are widespread among prokaryotes. a, Approximate maximum likelihood phylogenetic tree of 878 *dabA* homologues associated with PF10070.9 (see Methods). Homologues of *dabA* are found in more than 15 prokaryotic clades, including archaea. *H. neapolitanus dabA1* and *dabA2* represent two different groupings that are commonly found in proteobacteria. Inspecting the tree reveals several probable incidents of horizontal transfer, for example, between Proteobacteria and Firmicutes, and between Nitrospirae and Actinobacteria. Moreover, the genomes of several known pathogens contain a high-confidence *dabA* homologue, including *B. anthracis*, *V. cholerae* and *L. pneumophila*. Detailed annotations are provided in Supplementary Fig. 9. Scale bar, one substitution per site. Sequences used to generate the tree can be found in Supplementary File 5. **b**, Functional *DABs* are found in human pathogens. Colony-forming units (CFU) per optical density at 600 nm (OD_{600}) per ml were measured on lysogeny broth (LB) plates with induction in air. *DAB* operons from *B. anthracis* (Ba*DAB*) and *V. cholerae* (Vc*DAB*) rescued growth of CAfree cells. Hn*DAB2*, *H. neapolitanus DAB2* operon. Data are mean ± s.d. of six technical replicate platings. Consistent results were achieved in biologically independent platings of Ba*DAB* and Vc*DAB*.

Fig. 4a), DAB2 subunits do not co-purify with *E. coli* complex I (Fig. 4a) and DAB2 rescues CAfree growth in a complex I knock-out (Supplementary Fig. 6). We therefore propose that DAB activity is not coupled to electron flow through complex I but, rather, that DAB activity is coupled to a cation gradient across the membrane (Fig. 6).

DabA2–DabB2 functions robustly, as demonstrated by complementation of CAfree (Fig. 3b) and ¹⁴C uptake measurements (Fig. 3c). Indeed, we observed that DabA2–DabB2 functions substantially better than SbtA—a C_i transporter from cyanobacteria[9,40]—in *E. coli* (Fig. 3c). As *E. coli* and *H. neapolitanus* are proteobacteria, this observation could result from greater compatibility of proteobacterial proteins with *E. coli* expression. However, it may also be the case 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 chemolithoautotrophs throughout α-, β-, and γ-proteobacteria and the α-CCM was clearly horizontally transferred from proteobacteria to marine cyanobacteria[9]. DabA homologues are widespread among prokaryotes and were probably horizontally transferred multiple times (Fig. 5a). As 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 microorganisms, in which increased intracellular C_i could be useful[47].

Finally, DABs are present in a wide variety of bacteria and archaea[28]. High-confidence DabA homologues are found not only in large numbers of autotrophs but also in heterotrophs (Fig. 5a, Supplementary Fig. 9). Moreover, homologues are present in
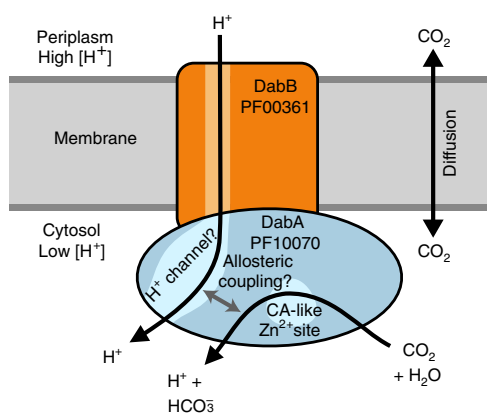
Fig. 6 | A hypothetical model of the unidirectional energy-coupled CA activity of DAB complexes. We propose that DabA–DabB complexes couple CA activity of DabA to a cation gradient across the cell membrane, generating HCO_3^- by the unidirectional hydration of CO_2. The cation gradient could be H^+ or Na^+. Energy-coupled CA activity is required for the DABs role as a C_i-uptake system in the proteobacterial CCM, as discussed above. As it seems that DabA2–DabB2 is not active as a purified complex outside of the membrane, it is assumed that the protein tightly couples the inflow of cations with CO_2 hydration so that there is no slippage—that is, uncoupled CA activity. Indeed, slippage would be counterproductive for the function of CCMs[7,14]. Notably, Zn^{2+} binding by the active-site aspartic acid residue of type II β-CAs (D353 in DabA2, Fig. 4a) is thought to allosterically regulate activity[43]. This Asp-mediated activity switch could therefore provide a means for allosteric coupling of a β-CA active site to distal ion transport.

the notable heterotrophic pathogens *V. cholerae*, *B. anthracis* and *L. pneumophila* (Fig. 5a). We showed that DABs from *V. cholerae* and *B. anthracis* are active in *E. coli* (Fig. 5b). This leads us to wonder: what do heterotrophic pathogens use C_i-uptake systems for? The activity of CA is essential for the growth of the heterotrophs *E. coli* and *Saccharomyces cerevisiae* in ambient CO_2 (refs. [39,48]). In the heterotrophic context, CA activity is thought to supply bicarbonate for biotin-dependent carboxylases in central metabolism, for which HCO_3^- is the substrate[39,48]. Bicarbonate levels have also been linked to virulence in both *V. cholerae* and *B. anthracis*[49,50]. Perhaps DAB-family C_i-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_i uptake in heterotrophic and pathogenic organisms.

Methods

Important strains and reagents. A detailed listing of key strains and reagents is provided in Supplementary File 1.

Bacterial strains and growth conditions. *E. coli* strain APA766 was used as the conjugation donor to transfer the Tn5 transposon to *H. neapolitanus* C2 by conjugation[31]. The *E. coli* double CA deletion strain CAfree (BW25113 Δ*can* Δ*cynT*) was generated by curing the KEIO collection *cynT* knockout (BW25113 Δ*cynT*, KEIO strain JW0330) of kanamycin resistance through pCP20-mediated FLP recombination and subsequent P1 transduction (and curing) of kanamycin resistance from the *can* knockout strain EDCM636 (MG1655 Δ*can*, Yale Coli Genomic Stock Center[39,51]). Then, complex I knockout strains (Δ(*nuoA-nuoN*)) were generated in the BW25113 and CAfree backgrounds. These strains were generated by lambda red mediated recombination of a Kan^R resistance cassette flanked by FRT sites into the *nuo* locus such that the entire operon was removed. The pSIM5 plasmid carrying the lambda red recombinase was heat cured at 42 °C. Lysogeny broth (LB) and LB agar were used as growth media for *E. coli* unless otherwise specified. *E. coli* strains were grown at 37 °C in the presence of 0.1 mg ml^{-1} carbenicillin, 0.06 mg ml^{-1} kanamycin or 0.025 mg ml^{-1} chloramphenicol as appropriate. *H. neapolitanus* was grown in DSMZ-68 medium at 30 °C and in the presence of 0.03 mg ml^{-1} kanamycin as appropriate.

Transposon mutagenesis and RB-TnSeq library production. A barcoded library of *H. neapolitanus* transposon mutants was generated by adapting the methods of Wetmore et al.[31] Conjugations were performed as follows. *H. neapolitanus* and APA766 were cultured and collected by centrifugation. Both cultures were washed once in 10 ml antibiotic-free growth medium per conjugation reaction and then resuspended in 100 µl antibiotic-free growth medium. Then, 5 OD$_{600}$ units of *H. neapolitanus* were mixed with 20 OD$_{600}$ units of APA766 on 3 0.45 µm Millipore MCE membrane filters and cultured overnight at 30 °C in 5% CO_2 on antibiotic-free LB agar plates containing 0.06 mg ml^{-1} diaminopimelic acid. Cells were scraped from the filters into 2 ml DSMZ-68 per conjugation (6 ml total) and collected in 2 ml microcentrifuge tubes. Recovered cells were pelleted by centrifugation at 16,000*g* for 1 min, washed in an equal volume of DSMZ-68, pelleted again at 9,000*g* for 1 min and resuspended in an equal volume of DSMZ-68 again before being plated onto ten separate DSMZ-68 kanamycin plates per conjugation (30 plates total). Plates were incubated at 30 °C under 5% CO_2 until colonies formed (around 7 d). Colonies were counted and scraped into 55 ml DSMZ-68. Then, two samples of 1.4 OD$_{600}$ units were taken and used to prepare genomic DNA (Qiagen DNeasy Blood and Tissue kit). Transposon insertions were amplified from genomic DNA, and transposons were mapped after Illumina sequencing using protocols and software according to Wetmore et al.[31] Aliquots of cells containing 1.6 OD$_{600}$ units were then flash-frozen in 50% glycerol for subsequent BarSeq experiments.

Essential gene assignment. Following the logic of Wetmore et al.[31] and Rubin et al.[52], we categorized genes as essential if we observed significantly fewer transposon insertions than would be expected by chance. If insertions occurred uniformly at random, the number of insertions per gene would be expected to follow a binomial distribution. The probability of observing a maximum of k insertions into a gene of length n is therefore expressed as:

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

where 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, that is, if the probability of observing k or fewer insertions was beneath 0.05 after Bonferroni correction. Genes were called ambiguously essential in two cases: (1) replicates were discordant or (2) zero insertions were observed but the gene was short enough that the formula could not yield a Bonferroni-corrected P-value 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 described by Wetmore et al.[31] This method enables pooled-library fitness experiments to be performed that compare different growth conditions by comparing barcode abundance changes to track changes in the abundance of the transposon mutants. In brief, a library aliquot was thawed and used to inoculate three 33 ml cultures. Cultures were grown to OD$_{600}$ of approximately 0.08 in 5% CO_2. Then, 20 ml was removed and collected by centrifugation as two t_0 (input) samples. Cultures were back-diluted 1:64 into 128 ml and incubated for 6.5–7.5 doublings under 5% CO_2 or ambient conditions. Then, 50 ml of the culture was collected by centrifugation. Genomic DNA was prepared and barcodes were amplified for fitness determination by Illumina sequencing as described previously[31]. Fitness values were calculated using existing software[31]. Genes were assigned an HCR phenotype if they had a fitness defect of twofold or greater in ambient CO_2 compared with their fitness in 5% CO_2 in two replicate experiments.

CAfree rescue experiments. Electrocompetent CAfree cells were prepared using standard protocols[53] and transformed with pFE plasmids expressing genes of interest by electroporation. CAfree precultures were grown overnight in 10% CO_2 and diluted into 96-well plates (3 µl cells in 250 µl medium). Growth curves were measured by culturing cells in a Tecan M1000 microplate reader under ambient conditions with continuous shaking, and OD$_{600}$ was measured every 15 min. Where samples are marked induced, 200 nM anhydrotetracycline (aTc) was added to the medium. Growth yields are calculated as the maximum OD$_{600}$ achieved after 24 h of growth and normalized to the yield of a wild-type control. Colony forming unit experiments were performed by back-diluting cultures to OD$_{600}$ 0.2 before performing 10× serial dilutions. Then, 3 µl of the samples with an OD$_{600}$ of 0.2 and each of the serial dilutions was spotted onto plates with 200 nM aTc and grown overnight in ambient conditions (400 ppm CO_2). The spot with the highest dilution that yielded more than one colony was counted and results from a minimum of six replicates were averaged for each strain.

Silicone-oil centrifugation measurement of C_i uptake. The silicone-oil filtration method was modified from Dobrinski et al.[54] and used to measure uptake of radiolabelled C_i. Assay tubes were generated using 0.6 ml microcentrifuge tubes containing 20 µl of dense kill solution (66.7% v/v 1 M glycine pH 10, 33.3% v/v Triton X-100) covered by 260 µl of silicone oil (4 parts AR20:3.5 parts AR200). Electrocompetent CAfree cells were prepared using standard protocols and

transformed with pFA-based plasmids that contained the genes of interest by electroporation. CAfree cultures were grown overnight in 10% CO_2, back diluted to an OD_{600} of 0.1 and allowed to grow to mid-log phase in 10% CO_2 in the presence of 200 nM aTc for induction. Cells were then collected by centrifugation, washed once in PBS (pH 7.55) and resuspended to an OD_{600} of 0.6 in PBS supplemented with 0.4% glucose. ^{14}C-labelled sodium bicarbonate (PerkinElmer) was added to a final concentration of 4.1 nM and an activity of 0.23 μCi. Cells were incubated with ^{14}C for 4 min before centrifugation at 17,000g for 4 min to separate cells from the buffer. Pellets were clipped into scintillation vials containing 5 ml Ultima Gold scintillation fluid and 300 μl 3 M NaOH using microcentrifuge tube clippers or medium dog toenail clippers. Counts were measured on a PerkinElmer scintillation counter. ^{14}C counts are normalized to 1 OD_{600} unit of cells added. During inhibition assays, cells were incubated in PBS pH 7.55 with 0.4% glucose and 0.4% DMSO and the inhibitor (100 μM CCCP) for 10 min before the assay.

Generation of the *dabA* phylogenetic tree. We searched the UniProt reference proteome database using the Pfam Hidden Markov Model PF10070.9 with a cut-off e value of 10^{-4}. Our search recovered 941 candidate DabA proteins. These sequences were aligned using MAFFT v.7.310 (ref. [55]) and manually pruned to remove fragments and poorly aligning sequences. The remaining 878 candidate *dabA* sequences were realigned with MAFFT and an approximate maximum likelihood phylogenetic tree was constructed using FastTree v.2.1 (ref. [56]). Taxonomy was assigned to nodes in the tree on the basis of NCBI taxonomy information for the genomes harbouring each sequence. Genomic neighbourhoods for each gene in the tree were determined using the EFIGNT online server[57] and genomes with a *dabB* gene within three genes of *dabA* and oriented in the same direction were considered to have full *DAB* operons. *dabA–dabB* fusions were found by visual inspection of genomic neighbourhoods from genomes that did not have separate *dabB* genes located close to *dabA*.

Generation of *dabB* phylogenetic tree. Homologues of *dabB* were collected manually by searching MicrobesOnline for close homologues of four PF00361 members in the *H. neapolitanus* genome (*dabB1*, *dabB2*, *Hneap_1953* and *Hneap_1130*) and other characterized PF00361 members including *Synechococcus elongatus ndhF1*, *S. elongatus ndhF3* and *S. elongatus ndhF4*. Genes were clustered to 95% similarity and genes with divergent operon structure were removed manually using the MicrobesOnline treeview[58]. *nuoL* from *E. 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 an approximate maximum likelihood phylogenetic tree was constructed using FastTree[59,56]. The tree was visualized using the Interactive Tree of Life[60].

Protein annotation and modelling of structural homology. Secondary structural annotations for DabA and DabB were generated using XtalPred[61]. Structural homology modelling of DabA was performed using Phyre2 and I-TASSER web servers with default parameters[62,63]. A list of close DabB homologues 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, whereas the MAFFT alignment generated during the creation of the DabA tree was used to calculate residue level conservation of DabA proteins (Supplementary Fig. 4b).

Purification of DAB2. Chemically competent BL21-AI *E. coli* were transformed with a pET14b-based vector containing the *dabA* and *dabB* genes. Then, 1 l of 2× YT medium was inoculated with 20 ml of an overnight culture of BL21-AI *E. coli* in LB and carbenicillin and allowed to grow to mid-log phase at 37 °C. When the mid-log phase was reached, cells were induced with 20 ml of 50 mg ml^{-1} 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^{-1} lysozyme and 0.8 ug ml^{-1} DNase I per litre of starting culture and then incubated at room temperature on a rocker for 20 min. Cells were lysed with four passes through a homogenizer (Avestin). The lysate was clarified at 15,000g for 30 min. Membranes were pelleted at 140,000g for 90 min. Membrane pellets were resuspended overnight in 25 ml TBS supplemented with 1 mM phenylmethylsulfonyl fluoride and 1% β-dodecyl-maltoside (DDM, Anatrace) per litre of culture[64]. Membranes were then repelleted at 140,000–200,000g for 60 min 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 and 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.

CA assays. CA-catalysed CO_2 hydration of purified DAB2 complex and hCA was measured using the buffer–indicator assay of Khalifah[65] using a KinTek AutoSF-120 stopped-flow spectrophotometer at 25 °C. The buffer–indicator pair used was TAPS and *m*-cresol purple, respectively, measured at a wavelength of 578 nm using a pathlength of 0.5 cm. The final buffer concentration after mixing was 50 mM TAPS pH 8.0 with the ionic strength adjusted to 50 mM with Na_2SO_4 and 50 μM of pH indicator. The 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_2 (32.9 mM) was prepared by bubbling CO_2 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_2 concentration of 16.5 mM. Progression curves were measured in seven replicates.

X-ray fluorescence spectroscopy for metal analysis. Protein (50–100 μg) in 20–200 μl of TBS and 0.03% DDM was precipitated by addition of four volumes of acetone and incubation at −20 °C for 1 h. Samples were centrifuged at 21,130g for 15 min 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 and 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 H524A, and 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. The experiments were performed at the Lawrence Berkeley National Laboratory Advanced Light Source Beamline 8.3.1.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
All of the Illumina sequencing data are accessible at the NCBI SRA (BioProject accession number: PRJNA546024). All other data are available on GitHub at https://github.com/jackdesmarais/DabTransporterPaper.

Code availability
All custom code is available on GitHub at https://github.com/jackdesmarais/DabTransporterPaper.

References

1. Bathellier, C., Tcherkez, G., Lorimer, G. H. & Farquhar, G. D. Rubisco is not really so bad. *Plant Cell Environ.* **41**, 705–716 (2018).
2. Flamholz, A. et al. Revisiting tradeoffs in Rubisco kinetic parameters. *Biochemistry-US* https://doi.org/10.1021/acs.biochem.9b00237 (2019).
3. Tcherkez, G. The mechanism of Rubisco-catalysed oxygenation. *Plant Cell Environ.* **39**, 983–997 (2016).
4. Bauwe, H., Hagemann, M. & Fernie, A. R. Photorespiration: players, partners and origin. *Trends Plant Sci.* **15**, 330–336 (2010).
5. 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. USA* **103**, 7246–7251 (2006).
6. 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. USA* **107**, 3475–3480 (2010).
7. Mangan, N. M., Flamholz, A., Hood, R. D., Milo, R. & Savage, D. F. pH determines the energetic efficiency of the cyanobacterial CO_2 concentrating mechanism. *Proc. Natl Acad. Sci. USA* **113**, E5354–E5362 (2016).
8. Raven, J. A., Beardall, J. & Sánchez-Baracaldo, P. The possible evolution and future of CO_2-concentrating mechanisms. *J. Exp. Bot.* **68**, 3701–3716 (2017).
9. 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_2 fixation in cyanobacteria and some proteobacteria. *Microbiol. Mol. Biol. Rev.* **77**, 357–379 (2013).
10. Long, B. M., Rae, B. D., Rolland, V., Förster, B. & Price, G. D. Cyanobacterial CO_2-concentrating mechanism components: function and prospects for plant metabolic engineering. *Curr. Opin. Plant Biol.* **31**, 1–8 (2016).
11. Price, G. D., Badger, M. R. & von Caemmerer, S. The prospect of using cyanobacterial bicarbonate transporters to improve leaf photosynthesis in C_3 crop plants. *Plant Physiol.* **155**, 20–26 (2011).
12. 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).

13. Long, B. M. et al. Carboxysome encapsulation of the CO_2-fixing enzyme Rubisco in tobacco chloroplasts. *Nat. Commun.* **9**, 3570 (2018).

14. Price, G. D. & Badger, M. R. Expression of human carbonic anhydrase in the cyanobacterium *Synechococcus* PCC7942 creates a high CO_2-requiring phenotype evidence for a central role for carboxysomes in the CO_2 concentrating mechanism. *Plant Physiol.* **91**, 505–513 (1989).

15. Hopkinson, B. M., Young, J. N., Tansik, A. L. & Binder, B. J. The minimal CO_2-concentrating mechanism of *Prochlorococcus* spp. MED4 is effective and efficient. *Plant Physiol.* **166**, 2205–2217 (2014).

16. 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).

17. Holthuijzen, Y. A., van Dissel-Emiliani, F. F. M., Kuenen, J. G. & Konings, W. N. Energetic aspects of CO_2 uptake in *Thiobacillus neapolitanus*. *Arch. Microbiol.* **147**, 285–290 (1987).

18. Price, G. D. & Badger, M. R. Isolation and characterization of high CO_2-requiring-mutants of the cyanobacterium *Synechococcus* PCC7942: two phenotypes that accumulate inorganic carbon but are apparently unable to generate CO_2 within the carboxysome. *Plant Physiol.* **91**, 514–525 (1989).

19. Marcus, Y., Schwarz, R., Friedberg, D. & Kaplan, A. High CO_2 requiring mutant of *Anacystis nidulans* R_2. *Plant Physiol.* **82**, 610–612 (1986).

20. Bonacci, W. et al. Modularity of a carbon-fixing protein organelle. *Proc. Natl Acad. Sci. USA* **109**, 478–483 (2012).

21. 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).

22. 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).

23. Shibata, M., Ohkawa, H., Katoh, H., Shimoyama, M. & Ogawa, T. Two CO_2 uptake systems in cyanobacteria: four systems for inorganic carbon acquisition in *Synechocystis* sp. strain PCC6803. *Funct. Plant Biol.* **29**, 123–129 (2002).

24. Price, G. D. Inorganic carbon transporters of the cyanobacterial CO_2 concentrating mechanism. *Photosynth. Res.* **109**, 47–57 (2011).

25. Heinhorst, S., Cannon, G. C. & Shively, J. M. in *Complex Intracellular Structures in Prokaryotes* (ed. Shively, J. M.) 141–165 (Springer, 2006).

26. Shively, J. M., Ball, F., Brown, D. H. & Saunders, R. E. Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of *Thiobacillus neapolitanus*. *Science* **182**, 584–586 (1973).

27. Cannon, G. C. et al. Microcompartments in prokaryotes: carboxysomes and related polyhedra. *Appl. Environ. Microbiol.* **67**, 5351–5361 (2001).

28. Mangiapia, M. et al. Proteomic and mutant analysis of the CO_2 concentrating mechanism of hydrothermal vent chemolithoautotroph *Thiomicrospira crunogena*. *J. Bacteriol.* **199**, e00871-16 (2017).

29. 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).

30. Scott, K. M. et al. Diversity in CO_2-concentrating mechanisms among chemolithoautotrophs from the genera *Hydrogenovibrio*, *Thiomicrorhabdus*, and *Thiomicrospira*, ubiquitous in sulfidic habitats worldwide. *Appl. Environ. Microbiol.* **85**, e02096-18 (2019).

31. Wetmore, K. M. et al. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *mBio* **6**, e00306–e00315 (2015).

32. Chaijarasphong, T. et al. Programmed ribosomal frameshifting mediates expression of the α-carboxysome. *J. Mol. Biol.* **428**, 153–164 (2016).

33. Cai, F. et al. The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO_2 leakage barrier. *PLoS ONE* **4**, e7521 (2009).

34. 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).

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 α-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. Mueller-Cajar, O. The diverse AAA+ machines that repair inhibited Rubisco active sites. *Front. Mol. Biosci.* **4**, 31 (2017).

38. 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).

39. Merlin, C., Masters, M., McAteer, S. & Coulson, A. Why is carbonic anhydrase essential to Escherichia coli? *J. Bacteriol.* **185**, 6415–6424 (2003).

40. 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).

41. Cronk, J. D., Endrizzi, J. A., Cronk, M. R., O'Neill, J. W. & Zhang, K. Y. Crystal structure of *E. coli* β-carbonic anhydrase, an enzyme with an unusual pH-dependent activity. *Protein Sci.* **10**, 911–922 (2001).

42. 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).

43. Cronk, J. D. et al. Identification of a novel noncatalytic bicarbonate binding site in eubacterial β-carbonic anhydrase. *Biochemistry* **45**, 4351–4361 (2006).

44. Shibata, M. et al. Distinct constitutive and low-CO_2-induced CO_2 uptake systems in cyanobacteria: genes involved and their phylogenetic relationship with homologous genes in other organisms. *Proc. Natl Acad. Sci. USA* **98**, 11789–11794 (2001).

45. 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_2 hydration in the cyanobacterium, *Synechococcus* sp. PCC7942. *Mol. Microbiol.* **43**, 425–435 (2002).

46. Battchikova, N., Eisenhut, M. & Aro, E. M. Cyanobacterial NDH-1 complexes: novel insights and remaining puzzles. *Biochim. Biophys. Acta* **1807**, 935–944 (2011).

47. Antonovsky, N. et al. Sugar synthesis from CO_2 in *Escherichia coli*. *Cell* **166**, 115–125 (2016).

48. 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).

49. 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).

50. Abuaita, B. H. & Withey, J. H. Bicarbonate induces *Vibrio cholerae* virulence gene expression by enhancing ToxT activity. *Infect. Immun.* **77**, 4111–4120 (2009).

51. 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).

52. Rubin, B. E. et al. The essential gene set of a photosynthetic organism. *Proc. Natl Acad. Sci. USA* **112**, E6634–E6643 (2015).

53. Oakes, B. L., Nadler, D. C. & Savage, D. F. in *Methods in Enzymology*, Vol. 546 (eds Doudna, J. A. & Sontheimer, E. J.) Ch. 23, 491–511 (Academic, 2014).

54. 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).

55. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

56. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).

57. Zallot, R., Oberg, N. O. & Gerlt, J. A. 'Democratized' genomic enzymology web tools for functional assignment. *Curr. Opin. Chem. Biol.* **47**, 77–85 (2018).

58. Dehal, P. S. et al. MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res.* **38**, D396–D400 (2010).

59. Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **27**, 135–145 (2018).

60. 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–W245 (2016).

61. Slabinski, L. et al. XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* **23**, 3403–3405 (2007).

62. 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).

63. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).

64. 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).

65. Khalifah, R. G. The carbon dioxide hydration activity of carbonic anhydrase. *J. Biol. Chem.* **246**, 2561–2573 (1971).

Acknowledgements

Author contributions

J.J.D., A.I.F. and D.F.S. conceived, designed and supervised this study, and wrote the final manuscript with input from all of the authors; J.J.D. performed and analysed most of the biochemical experiments. C.B. performed CA activity assays. J.J.D., E.J.D. and K.W. generated biochemical reagents and strains. T.G.L. performed size-exclusion chromatography. J.J.D., T.G.L., L.M.O. and J.Y.W. developed the purification strategy. J.J.D., T.G.L. and L.M.O. developed the X-ray fluorescence assays. J.J.D. and A.W.C. generated the RB-TnSeq library. J.J.D. and S.D. generated the phylogenetic trees. All of the authors reviewed and approved the final manuscript.

Competing interests

The Regents of the University of California have filed a patent related to this work that lists J.J.D., A.I.F. and D.F.S. as 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.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-019-0520-8.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.F.S.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): David Savage

Last updated by author(s): Jun 3, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used
Data analysis	MAFFT 7.310, ITOl V4, clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/), xtalpred (http://xtalpred.godziklab.org/XtalPred-cgi/xtal.pl), FastTree 2.1 , feba (https://bitbucket.org/berkeleylab/feba) Custom python 3.6.5 scripts (https://github.com/jackdesmarais/DabTransporterPaper) using the following packages: biopython 1.72, ipykernel 4.8.2, ipython 6.4.0, ipython-genutils 0.2.0, ipywidgets 7.2.1, jupyter 1.0.0, jupyter-client 5.2.3, jupyter-console 5.2.0, jupytercore 4.4.0, matplotlib 2.2.2, numpy 1.14.5, pandas 0.23.1, scipy 1.1.0, seaborn 0.8.1, xlrd 1.1.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All illumina sequencing data is accessible as an NCBI SRA (BioProject accession number: PRJNA546024). All other data is available on github at https://github.com/jackdesmarais/DabTransporterPaper.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was chosen following previous publications in the field.
Data exclusions	Data were not excluded from analysis.
Replication	Attempts at replication of the work were successful. Gene essentially analysis and HCR analysis were performed in technical duplicate and results were concordant. HCR analysis also had high levels of internal biological replication. CAfree rescue by DAB2 was observed in two independent transforms as well as in one transformant with a different plasmid backbone and was biologically replicated many times. 14C uptake and CCCP inhibition was observed in 3 independent experiments. SDS-PAGE gels and FPLC were replicated in at least 3 independent purifications. The failure of point mutants to rescue CAfree was confirmed in two independent transformants of each point mutant. X-ray fluorescence measurements were technically replicated, and consistent results were seen for H524A in an independent purification. CA activity assays were replicated in 2 independent purifications. pH dependence of CAfree rescue experiments were performed 4 times with concordant results. Complex I knockout CAfree growth experiments were performed twice with consistent results. DAB homolog CAfree rescue was confirmed in two independent transformants for each homolog. Phylogenetic analysis and gene set enrichment are not typically replicated, and no replication of such analysis was performed here.
Randomization	Strains used in the study were picked at random off of an agar plate.
Blinding	Libraries were pooled so individual strains were never handled. Library data was collected and analyzed in an automated fashion so there was no one to blind for those assays. All growth curves and growth yields had automated data collection and analysis, so there was no one to blind. 14C uptake assays, spotting assays, and biochemical assays were not blinded.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging