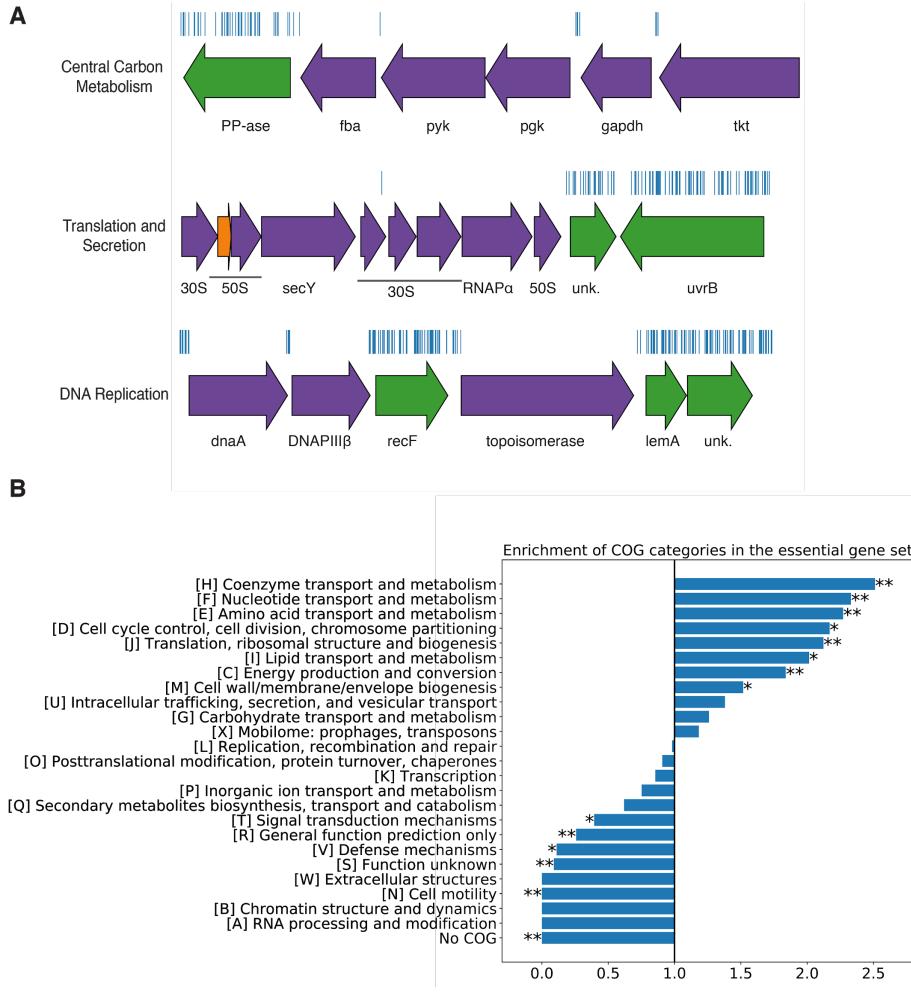


868 **Supplementary Information:**

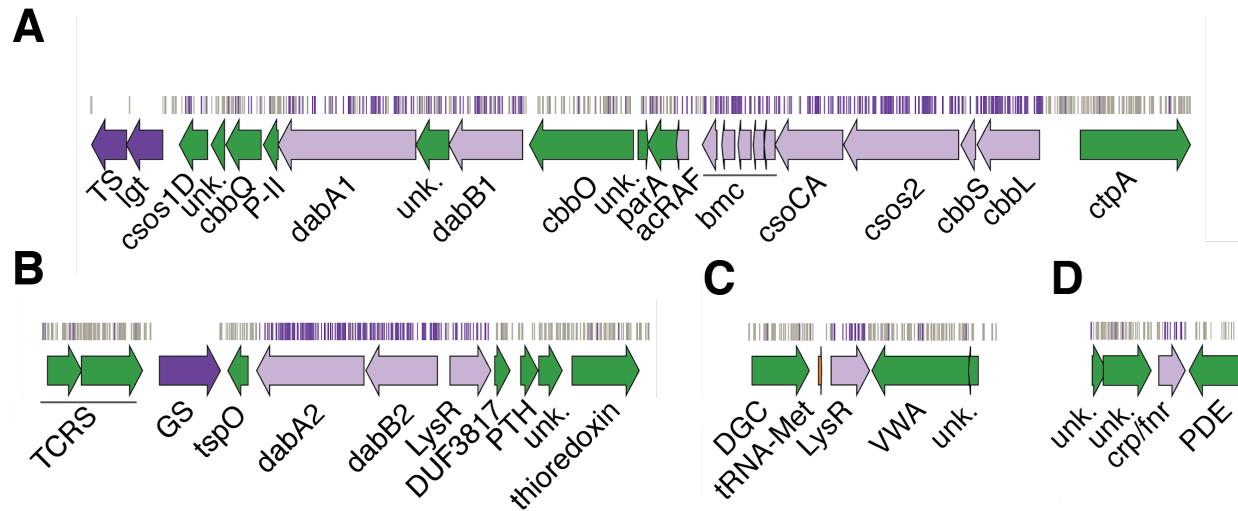
- 869 **Supplemental File 1.** Important strains and reagents.
- 870 **Supplemental File 2.** Transposon insertion information and essentiality determination by gene.
- 871 **Supplemental File 3.** Fitness effects and HCR phenotype by gene.
- 872 **Supplemental File 4.** Genes used to generate figure S3A.
- 873 **Supplemental File 5.** Genes used to generate figure 5A.



874

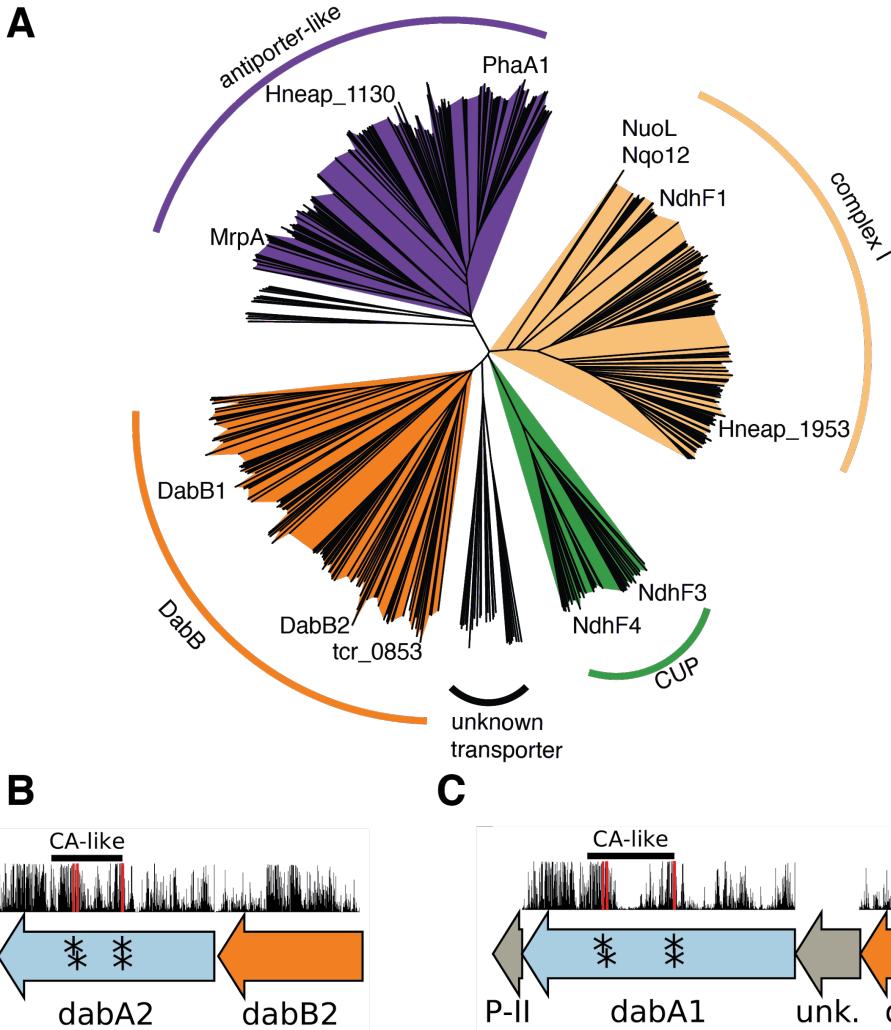
875 **Figure S1 The essential gene set is enriched for COGs associated with essential cellular**
 876 **processes.** A. Representative essential genes and nonessential genes in the *Hnea* genome. The blue
 877 track indicates the presence of an insertion. Genes in purple were called essential and genes in green are
 878 nonessential. Genes labeled "unk." are hypothetical proteins. The first genomic locus contains 5 essential
 879 genes involved in glycolysis or the CBB cycle including pyruvate kinase (pyk) and transketolase (tkt). The
 880 8 essential genes in the second locus encode 30S and 50S subunits of the ribosome, the secY secretory
 881 channel, and an RNA polymerase subunit. Essential genes in the third example locus include
 882 topoisomerase and DNA polymerase III β . B. COG enrichments were calculated by dividing the fraction of
 883 genes in the essential gene set associated with this COG category by the fraction of genes in the genome
 884 associated with this category. ** denotes that this COG is enriched (or depleted) with Bonferroni
 885 corrected $P < 0.05$ by a hypergeometric test, and *** denotes $P < 5 \times 10^{-4}$. In panel A, the following
 886 abbreviations are used: exopolyphosphatase (PP-ase), fructose-bisphosphate aldolase class II (fba),
 887 pyruvate kinase (pyk), phosphoglycerate kinase (pgk), type I glyceraldehyde-3-phosphate dehydrogenase
 888 (gapdh), transketolase (tkt), 30S ribosomal protein (30S), 50S ribosomal protein (50S), preprotein
 889 translocase subunit SecY (SecY), DNA-directed RNA polymerase subunit alpha (RNAPa), hypothetical
 890 protein (unk.), excinuclease ABC subunit UvrB (UvrB), chromosomal replication initiator protein dnaA

891 (dnaA), DNA polymerase III subunit beta (DNAPIII β), DNA replication and repair protein recF (recF), DNA
892 topoisomerase (ATP-hydrolyzing) subunit B (topoisomerase), lemA family protein (LemA).



893
894 **Figure S2 Genomic context of *Hnea* HCR genes identified in our genome-wide screen.** Panels **A-D**
895 show regions of the *Hnea* genome containing genes annotated as HCR. Essential genes are in dark
896 purple, HCR genes are in light purple, and other genes are in green. The top tracks show the presence of
897 an insertion in that location. Insertions are colored grey unless they display a twofold or greater
898 fitness defect in ambient CO₂, in which case they are colored purple. **A.** The gene cluster containing the
899 carboxysome operon and a second CCM-associated operon. This second operon contains acRAF, a
900 FormIC associated cbbQOQ-type Rubisco activase and dabAB1. **B.** The DAB2 operon and surrounding
901 genomic context. **C.** The genomic context of a lysR-type transcriptional regulator that shows an HCR
902 phenotype. **D** Genomic context of a crp/fnr-type transcriptional regulator that displays an HCR phenotype.
Abbreviations for Figure S2: thymidylate synthase (TS), prolipoprotein diacylglycerol transferase (lgt),
Rubisco activase Rubisco activase subunits (cbbQOQ), nitrogen regulatory protein P-II (P-II), ParA family
protein (parA), csos1CAB and csos4AB (bmc), copper-translocating P-type ATPase (ctpA), DNA-binding
response regulator and two-component sensor histidine kinase (TCRS), glutamate--ammonia ligase (GS),
tryptophan-rich sensory protein (tspO), DUF3817 domain-containing protein (DUF3817), aminoacyl-tRNA
hydrolase (PTH), thioredoxin domain-containing protein (thioredoxin), sensor domain-containing
diguanylate cyclase (DGC), methionine tRNA (tRNA-Met), VWA domain-containing protein (VWA),
diguanylate phosphodiesterase (PDE).

911

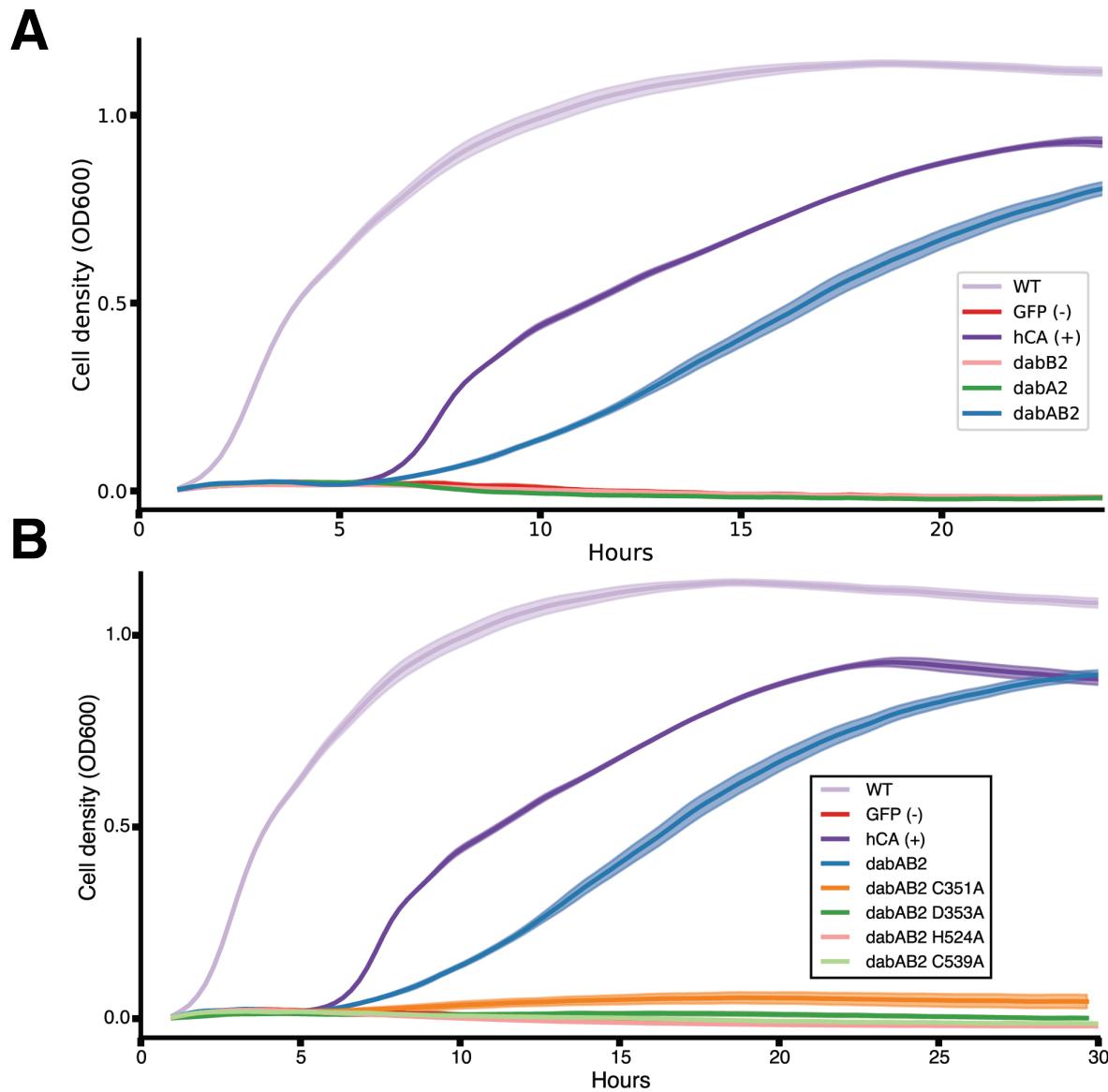


912
913 **Figure S3 PF0361 contains multiple subfamilies, but some regions of DAB subunits are highly**
914 **conserved.** A. PF0361 is a large and diverse protein family containing multiple subgroups with different
915 documented activities. These subfamilies include Mrp-family cation antiporters, proton translocating
916 subunits of complex I, membrane subunits of CUP (CO₂ uptake protein) complexes, and DabB proteins.
917 These subfamilies are highly diverged and perform a variety of activities. This means that it is not
918 possible to draw conclusions about the mechanism of DAB complexes just from their homology to
919 PF0361. This panel contains a nearest neighbor tree of PF0361 genes. Clades were colored according to
920 the presence of genes with known functions. The purple clade contains the *Bacillus subtilis* and
921 *Staphylococcus aureus* MrpA cation antiporter subunits and the *Sinorhizobium meliloti* antiporter PhaA1.
922 The light orange clade contains the known cation translocating subunits of complex I: nuoL from
923 *Escherichia coli*, Nqo12 from *Thermus thermophilus*, and NdF1 from both *Synechococcus elongatus*
924 PCC7942 and *Thermosynechococcus elongatus* BP-1. The green clade contains CUP-associated
925 membrane subunits ndhF3 from both *Synechococcus elongatus* PCC7942 and *Thermosynechococcus*
926 *elongatus* BP-1 and ndhF4 from the same two species. The dark orange clade includes DabB1-2
927 and tcr_0853 from *Thiomicrospira crunogena*. We note that the clade containing DabB1-2 is distinct from
928 that containing known complex I subunits or to mrp-family antiporters. This tree is consistent with our

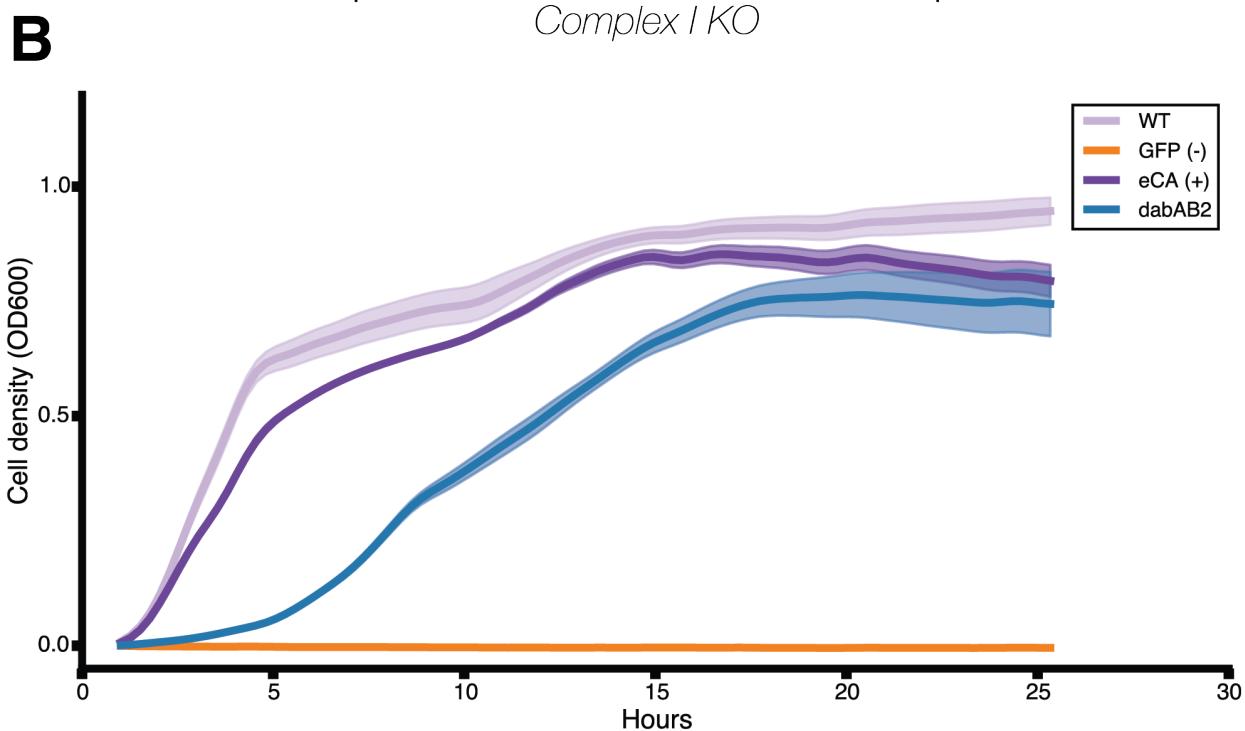
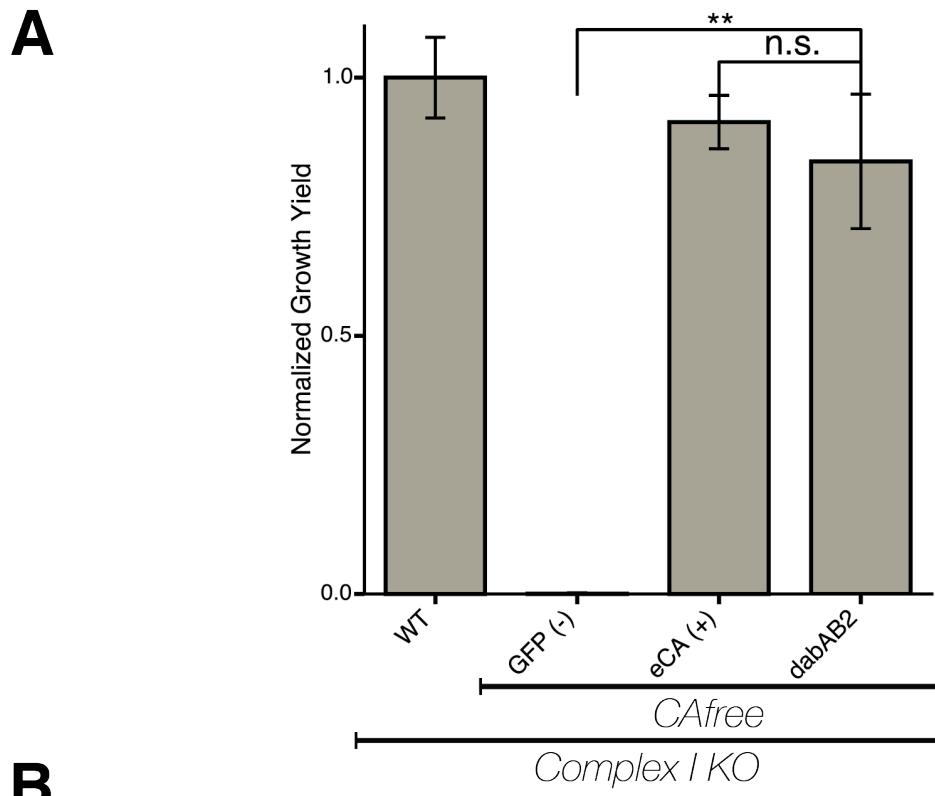
929 model, where DabB is not bound to a redox-coupled complex but rather couples redox-independent
930 cation transport to CA activity (as shown in Figure 5). No conclusions should be drawn from the number
931 of sequences in each clade as an exhaustive search for homologs was not performed to ensure that all
932 members of each clade are represented. **B** and **C** As noted in the text and shown in Figure 2B, DAB1 is a
933 segment of an 11-gene operon directly downstream of the carboxysome operon that contains CCM-
934 associated genes. Both DAB1 (**B**) and DAB2 (**C**) “operons” contain two distinct genes that we label DabB
935 and DabA. DabA is annotated as Domain of Unknown Function 2309 (DUF2309, PFAM:PF10070) and
936 appears to be a soluble protein. Approximately one third of dabA is distantly homologous to a type II β -
937 CA. CA-like regions are marked with a line, and the four residues expected to be involved in binding the
938 catalytic zinc ion are marked by asterisks. The height of the asterisks has been varied to make them
939 distinguishable despite proximity in sequence space. DabB is homologous to a cation transporter in the
940 same family as the H⁺ pumping subunits of respiratory complex I (PFAM:PF00361). The DAB1 operon
941 also contains a protein of unknown function between DabA1 and DabB1. This protein has distant
942 homology to DabA1 but is truncated to half the length. Vertical bars above the genes indicate percent
943 conservation of that particular amino acid position in a multiple sequence alignment (Methods). Active site
944 residues are in red. All active site residues are highly conserved with percent identities of greater than
945 99%. One active site cysteine and the active site aspartate residue are the two most conserved residues
946 in DabA with 99.9% identity each.

947

948

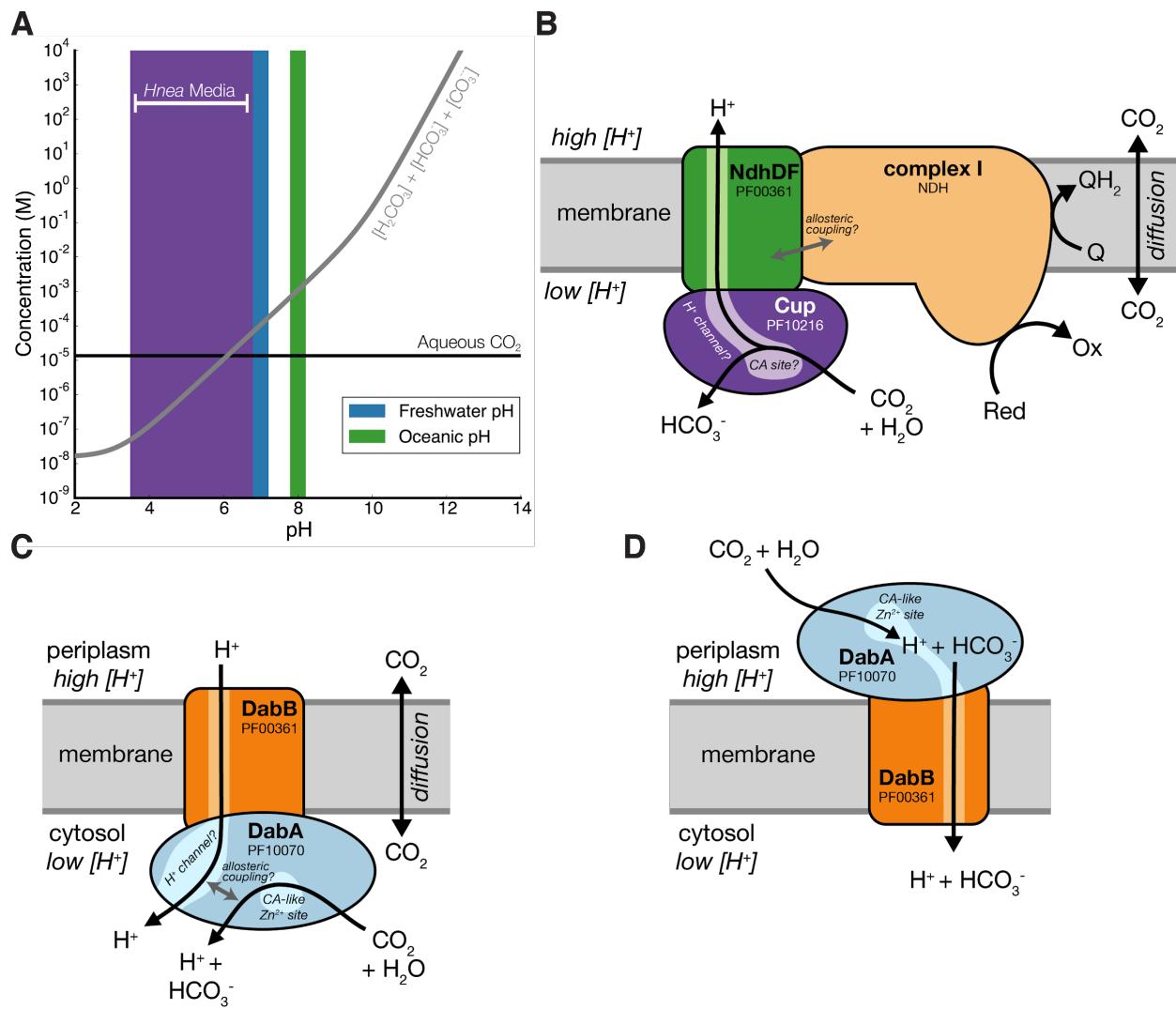


951 **Figure S4. Expression of DabAB2 rescues growth of CAfree *E. coli* in ambient CO₂. A.** These
 952 growth curves were used to generate the growth yield values in Figure 3B. Mean OD₆₀₀ is plotted +/-
 953 standard error for four replicate cultures. Wild-type *E. coli* (BW25113) and CAfree strains expressing
 954 either dabAB2 or human carbonic anhydrase II (hCA) grow in ambient CO₂ while CAfree expressing GFP,
 955 dabB2 alone, or dabA2 alone fail to grow. **B.** These growth curves were used to generate the growth yield
 956 values in Figure 4B. Mean OD₆₀₀ is plotted +/- standard error of four replicate cultures. Wild type cells
 957 and CAfree expressing either DabAB2 or human carbonic anhydrase II (hCA) grow robustly. CAfree cells
 958 expressing putative active site mutants of DabAB2 (C351, D353, H524, or C539) grow as poorly as the
 959 negative control – CAfree expressing superfolder GFP in the same plasmid backbone.



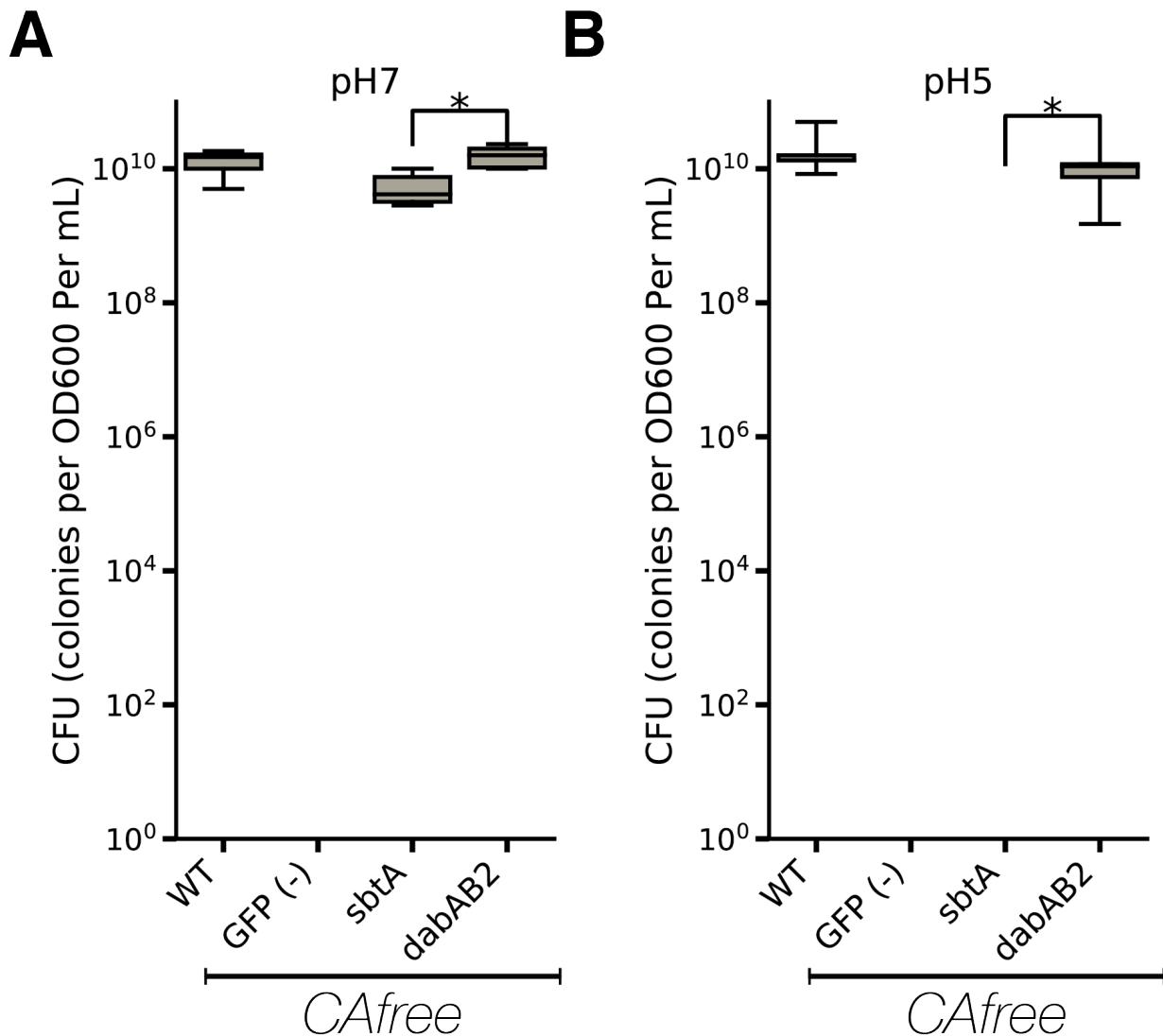
961
962 **Figure S5. DAB2 function is not dependant on complex I.** A. DAB2 is still able to rescue growth of
963 CAfree cells in the absence of Complex I ($\Delta(nuoA-nuoN)$). Error bars represent standard deviation of six
964 replicate cultures. “**” denotes that means differ with bonferroni corrected $P < 0.05$ by a two-tailed T-test,
965 and “***” denotes $P < 5 \times 10^{-4}$. B. These growth curves were used to generate the growth yield values in

966 Figure S5A. Mean OD600 is plotted +/- standard error of six replicate cultures. All strains are Complex I
 967 knockout strains. DAB2 is still able to rescue growth of CAfree cells in the absence of Complex I.
 968

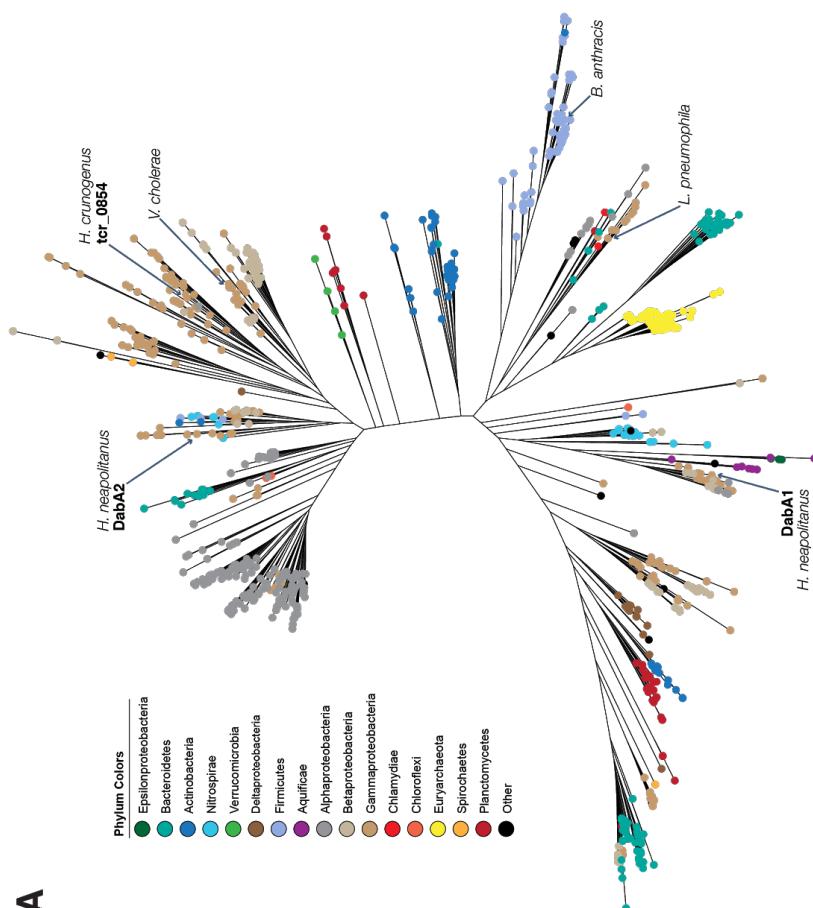


969
 970 **Figure S6. Comparison of models of vectorial CA activity for DABs and the Cyanobacterial CUP**
 971 **systems.** **A.** Equilibrium concentrations of dissolved inorganic carbon as a function of pH. In this plot we
 972 assume the growth medium is in Henry's law equilibrium with present-day atmosphere (400 PPM CO_2) at
 973 25 °C giving a soluble CO_2 concentration of roughly 15 μM . The equilibrium concentrations of hydrated C_i
 974 species (H_2CO_3 , HCO_3^- , CO_3^{2-}) is determined by the pH. As such, the organisms will "see" a C_i species in
 975 very different ratios depending on the environmental pH. In a oceanic pH near 8, HCO_3^- dominates the C_i
 976 pool. HCO_3^- is also the dominant constituent of the C_i pool in freshwater, but less so (by a factor of ~10
 977 since freshwater and oceanic environments differ by about 1 pH unit). In acid conditions (pH < 6.1) CO_2
 978 will be the dominant constituent of the C_i pool. The pH of our *Hnea* culture media ranges from 6.8 (when
 979 freshly made) to ~3.5 when cells reach stationary phase (*Hnea* make H_2SO_4 as a product of their sulfur
 980 oxidizing metabolism). As such we expect that *Hnea* regularly experiences environments wherein it is
 981 advantageous to pump CO_2 and not HCO_3^- . **B.** CupA/B proteins are CA-like subunits of a class of

982 cyanobacterial Ci uptake systems. Cup-type systems are believed to couple electron transfer to vectorial
983 CA activity and, potentially, outward-directed proton pumping. This model is based on the observation
984 that Cup systems displace the two distal H⁺-pumping subunits of the cyanobacterial complex I and
985 replace them with related subunits that bind CupA/B (illustrated in green as NdhD/F). **C**. As our data are
986 consistent with DAB2 functioning as a standalone complex (i.e. DabAB do not appear to bind or require
987 the *E. coli* complex I), we propose a different model for DAB function where energy for unidirectional
988 hydration of CO₂ is drawn from the movement of cations along their electrochemical gradient (right panel
989 above). **D**. An alternative model for DAB activity is that DabA is localized to the periplasm and DabB is
990 functioning as a H⁺ : HCO₃⁻ symporter. In this model DabA CA activity is made vectorial by removal of
991 products. Energy is provided in the form of the PMF driving H⁺ (and therefore HCO₃⁻) uptake. This model
992 is not preferred because no secretion signals were observed in the DabA sequence. Moreover, the
993 *Acidimicrobium ferrooxidans* genome contains an apparent DabA:DabB fusion protein. The predicted
994 architecture the fusion would place DabA in the cytoplasm.

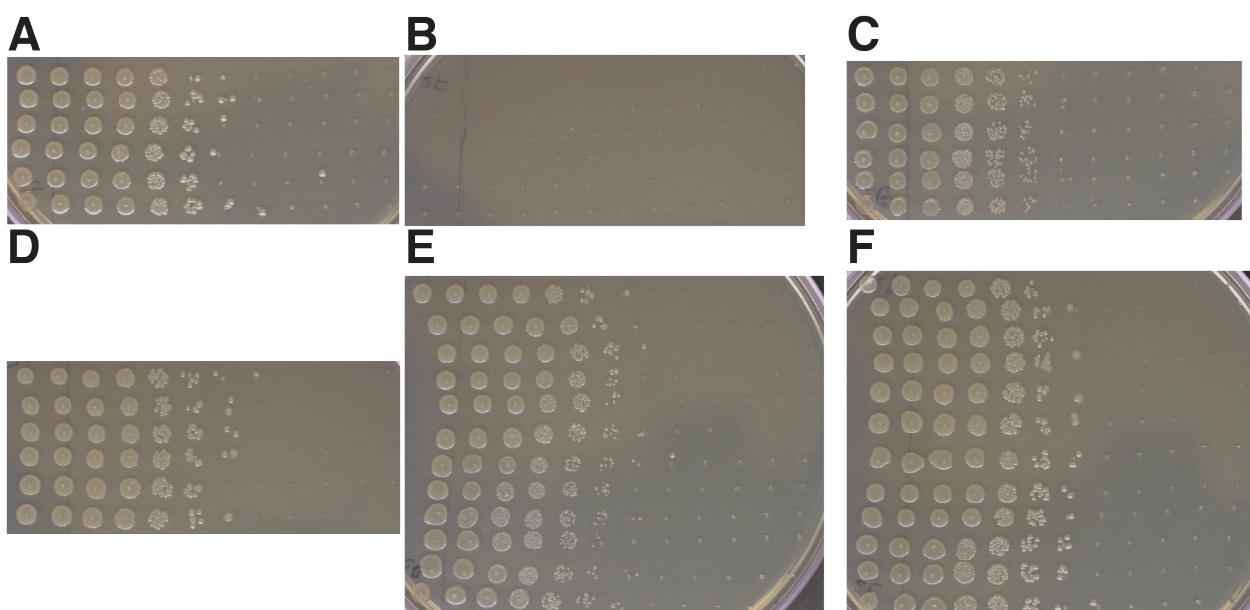


995
996 **pH independence of *dabAB2* rescue of CAfree** Colony forming units per OD600 per ml were measured
997 on LB plates with induction in air at both pH 7 (A.) and 5 (B.). *dabAB2* rescued growth at both pH 7 and
998 pH 5, *sbtA* only rescued growth at pH 7. Whiskers represent the range of the data, the box represents the
999 interquartile range, and the middle line represents the median. Data is from 6 replicate platings of all
1000 conditions. In all panels “*” denotes that means differ with bonferroni corrected P < 0.05 by a two-tailed T-
1001 test.

B**A**

1003 **Figure S8. Fully annotated approximate maximum likelihood phylogenetic trees of DabA.** **A.** A
1004 phylogenetic tree emphasizing the clades containing high-confidence DabA homologs. DabA homologs
1005 are found in > 15 prokaryotic clades, including some archaea. *Hnea* DabA1 and DabA2 represent two
1006 different groupings that are commonly found in proteobacteria. The tcr_0854 gene of *H. crunogenus* is
1007 more closely related to DabA2 than DabA1. Inspecting the tree reveals several likely incidents of
1008 horizontal transfer, e.g. between proteobacteria and Firmicutes, Nitrospirae and Actinobacteria.
1009 Moreover, the genomes of several known pathogens contain a high-confidence DabA homolog, including
1010 *B. anthracis*, *L. pneumophila*, *V. cholerae*. **B.** Association of various Rubisco isoforms with DabA
1011 homologs. Many organisms that have DabA also have a Rubisco. However, there are numerous
1012 examples of DabA homologs that are found in genomes with no Rubisco (denoted by leaves with no
1013 colored marking), suggesting that this uptake system might play a role in heterotrophic metabolism. DabA
1014 is most-frequently associated with Form I Rubiscos (red and purple leaves in panel B), which is sensible
1015 because all known bacterial CCMs involve a Form I Rubisco exclusively. Some DabA-bearing genomes
1016 have only a Form II Rubisco (blue) and the Euryarchaeota genomes have that DabA have a Form III
1017 Rubisco (green) or none at all.

1018



1019
1020 **Figure S9. Plates used for determining CFU counts for Figure 5B.** **A.** Wt positive control. **B.** CAfree
1021 sfGFP negative control does not rescue. **C.** CAfree hCA positive control rescues growth. **D.** CAfree DAB2
1022 rescues growth. **E.** baDAB from *Bacillus anthracis* rescues growth of CAfree. **F.** vcDAB from *Vibrio*
1023 *cholera* rescues growth of CAfree. Panels **A-D** represent 6 technical replicates of the plating. Panels **E**
1024 and **F** represent 6 technical replicates each of 2 biological replicates. In all panels, the first spot

1025 represents 3ul of an OD 0.2 culture grown at 10% CO₂ each subsequent spot is 3 ul of a 1:10 dilution of

1026 the previous spot.

1027