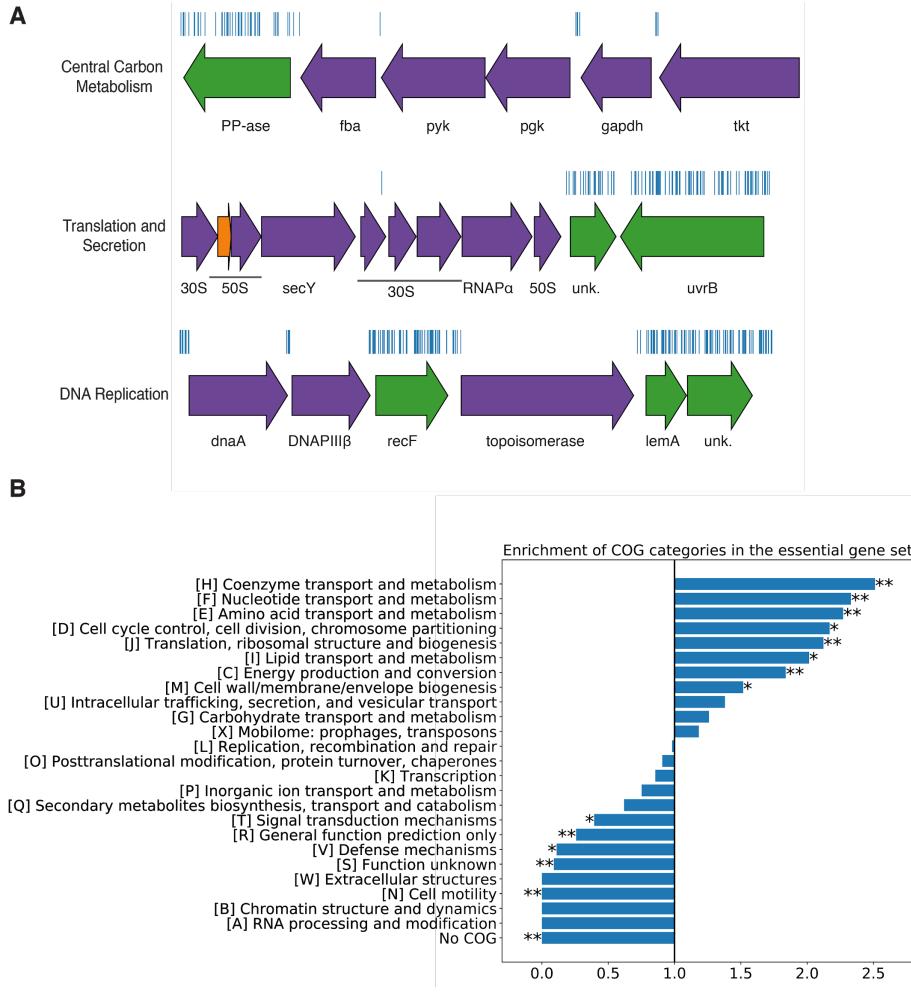


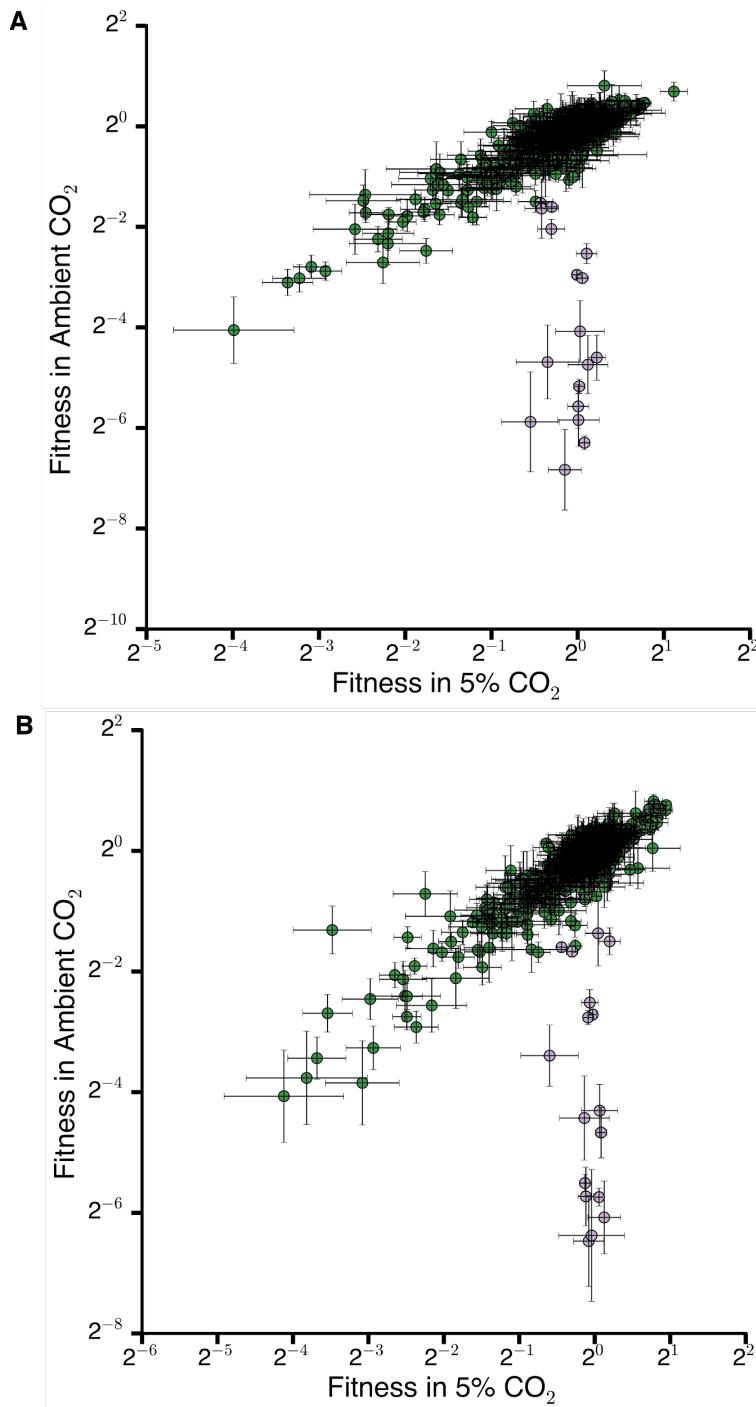
844 **Supplementary Information:**

- 845 **Supplemental File 1.** Important strains and reagents.
- 846 **Supplemental File 2.** Transposon insertion information and essentiality determination by gene.
- 847 **Supplemental File 3.** Fitness effects and HCR phenotype by gene.
- 848 **Supplemental File 4.** Genes used to generate figure S4A.
- 849 **Supplemental File 5.** Genes used to generate figure 5A.



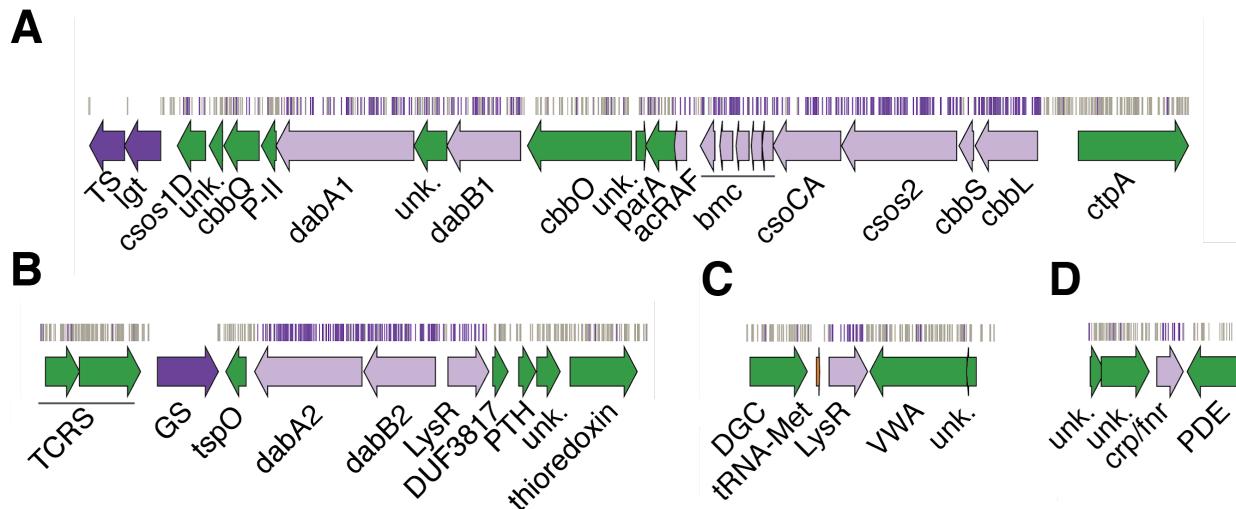
850
 851 **Figure S1 The essential gene set is enriched for COGs associated with essential cellular
 852 processes.** **A.** Representative essential genes and nonessential genes in the *Hnea* genome. The blue
 853 track indicates the presence of an insertion. Genes in purple were called essential and genes in green are
 854 nonessential. Genes labeled "unk." are hypothetical proteins. The first genomic locus contains 5 essential
 855 genes involved in glycolysis or the CBB cycle including pyruvate kinase (pyk) and transketolase (tkt). The
 856 8 essential genes in the second locus encode 30S and 50S subunits of the ribosome, the secY secretory
 857 channel, and an RNA polymerase subunit. Essential genes in the third example locus include
 858 topoisomerase and DNA polymerase III β . **B.** COG enrichments were calculated by dividing the fraction of
 859 genes in the essential gene set associated with this COG category by the fraction of genes in the genome
 860 associated with this category. ** denotes that this COG is enriched (or depleted) with Bonferroni
 861 corrected $p < 0.05$ by a hypergeometric test, and *** denotes $p < 5 \times 10^{-4}$. Exact p values are as follows
 862 for each category, No COG: 7×10^{-68} , C: 1.1×10^{-5} , D: 1×10^{-2} , E: $< 7 \times 10^{-68}$, F: 4.3×10^{-6} , H: $< 7 \times 10^{-68}$, I: 6.3×10^{-4} ,
 863 J: $< 7 \times 10^{-68}$, M: 6.9×10^{-3} , N: 6.5×10^{-6} , R: 6.17×10^{-5} , S: 7.7×10^{-8} , T: 2.8×10^{-2} , V: 1.1×10^{-2} . In panel A, the following
 864 abbreviations are used: exopolyphosphatase (PP-ase), fructose-bisphosphate aldolase class II (fba),
 865 pyruvate kinase (pyk), phosphoglycerate kinase (pgk), type I glyceraldehyde-3-phosphate dehydrogenase
 866 (gapdh), transketolase (tkt), 30S ribosomal protein (30S), 50S ribosomal protein (50S), preprotein

867 translocase subunit SecY (SecY), DNA-directed RNA polymerase subunit alpha (RNAP α), hypothetical
 868 protein (unk.), excinuclease ABC subunit UvrB (UvrB), chromosomal replication initiator protein dnaA
 869 (dnaA), DNA polymerase III subunit beta (DNAPIII β), DNA replication and repair protein recF (recF), DNA
 870 topoisomerase (ATP-hydrolyzing) subunit B (topoisomerase), lemA family protein (LemA).



871
 872 **Figure S2 Gene fitnesses measurements for each replicates.** Fitness effects of gene knockouts in
 873 $5\% \text{CO}_2$ as compared to ambient CO_2 . The effects of single transposon insertions into a gene are

874 averaged to produce the gene-level fitness value plotted. Error bars represent one standard error of the
 875 mean. We define HCR mutants as those displaying a twofold fitness defect in ambient CO₂ relative to 5%
 876 CO₂. HCR genes are colored light purple. Panel **A** contains data from the first replicate experiment and
 877 panel **B** contains data from the second replicate experiment.



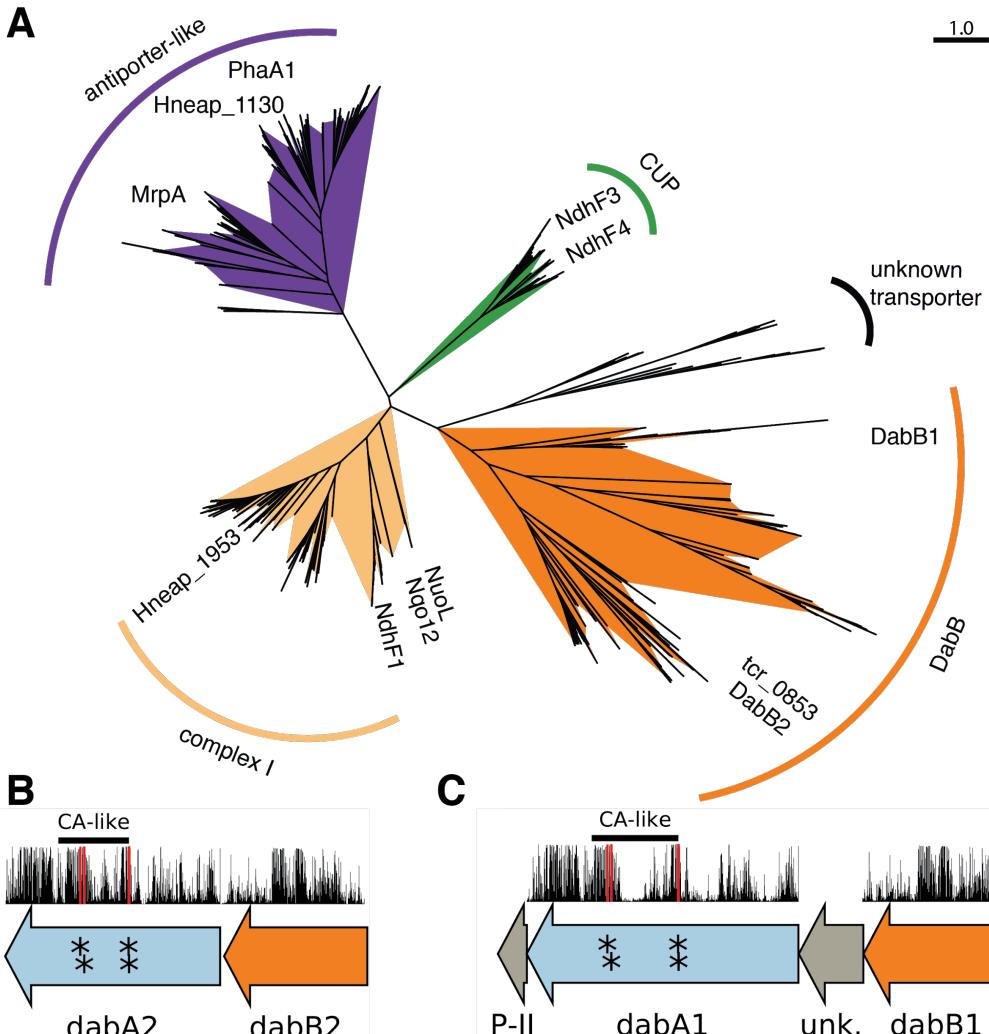
878

879 **Figure S3 Genomic context of *Hneaa* HCR genes identified in our genome-wide screen.** Panels **A-D**
 880 show regions of the *Hneaa* genome containing genes annotated as HCR. Essential genes are in dark
 881 purple, HCR genes are in light purple, and other genes are in green. The top tracks show the presence of
 882 an insertion in that location. Insertions are colored grey unless they display a twofold or greater
 883 fitness defect in ambient CO₂, in which case they are colored purple. **A.** The gene cluster containing the
 884 carboxysome operon (HNEAP_RS04660-HNEAP_RS04620) and a second CCM-associated operon. This
 885 second operon contains acRAF (HNEAP_RS04615), a FormIC associated cbbOQ-type Rubisco activase
 886 (HNEAP_RS04575 and HNEAP_RS04600), parA (HNEAP_RS04610), P-II (HNEAP_RS04580) and
 887 dabAB1 (dabA1: HNEAP_RS04585 and dabB1: HNEAP_RS04620). **B.** The DAB2 operon and
 888 surrounding genomic context (lysR: HNEAP_RS01040, dabA2: HNEAP_RS01030, and dabB2:
 889 HNEAP_RS01035). **C.** The genomic context of a lysR-type transcriptional regulator (HNEAP_RS05490)
 890 that shows an HCR phenotype. **D** Genomic context of a crp/fnr-type transcriptional regulator that displays
 891 an HCR phenotype (HNEAP_RS07320). Accession numbers and gi numbers for selected genes can be
 892 found in Table S1. Abbreviations for Figure S3: thymidylate synthase (TS), prolipoprotein diacylglycerol
 893 transferase (lgt), Rubisco activase Rubisco activase subunits (cbbOQ), nitrogen regulatory protein P-II (P-
 894 II), ParA family protein (parA), csos1CAB and csos4AB (bmc), copper-translocating P-type ATPase
 895 (ctpA), DNA-binding response regulator and two-component sensor histidine kinase (TCRS), glutamate--
 896 ammonia ligase (GS), tryptophan-rich sensory protein (tspO), DUF3817 domain-containing protein
 897 (DUF3817), aminoacyl-tRNA hydrolase (PTH), thioredoxin domain-containing protein (thioredoxin),
 898 sensor domain-containing diguanylate cyclase (DGC), methionine tRNA (tRNA-Met), VWA domain-
 899 containing protein (VWA), diguanylate phosphodiesterase (PDE).

Locus Id	NCBI Accession number	NCBI gi number	Gene description	Has HCR phenotype
HNEAP_RS01030	WP_012823110.1	502585319	DabA2	TRUE
HNEAP_RS01035	WP_012823111.1	502585320	DabB2	TRUE
HNEAP_RS01040	WP_012823112.1	502585321	LysR	TRUE
HNEAP_RS04565	WP_012823782.1	502586009	Csos1D	FALSE
HNEAP_RS04570	WP_012823783.1	502586011	unk.	FALSE
HNEAP_RS04575	WP_012823784.1	502586012	CbbQ	FALSE
HNEAP_RS04580	WP_012823785.1	502586013	p-II	FALSE
HNEAP_RS04585	WP_012823786.1	502586014	DabA1	TRUE
HNEAP_RS04590	WP_012823787.1	502586015	unk.	FALSE
HNEAP_RS04595	WP_012823788.1	502586016	DabB1	TRUE
HNEAP_RS04600	WP_012823789.1	502586017	CbbO	FALSE
HNEAP_RS04605	WP_041600361.1	753844744	unk.	FALSE
HNEAP_RS04610	WP_049772467.1	908628434	ParA	FALSE
HNEAP_RS04615	WP_012823792.1	502586020	acRAF	TRUE
HNEAP_RS04620	WP_012823793.1	502586021	Csos1B	TRUE
HNEAP_RS04625	WP_012823794.1	502586022	Csos1A	TRUE
HNEAP_RS04630	WP_012823795.1	502586023	Csos1C	TRUE
HNEAP_RS04635	WP_012823796.1	502586024	Csos4B	TRUE
HNEAP_RS04640	WP_012823797.1	502586025	Csos4A	TRUE
HNEAP_RS04645	WP_012823798.1	502586026	CsosCA	TRUE
HNEAP_RS04650	WP_081441107.1	1174219926	Csos2	TRUE
HNEAP_RS04655	WP_012823800.1	502586028	CbbS	TRUE
HNEAP_RS04660	WP_012823801.1	502586029	CbbL	TRUE
HNEAP_RS05490	WP_012823963.1	502586200	LysR	TRUE
HNEAP_RS07320	WP_081441122.1	1174219941	Crp/Fnr	TRUE

900
901
902

Table S1 Genes from HCR operons. This table includes genes from the the HCR operons with their phenotype and identifying information. "unk." indicates a hypothetical protein.



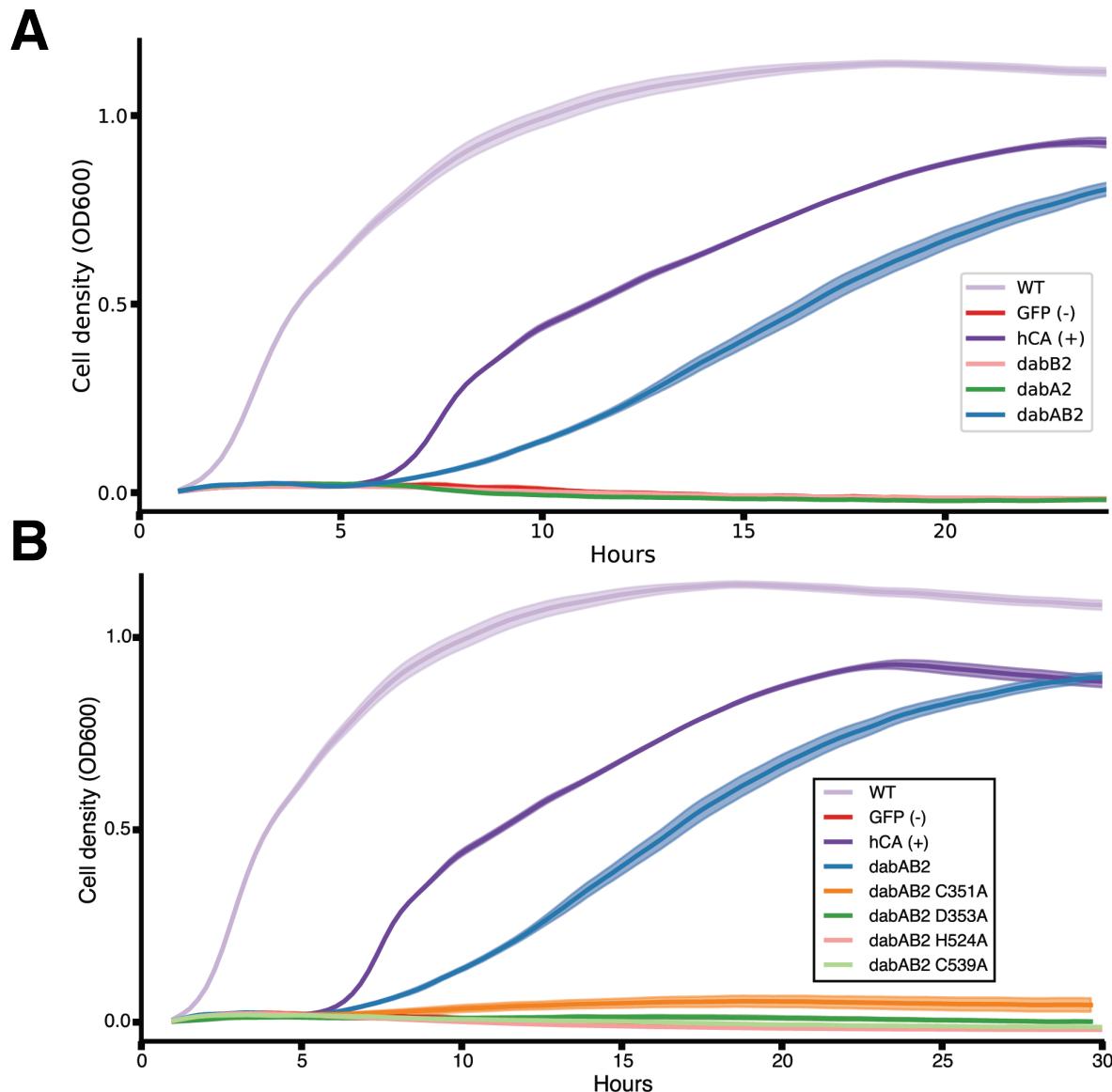
903
904 **Figure S4 PF0361 contains multiple subfamilies, but some regions of DAB subunits are highly**
905 **conserved.** A. PF0361 is a large and diverse protein family containing multiple subgroups with different
906 documented activities. These subfamilies include Mrp-family cation antiporters, proton translocating
907 subunits of complex I, membrane subunits of CUP (CO₂ uptake protein) complexes, and DabB proteins.
908 These subfamilies are highly diverged and perform a variety of activities. This means that it is not
909 possible to draw conclusions about the mechanism of DAB complexes just from their homology to
910 PF0361. This panel contains an approximate maximum likelihood tree of PF0361 genes. Clades were
911 colored according to the presence of genes with known functions. The purple clade contains the *Bacillus*
912 *subtilis* and *Staphylococcus aureus* MrpA cation antiporter subunits and the *Sinorhizobium meliloti*
913 antiporter PhaA1. The light orange clade contains the known cation translocating subunits of complex I:
914 nuoL from *Escherichia coli*, Nqo12 from *Thermus thermophilus*, and NdhF1 from both *Synechococcus*
915 *elongatus* PCC7942 and *Thermosynechococcus elongatus* BP-1. The green clade contains CUP-
916 associated membrane subunits ndhF3 from both *Synechococcus elongatus* PCC7942 and
917 *Thermosynechococcus elongatus* BP-1 and ndhF4 from the same two species. The dark orange
918 clade includes DabB1-2 and tcr_0853 from *Thiomicrospira crunogena*. We note that the clade containing
919 DabB1-2 is distinct from that containing known complex I subunits or to mrp-family antiporters. This tree is

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

938

939

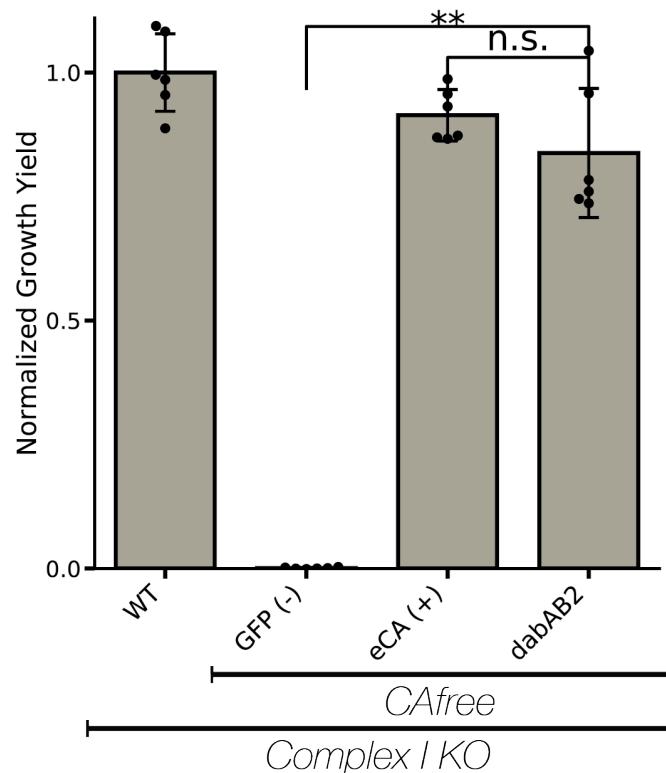
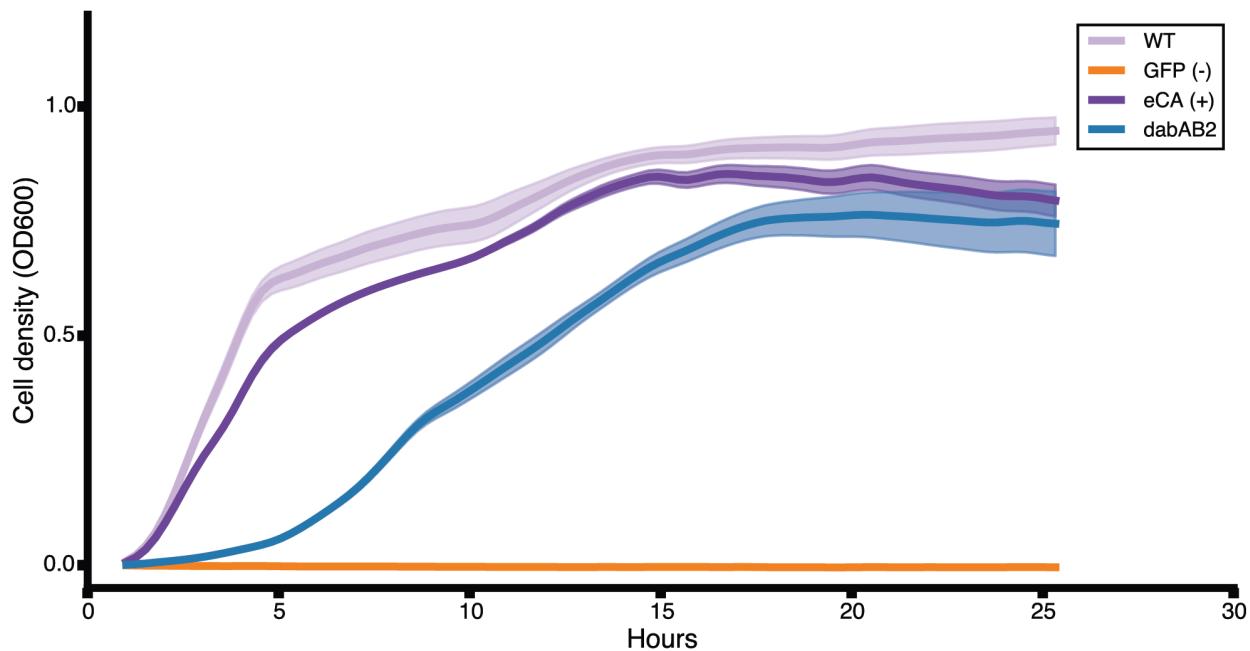
940



941

942 **Figure S5. Expression of DabAB2 rescues growth of CAfree *E. coli* in ambient CO₂. A.** These
 943 growth curves were used to generate the growth yield values in Figure 3B. Mean OD₆₀₀ is plotted +/-
 944 standard error for four replicate cultures. Wild-type *E. coli* (BW25113) and CAfree strains expressing
 945 either dabAB2 or human carbonic anhydrase II (hCA) grow in ambient CO₂ while CAfree expressing GFP,
 946 dabB2 alone, or dabA2 alone fail to grow. **B.** These growth curves were used to generate the growth yield
 947 values in Figure 4B. Mean OD₆₀₀ is plotted +/- standard error of four replicate cultures. Wild type cells
 948 and CAfree expressing either DabAB2 or human carbonic anhydrase II (hCA) grow robustly. CAfree cells
 949 expressing putative active site mutants of DabAB2 (C351, D353, H524, or C539) grow as poorly as the
 950 negative control – CAfree expressing superfolder GFP in the same plasmid backbone.

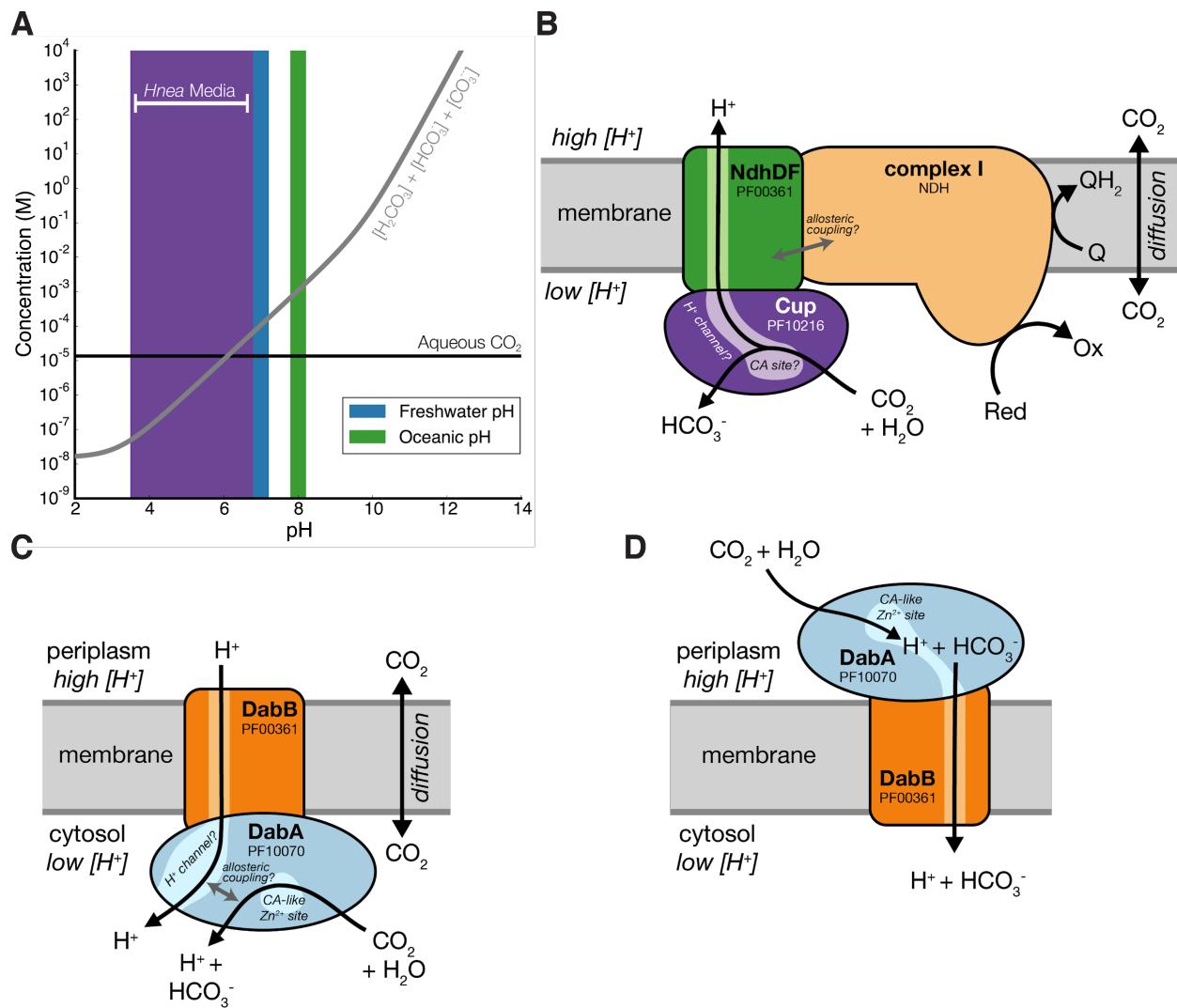
951

A**B**

952

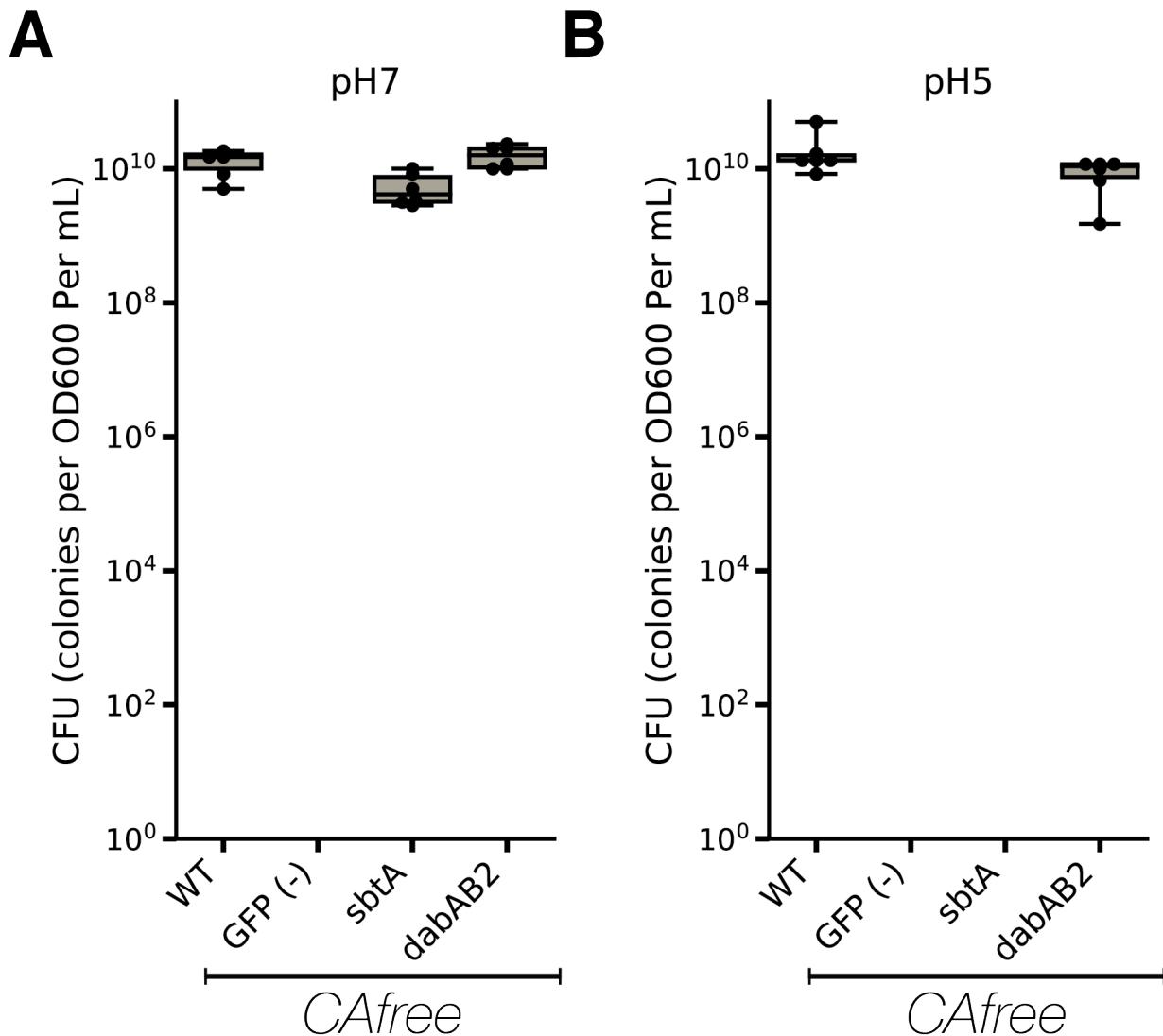
953 **Figure S6. DAB2 function is not dependant on complex 1.** A. DAB2 is still able to rescue growth of
954 CAfree cells in the absence of Complex I ($\Delta(nuoA-nuoN)$). dabAB2 rescues better than GFP ($t=15.7$,
955 $p=2.37 \times 10^{-8}$). Error bars represent standard deviation of six replicate cultures. "n.s." denotes means do
956 not differ significantly, "*" denotes that means differ with bonferroni corrected $p < 0.05$ by a two-tailed t-
957 test, and "***" denotes $p < 5 \times 10^{-4}$. B. These growth curves were used to generate the growth yield values

958 in Figure S6A. Mean OD₆₀₀ is plotted +/- standard error of six replicate cultures. All strains are Complex I
 959 knockout strains. DAB2 is still able to rescue growth of CAfree cells in the absence of Complex I.
 960



961
 962 **Figure S7. Comparison of models of vectorial CA activity for DABs and the Cyanobacterial CUP**
 963 **systems.** A. Equilibrium concentrations of dissolved inorganic carbon as a function of pH. In this plot we
 964 assume the growth medium is in Henry's law equilibrium with present-day atmosphere (400 PPM CO_2) at
 965 25 °C giving a soluble CO_2 concentration of roughly 15 μM . The equilibrium concentrations of hydrated C_i
 966 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
 967 very different ratios depending on the environmental pH. In a oceanic pH near 8, HCO_3^- dominates the C_i
 968 pool. HCO_3^- is also the dominant constituent of the C_i pool in freshwater, but less so (by a factor of ~10
 969 since freshwater and oceanic environments differ by about 1 pH unit). In acid conditions (pH < 6.1) CO_2
 970 will be the dominant constituent of the C_i pool. The pH of our *Hnea* culture media ranges from 6.8 (when
 971 freshly made) to ~3.5 when cells reach stationary phase (*Hnea* make H_2SO_4 as a product of their sulfur
 972 oxidizing metabolism). As such we expect that *Hnea* regularly experiences environments wherein it is

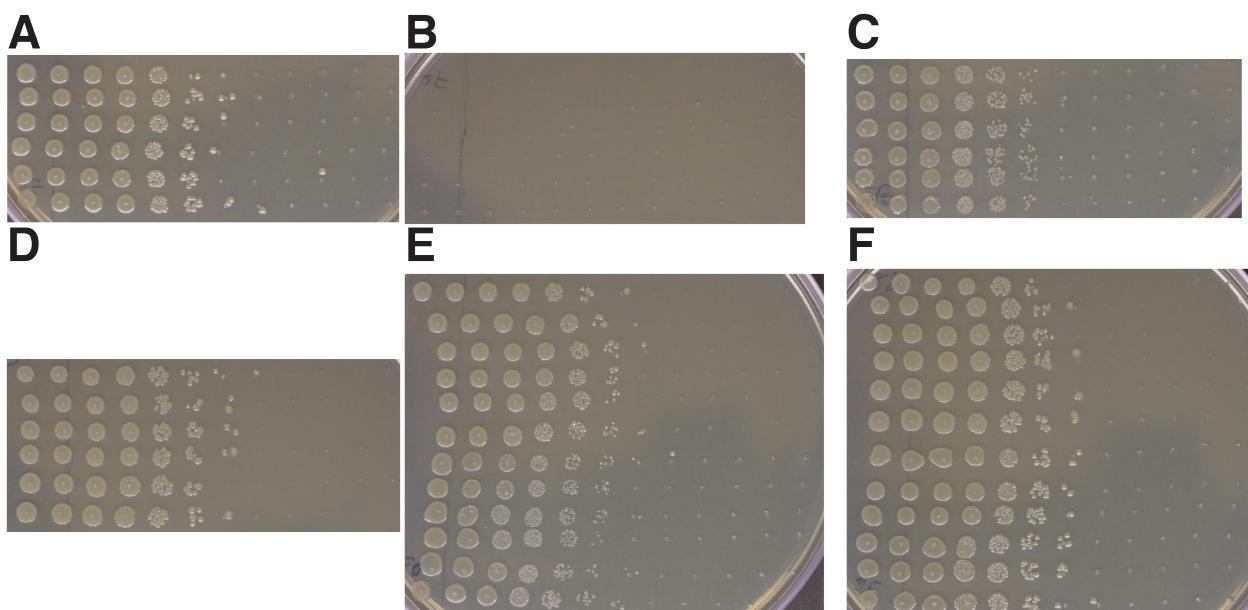
973 advantageous to pump CO₂ and not HCO₃⁻. **B.** CupA/B proteins are CA-like subunits of a class of
974 cyanobacterial Ci uptake systems. Cup-type systems are believed to couple electron transfer to vectorial
975 CA activity and, potentially, outward-directed proton pumping. This model is based on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



987
988
989
990
991
992

Figure S8. pH independence of *dabAB2* rescue of CAfree Colony forming units per OD600 per ml were measured on LB plates with induction in air at both pH 7 (A.) and 5 (B.). *dabAB2* rescued growth at both pH7 and pH 5, *sbtA* only rescued growth at pH 7. Whiskers represent the range of the data, the box represents the interquartile range, and the middle line represents the median. Data is from 6 replicate platings of all conditions.

994 **Figure S9. Fully annotated approximate maximum likelihood phylogenetic trees of DabA.** **A.** A
995 phylogenetic tree emphasizing the clades containing high-confidence DabA homologs. DabA homologs
996 are found in > 15 prokaryotic clades, including some archaea. *Hnea* DabA1 and DabA2 represent two
997 different groupings that are commonly found in proteobacteria. The tcr_0854 gene of *H. crunogenus* is
998 more closely related to DabA2 than DabA1. Inspecting the tree reveals several likely incidents of
999 horizontal transfer, e.g. between proteobacteria and Firmicutes, Nitrospirae and Actinobacteria.
1000 Moreover, the genomes of several known pathogens contain a high-confidence DabA homolog, including
1001 *B. anthracis*, *L. pneumophila*, *V. cholerae*. **B.** Association of various Rubisco isoforms with DabA
1002 homologs. Many organisms that have DabA also have a Rubisco. However, there are numerous
1003 examples of DabA homologs that are found in genomes with no Rubisco (denoted by leaves with no
1004 colored marking), suggesting that this uptake system might play a role in heterotrophic metabolism. DabA
1005 is most-frequently associated with Form I Rubiscos (red and purple leaves in panel B), which is sensible
1006 because all known bacterial CCMs involve a Form I Rubisco exclusively. Some DabA-bearing genomes
1007 have only a Form II Rubisco (blue) and the Euryarchaeota genomes have that DabA have a Form III
1008 Rubisco (green) or none at all. For both panels, scale bars indicate one substitution per site.
1009



1010
1011 **Figure S10. Plates used for determining CFU counts for Figure 5B.** **A.** Wt positive control. **B.** CAfree
1012 sfGFP negative control does not rescue. **C.** CAfree hCA positive control rescues growth. **D.** CAfree DAB2
1013 rescues growth. **E.** baDAB from *Bacillus anthracis* rescues growth of CAfree. **F.** vcDAB from *Vibrio*
1014 *cholera* rescues growth of CAfree. Panels **A-D** represent 6 technical replicates of the plating. Panels **E**
1015 and **F** represent 6 technical replicates each of 2 biological replicates. In all panels, the first spot
1016 represents 3 ul of an OD 0.2 culture grown at 10% CO₂ each subsequent spot is 3 ul of a 1:10 dilution of
1017 the previous spot.
1018