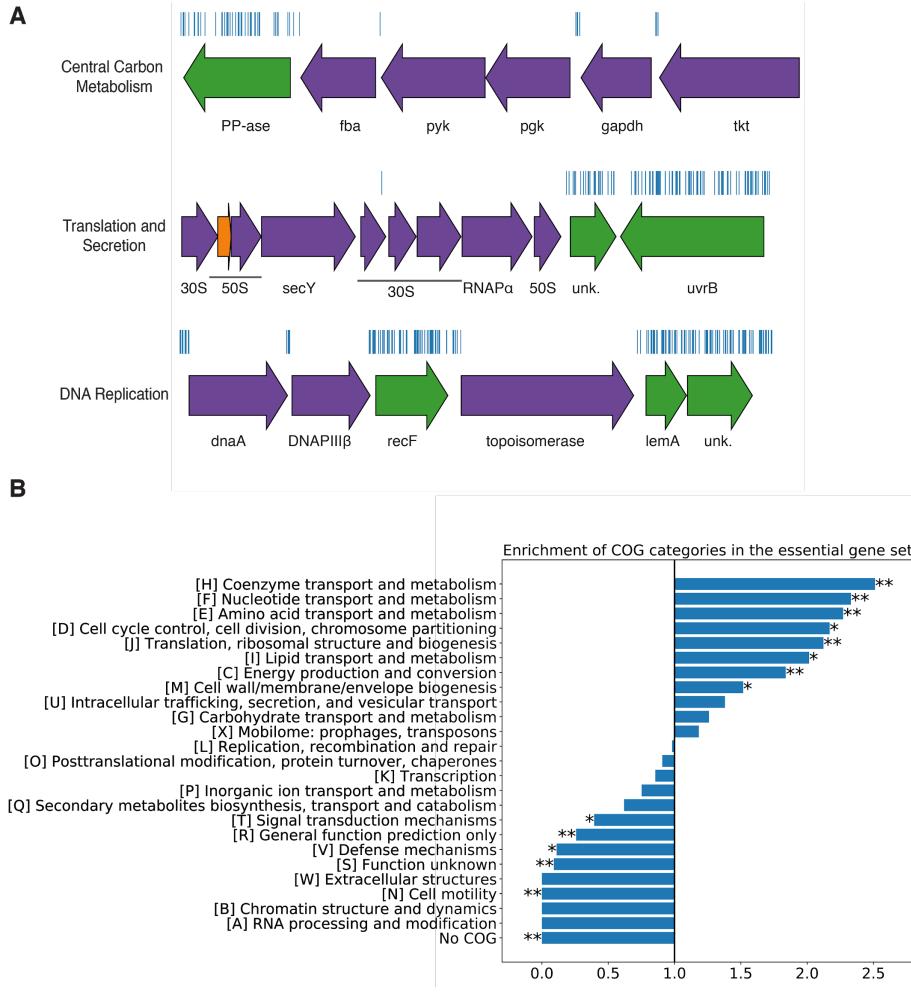


884 **Supplementary Information:**

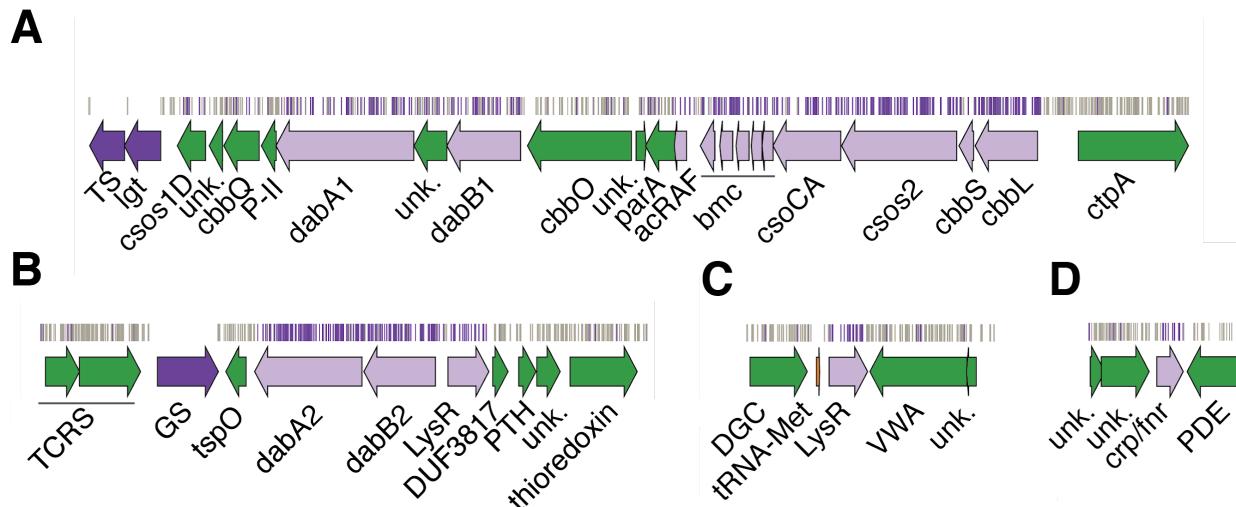
- 885 **Supplemental File 1.** Important strains and reagents.
- 886 **Supplemental File 2.** Transposon insertion information and essentiality determination by gene.
- 887 **Supplemental File 3.** Fitness effects and HCR phenotype by gene.
- 888 **Supplemental File 4.** Genes used to generate figure S3A.
- 889 **Supplemental File 5.** Genes used to generate figure 5A.



890

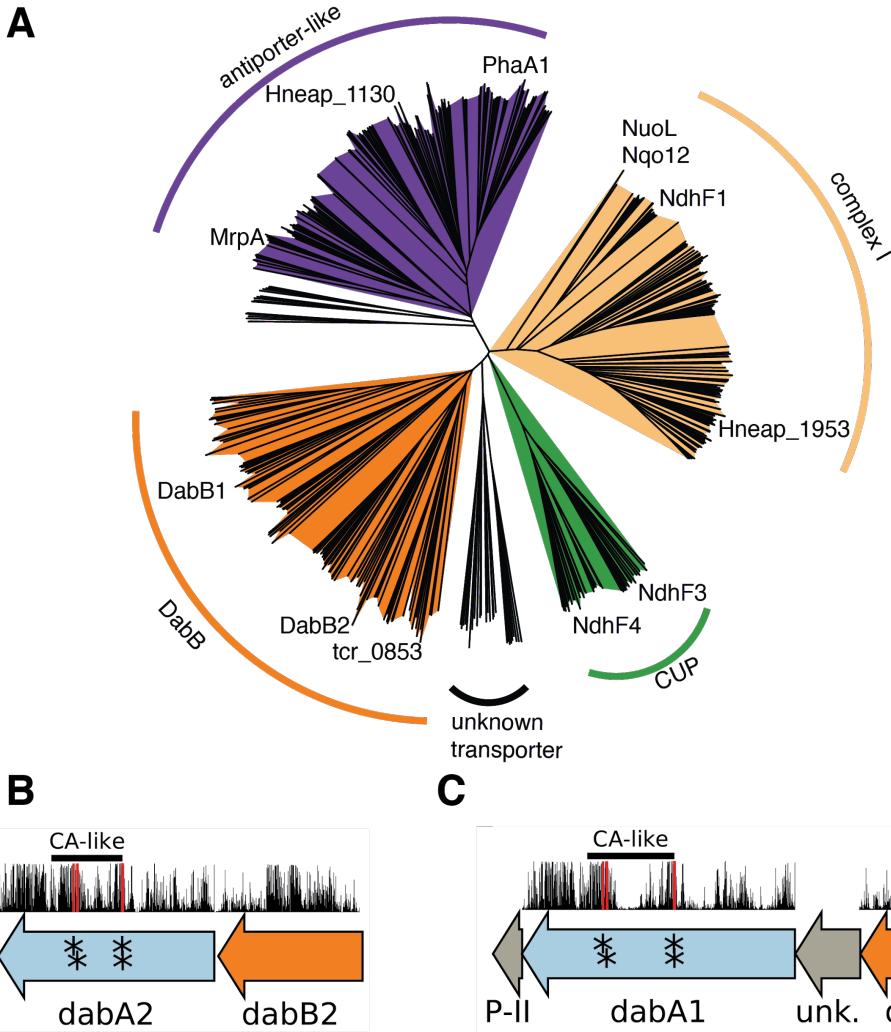
**Figure S1 The essential gene set is enriched for COGs associated with essential cellular processes.** **A.** Representative essential genes and nonessential genes in the *Hnea* genome. The blue track indicates the presence of an insertion. Genes in purple were called essential and genes in green are 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 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  $\beta$ . **B.** COG enrichments were calculated by dividing the fraction of genes in the essential gene set associated with this COG category by the fraction of genes in the genome associated with this category. \*\* denotes that this COG is enriched (or depleted) with Bonferroni corrected  $P < 0.05$  by a hypergeometric test, and \*\*\* denotes  $P < 5 \times 10^{-4}$ . Exact p values are as follows for each category, No COG:  $7 \times 10^{-68}$ , C:  $1.1 \times 10^{-5}$ , D:  $1 \times 10^{-2}$ , E:  $< 7 \times 10^{-68}$ , F:  $4.3 \times 10^{-6}$ , H:  $< 7 \times 10^{-68}$ , I:  $6.3 \times 10^{-4}$ , J:  $< 7 \times 10^{-68}$ , M:  $6.9 \times 10^{-3}$ , N:  $6.5 \times 10^{-6}$ , R:  $6.17 \times 10^{-5}$ , S:  $7.7 \times 10^{-8}$ , T:  $2.8 \times 10^{-2}$ , V:  $1.1 \times 10^{-2}$ . In panel A, the following abbreviations are used: exopolyphosphatase (PP-ase), fructose-bisphosphate aldolase class II (fba), pyruvate kinase (pyk), phosphoglycerate kinase (pgk), type I glyceraldehyde-3-phosphate dehydrogenase (gapdh), transketolase (tkt), 30S ribosomal protein (30S), 50S ribosomal protein (50S), preprotein

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



911  
 912 **Figure S2 Genomic context of *Hnea* HCR genes identified in our genome-wide screen.** Panels A-D  
 913 show regions of the *Hnea* genome containing genes annotated as HCR. Essential genes are in dark  
 914 purple, HCR genes are in light purple, and other genes are in green. The top tracks show the presence of  
 915 an insertion in that location. Insertions are colored grey unless they display a twofold or greater  
 916 fitness defect in ambient CO<sub>2</sub>, in which case they are colored purple. **A.** The gene cluster containing the  
 917 carboxysome operon and a second CCM-associated operon. This second operon contains acRAF, a  
 918 FormIC associated cbbQOQ-type Rubisco activase and dabAB1. **B.** The DAB2 operon and surrounding  
 919 genomic context. **C.** The genomic context of a lysR-type transcriptional regulator that shows an HCR  
 920 phenotype. **D** Genomic context of a crp/fnr-type transcriptional regulator that displays an HCR phenotype.  
 921 Abbreviations for Figure S2: thymidylate synthase (TS), prolipoprotein diacylglycerol transferase (Igt),  
 922 Rubisco activase Rubisco activase subunits (cbbQOQ), nitrogen regulatory protein P-II (P-II), ParA family  
 923 protein (parA), csos1CAB and csos4AB (bmc), copper-translocating P-type ATPase (ctpA), DNA-binding  
 924 response regulator and two-component sensor histidine kinase (TCRS), glutamate--ammonia ligase (GS),  
 925 tryptophan-rich sensory protein (tspO), DUF3817 domain-containing protein (DUF3817), aminoacyl-tRNA  
 926 hydrolase (PTH), thioredoxin domain-containing protein (thioredoxin), sensor domain-containing  
 927 diguanylate cyclase (DGC), methionine tRNA (tRNA-Met), VWA domain-containing protein (VWA),  
 928 diguanylate phosphodiesterase (PDE).

929



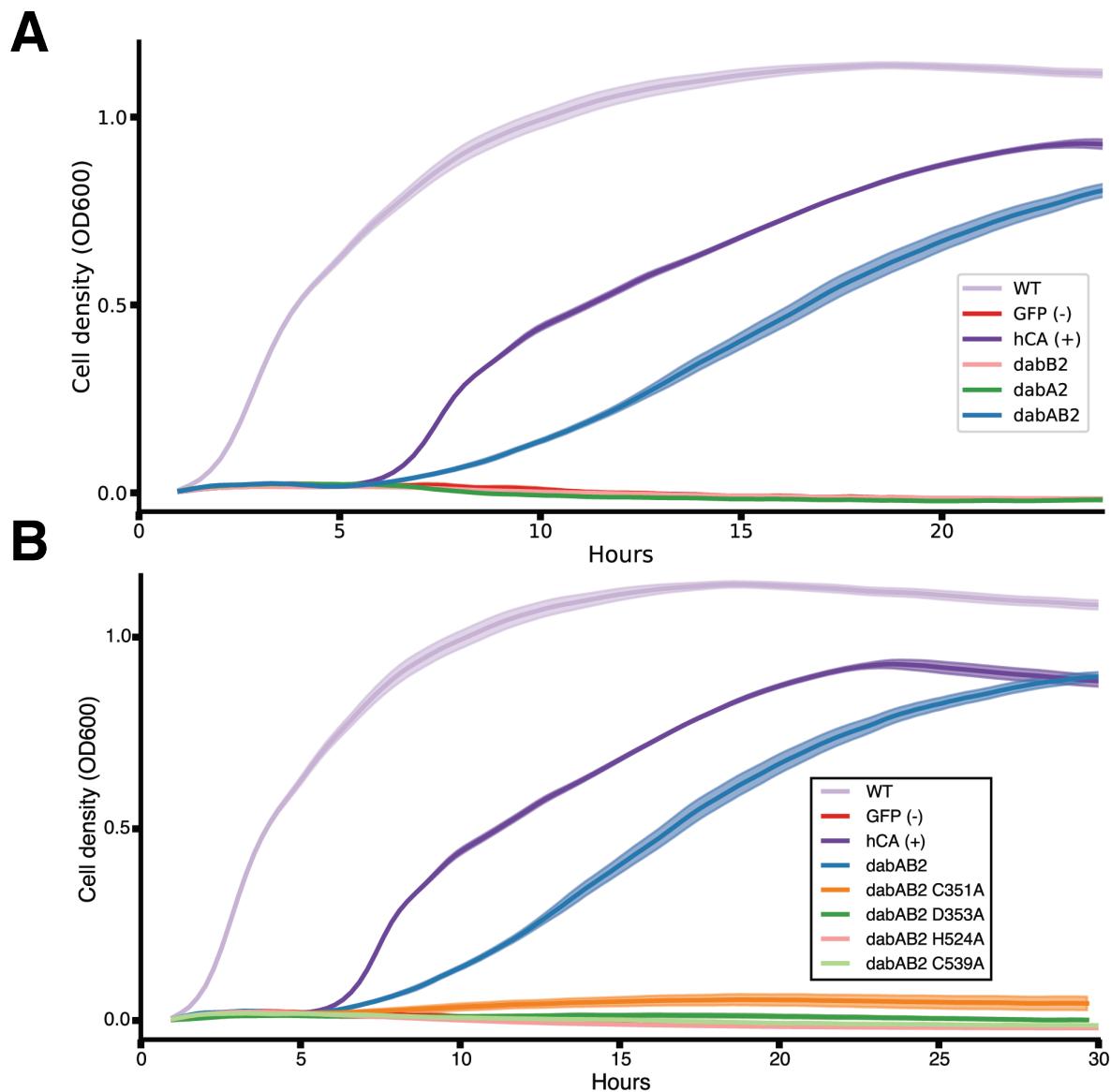
930  
931 **Figure S3 PF0361 contains multiple subfamilies, but some regions of DAB subunits are highly**  
932 **conserved.** A. PF0361 is a large and diverse protein family containing multiple subgroups with different  
933 documented activities. These subfamilies include Mrp-family cation antiporters, proton translocating  
934 subunits of complex I, membrane subunits of CUP (CO<sub>2</sub> uptake protein) complexes, and DabB proteins.  
935 These subfamilies are highly diverged and perform a variety of activities. This means that it is not  
936 possible to draw conclusions about the mechanism of DAB complexes just from their homology to  
937 PF0361. This panel contains a nearest neighbor tree of PF0361 genes. Clades were colored according to  
938 the presence of genes with known functions. The purple clade contains the *Bacillus subtilis* and  
939 *Staphylococcus aureus* MrpA cation antiporter subunits and the *Sinorhizobium meliloti* antiporter PhaA1.  
940 The light orange clade contains the known cation translocating subunits of complex I: nuoL from  
941 *Escherichia coli*, Nqo12 from *Thermus thermophilus*, and NdF1 from both *Synechococcus elongatus*  
942 PCC7942 and *Thermosynechococcus elongatus* BP-1. The green clade contains CUP-associated  
943 membrane subunits ndhF3 from both *Synechococcus elongatus* PCC7942 and *Thermosynechococcus*  
944 *elongatus* BP-1 and ndhF4 from the same two species. The dark orange clade includes DabB1-2  
945 and tcr\_0853 from *Thiomicrospira crunogena*. We note that the clade containing DabB1-2 is distinct from  
946 that containing known complex I subunits or to mrp-family antiporters. This tree is consistent with our

947 model, where DabB is not bound to a redox-coupled complex but rather couples redox-independent  
948 cation transport to CA activity (as shown in Figure 5). No conclusions should be drawn from the number  
949 of sequences in each clade as an exhaustive search for homologs was not performed to ensure that all  
950 members of each clade are represented. **B** and **C** As noted in the text and shown in Figure 2B, DAB1 is a  
951 segment of an 11-gene operon directly downstream of the carboxysome operon that contains CCM-  
952 associated genes. Both DAB1 (**B**) and DAB2 (**C**) “operons” contain two distinct genes that we label DabB  
953 and DabA. DabA is annotated as Domain of Unknown Function 2309 (DUF2309, PFAM:PF10070) and  
954 appears to be a soluble protein. Approximately one third of dabA is distantly homologous to a type II  $\beta$ -  
955 CA. CA-like regions are marked with a line, and the four residues expected to be involved in binding the  
956 catalytic zinc ion are marked by asterisks. The height of the asterisks has been varied to make them  
957 distinguishable despite proximity in sequence space. DabB is homologous to a cation transporter in the  
958 same family as the H<sup>+</sup> pumping subunits of respiratory complex I (PFAM:PF00361). The DAB1 operon  
959 also contains a protein of unknown function between DabA1 and DabB1. This protein has distant  
960 homology to DabA1 but is truncated to half the length. Vertical bars above the genes indicate percent  
961 conservation of that particular amino acid position in a multiple sequence alignment (Methods). Active site  
962 residues are in red. All active site residues are highly conserved with percent identities of greater than  
963 99%. One active site cysteine and the active site aspartate residue are the two most conserved residues  
964 in DabA with 99.9% identity each.

965

966

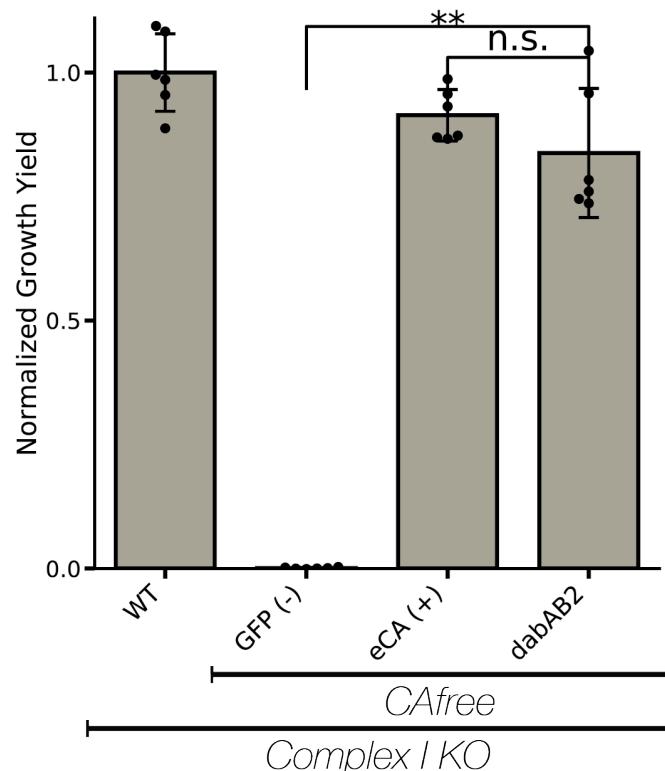
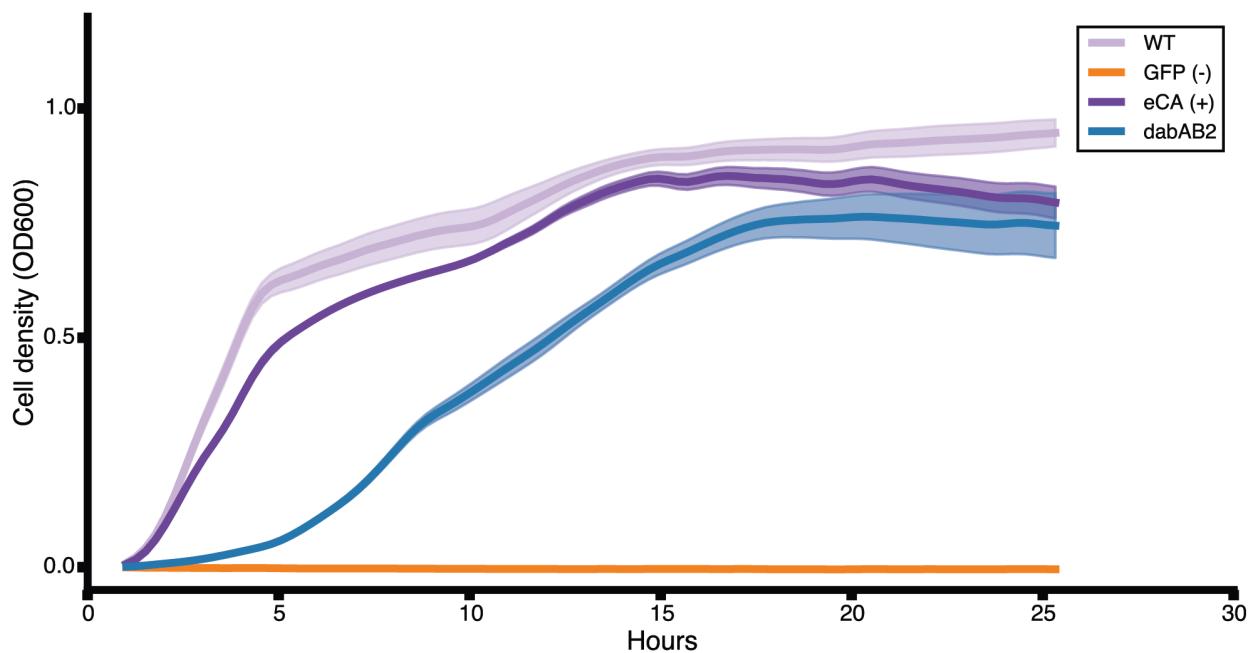
967



968

969 **Figure S4. Expression of DabAB2 rescues growth of CAfree *E. coli* in ambient CO<sub>2</sub>. A.** These  
 970 growth curves were used to generate the growth yield values in Figure 3B. Mean OD600 is plotted +/-  
 971 standard error for four replicate cultures. Wild-type *E. coli* (BW25113) and CAfree strains expressing  
 972 either dabAB2 or human carbonic anhydrase II (hCA) grow in ambient CO<sub>2</sub> while CAfree expressing GFP,  
 973 dabB2 alone, or dabA2 alone fail to grow. **B.** These growth curves were used to generate the growth yield  
 974 values in Figure 4B. Mean OD600 is plotted +/- standard error of four replicate cultures. Wild type cells  
 975 and CAfree expressing either DabAB2 or human carbonic anhydrase II (hCA) grow robustly. CAfree cells  
 976 expressing putative active site mutants of DabAB2 (C351, D353, H524, or C539) grow as poorly as the  
 977 negative control – CAfree expressing superfolder GFP in the same plasmid backbone.

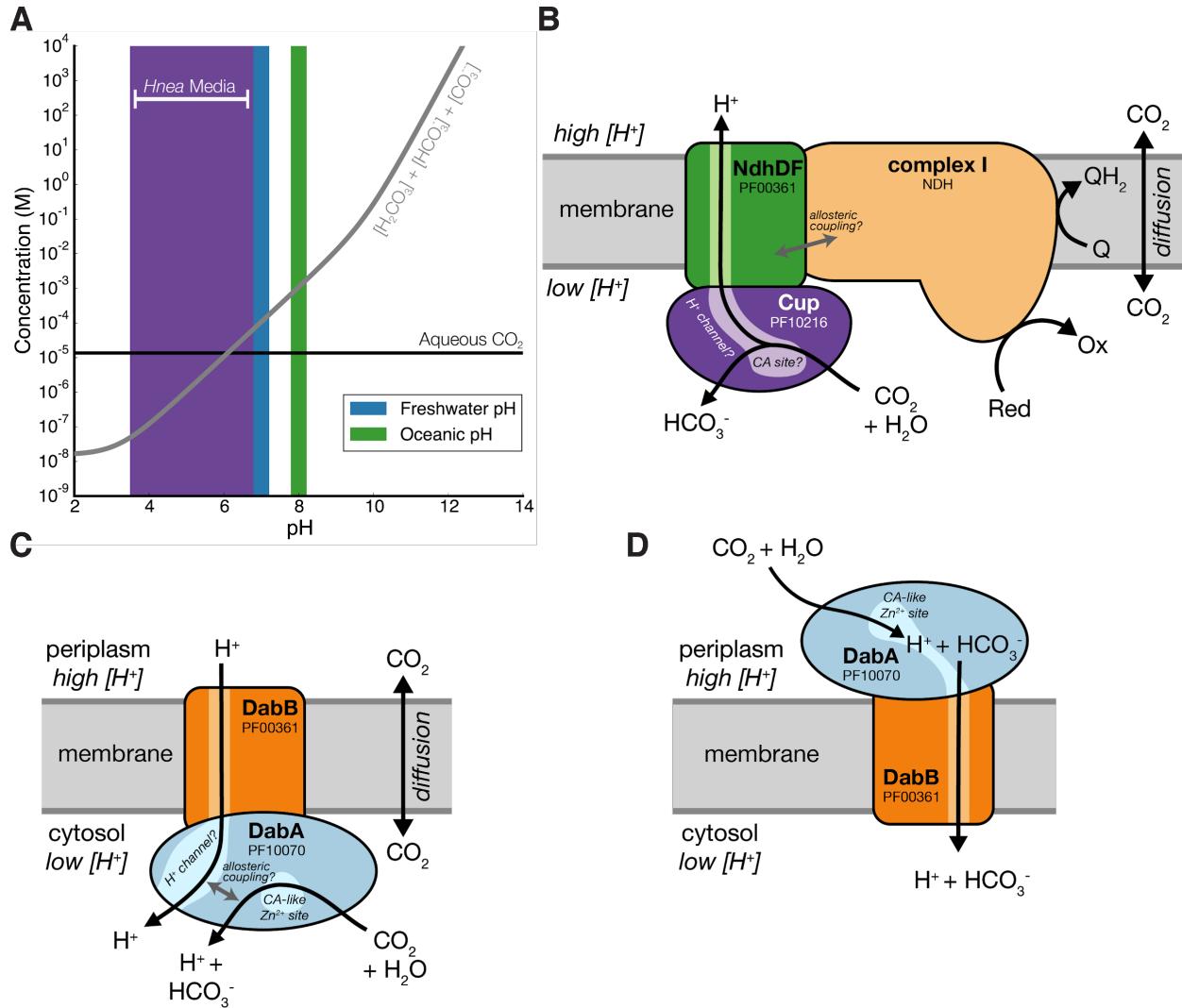
978

**A****B**

979

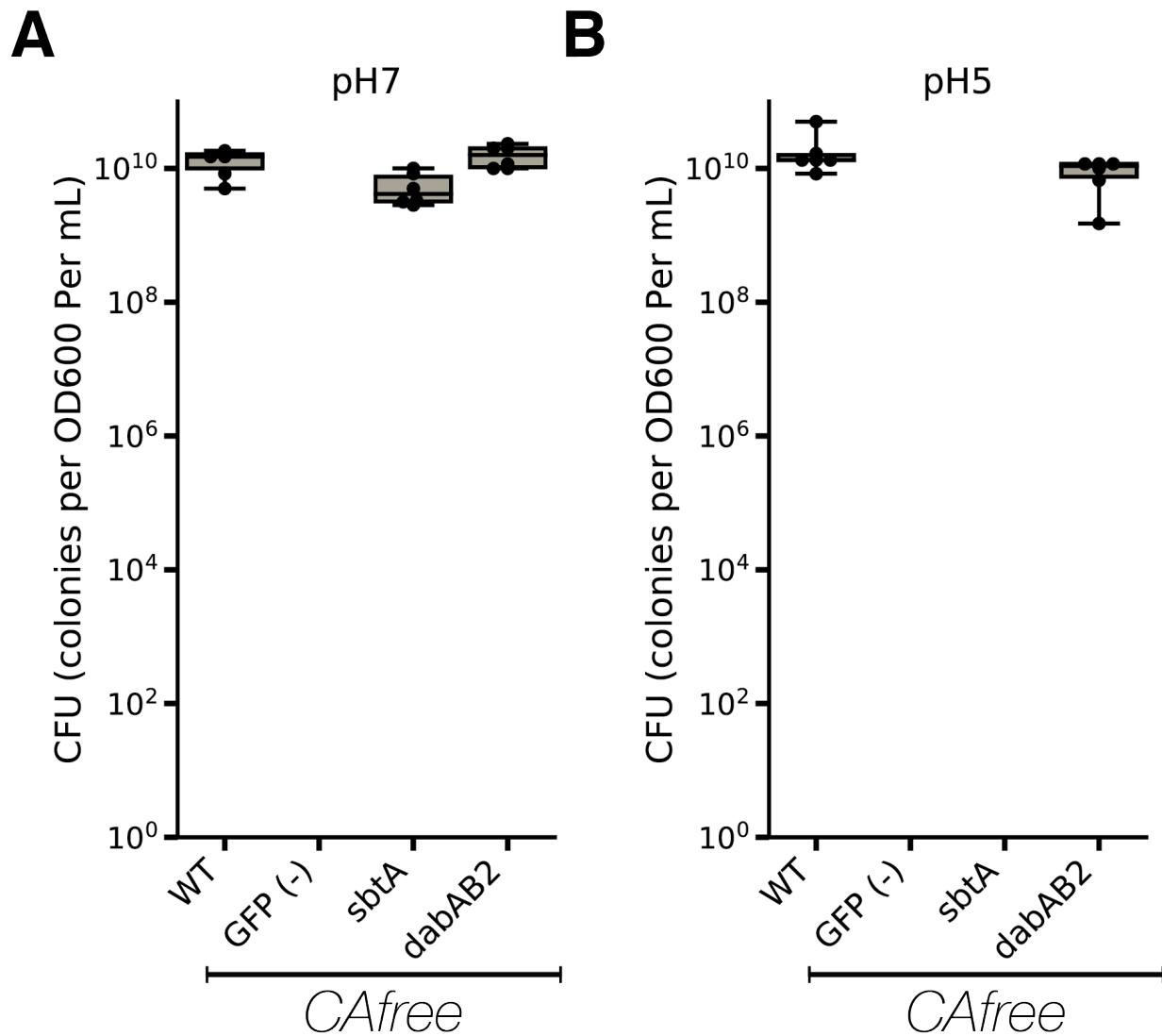
980 **Figure S5. DAB2 function is not dependant on complex 1.** A. DAB2 is still able to rescue growth of  
981 CAfree cells in the absence of Complex I ( $\Delta(nuoA-nuoN)$ ). dabAB2 rescues better than GFP ( $t=15.7$ ,  
982  $p=2.37 \times 10^{-8}$ ). Error bars represent standard deviation of six replicate cultures. “n.s.” denotes means do  
983 not differ significantly, “\*\*” denotes that means differ with bonferroni corrected  $P < 0.05$  by a two-tailed T-  
984 test, and “\*\*\*” denotes  $P < 5 \times 10^{-4}$ . B. These growth curves were used to generate the growth yield values

985 in Figure S5A. Mean OD600 is plotted +/- standard error of six replicate cultures. All strains are Complex I  
986 knockout strains. DAB2 is still able to rescue growth of CAfree cells in the absence of Complex I.  
987

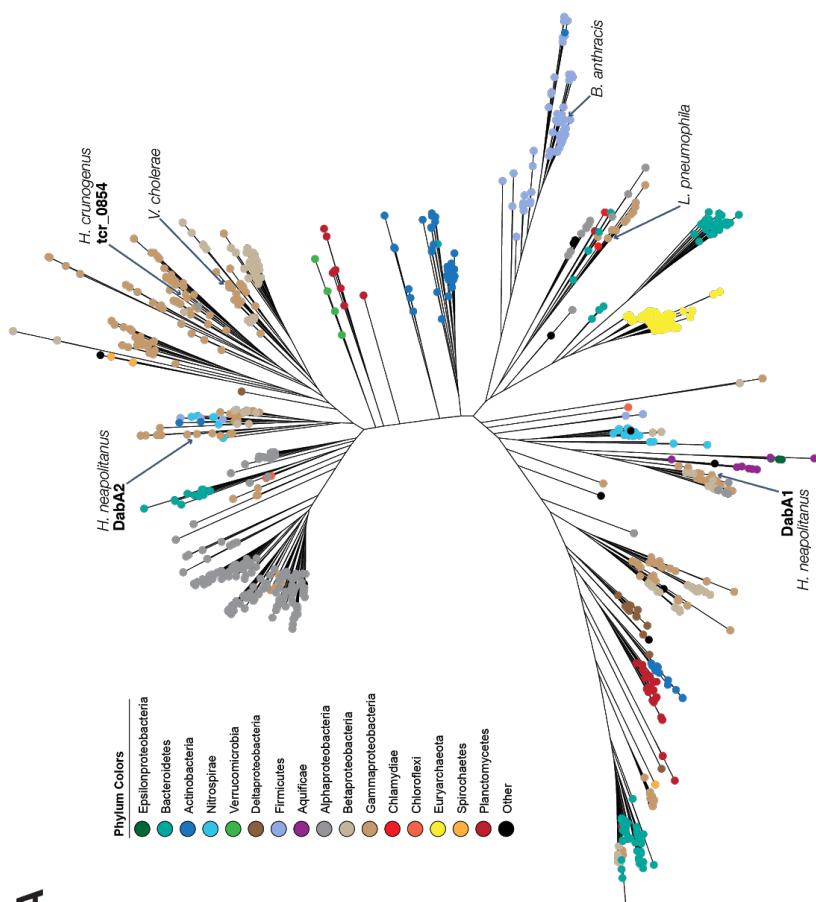
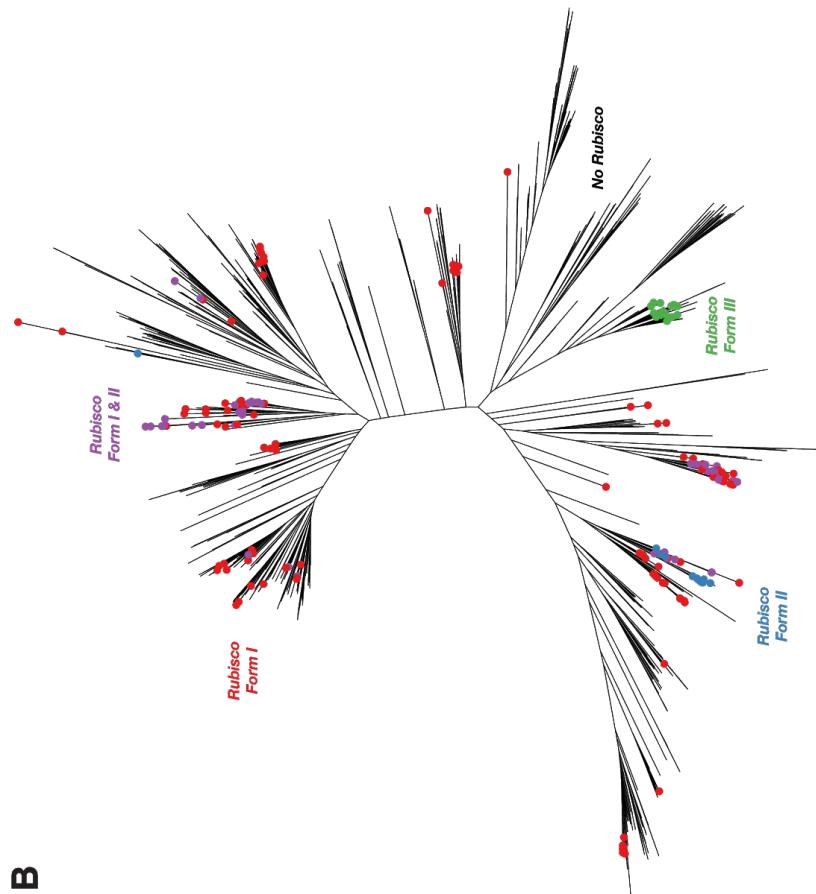


988  $\text{HCO}_3^-$   
 989 **Figure S6. Comparison of models of vectorial CA activity for DABs and the Cyanobacterial CUP**  
 990 **systems. A.** Equilibrium concentrations of dissolved inorganic carbon as a function of pH. In this plot we  
 991 assume the growth medium is in Henry's law equilibrium with present-day atmosphere (400 PPM  $\text{CO}_2$ ) at  
 992 25 °C giving a soluble  $\text{CO}_2$  concentration of roughly 15  $\mu\text{M}$ . The equilibrium concentrations of hydrated  $\text{C}_i$   
 993 species ( $\text{H}_2\text{CO}_3$ ,  $\text{HCO}_3^-$ ,  $\text{CO}_3^{2-}$ ) is determined by the pH. As such, the organisms will "see" a  $\text{C}_i$  species in  
 994 very different ratios depending on the environmental pH. In a oceanic pH near 8,  $\text{HCO}_3^-$  dominates the  $\text{C}_i$   
 995 pool.  $\text{HCO}_3^-$  is also the dominant constituent of the  $\text{C}_i$  pool in freshwater, but less so (by a factor of ~10  
 996 since freshwater and oceanic environments differ by about 1 pH unit). In acid conditions (pH < 6.1)  $\text{CO}_2$   
 997 will be the dominant constituent of the  $\text{C}_i$  pool. The pH of our *Hnea* culture media ranges from 6.8 (when  
 998 freshly made) to ~3.5 when cells reach stationary phase (*Hnea* make  $\text{H}_2\text{SO}_4$  as a product of their sulfur  
 999 oxidizing metabolism). As such we expect that *Hnea* regularly experiences environments wherein it is  
 1000 advantageous to pump  $\text{CO}_2$  and not  $\text{HCO}_3^-$ . **B.** CupA/B proteins are CA-like subunits of a class of

1001 cyanobacterial Ci uptake systems. Cup-type systems are believed to couple electron transfer to vectorial  
1002 CA activity and, potentially, outward-directed proton pumping. This model is based on the observation  
1003 that Cup systems displace the two distal H<sup>+</sup>-pumping subunits of the cyanobacterial complex I and  
1004 replace them with related subunits that bind CupA/B (illustrated in green as NdhD/F). **C**. As our data are  
1005 consistent with DAB2 functioning as a standalone complex (i.e. DabAB do not appear to bind or require  
1006 the *E. coli* complex I), we propose a different model for DAB function where energy for unidirectional  
1007 hydration of CO<sub>2</sub> is drawn from the movement of cations along their electrochemical gradient (right panel  
1008 above). **D**. An alternative model for DAB activity is that DabA is localized to the periplasm and DabB is  
1009 functioning as a H<sup>+</sup> : HCO<sub>3</sub><sup>-</sup> symporter. In this model DabA CA activity is made vectorial by removal of  
1010 products. Energy is provided in the form of the PMF driving H<sup>+</sup> (and therefore HCO<sub>3</sub><sup>-</sup>) uptake. This model  
1011 is not preferred because no secretion signals were observed in the DabA sequence. Moreover, the  
1012 *Acidimicrobium ferrooxidans* genome contains an apparent DabA:DabB fusion protein. The predicted  
1013 architecture the fusion would place DabA in the cytoplasm.

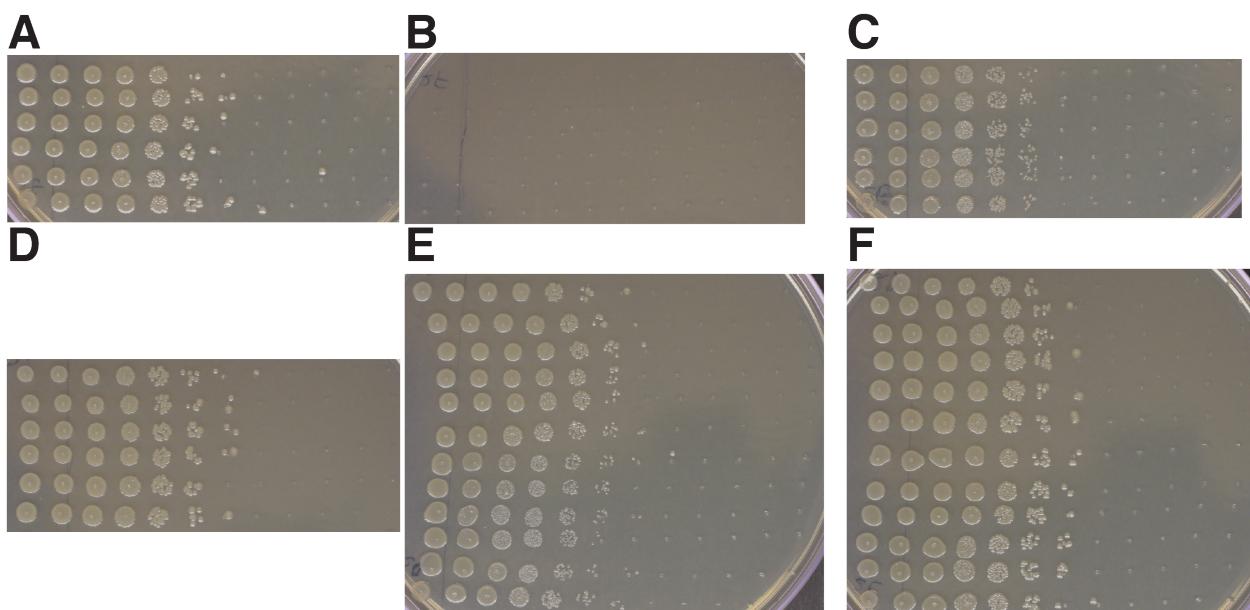


1014  
1015 **Figure S7. pH independence of *dabAB2* rescue of CAfree** Colony forming units per OD600 per ml  
1016 were measured on LB plates with induction in air at both pH 7 (A.) and 5 (B.). *dabAB2* rescued growth at  
1017 both pH7 and pH 5, *sbtA* only rescued growth at pH 7. Whiskers represent the range of the data, the box  
1018 represents the interquartile range, and the middle line represents the median. Data is from 6 replicate  
1019 platings of all conditions.

**B****A**

1021 **Figure S8. Fully annotated approximate maximum likelihood phylogenetic trees of DabA.** **A.** A  
1022 phylogenetic tree emphasizing the clades containing high-confidence DabA homologs. DabA homologs  
1023 are found in > 15 prokaryotic clades, including some archaea. *Hnea* DabA1 and DabA2 represent two  
1024 different groupings that are commonly found in proteobacteria. The tcr\_0854 gene of *H. crunogenus* is  
1025 more closely related to DabA2 than DabA1. Inspecting the tree reveals several likely incidents of  
1026 horizontal transfer, e.g. between proteobacteria and Firmicutes, Nitrospirae and Actinobacteria.  
1027 Moreover, the genomes of several known pathogens contain a high-confidence DabA homolog, including  
1028 *B. anthracis*, *L. pneumophila*, *V. cholerae*. **B.** Association of various Rubisco isoforms with DabA  
1029 homologs. Many organisms that have DabA also have a Rubisco. However, there are numerous  
1030 examples of DabA homologs that are found in genomes with no Rubisco (denoted by leaves with no  
1031 colored marking), suggesting that this uptake system might play a role in heterotrophic metabolism. DabA  
1032 is most-frequently associated with Form I Rubiscos (red and purple leaves in panel B), which is sensible  
1033 because all known bacterial CCMs involve a Form I Rubisco exclusively. Some DabA-bearing genomes  
1034 have only a Form II Rubisco (blue) and the Euryarchaeota genomes have that DabA have a Form III  
1035 Rubisco (green) or none at all.

1036



1037  
1038 **Figure S9. Plates used for determining CFU counts for Figure 5B.** **A.** Wt positive control. **B.** CAfree  
1039 sfGFP negative control does not rescue. **C.** CAfree hCA positive control rescues growth. **D.** CAfree DAB2  
1040 rescues growth. **E.** baDAB from *Bacillus anthracis* rescues growth of CAfree. **F.** vcDAB from *Vibrio*  
1041 *cholera* rescues growth of CAfree. Panels **A-D** represent 6 technical replicates of the plating. Panels **E**  
1042 and **F** represent 6 technical replicates each of 2 biological replicates. In all panels, the first spot

1043 represents 3ul of an OD 0.2 culture grown at 10% CO<sub>2</sub> each subsequent spot is 3 ul of a 1:10 dilution of  
1044 the previous spot.  
1045