Supplementary Information:

Supplemental File 1. Important strains and reagents.

Supplemental File 2. Transposon insertion information and essentiality determination by gene.

Supplemental File 3. Fitness effects and HCR phenotype by gene.

Supplemental File 4. Genes used to generate figure S3A.

Supplemental File 5. Genes used to generate figure 5A.

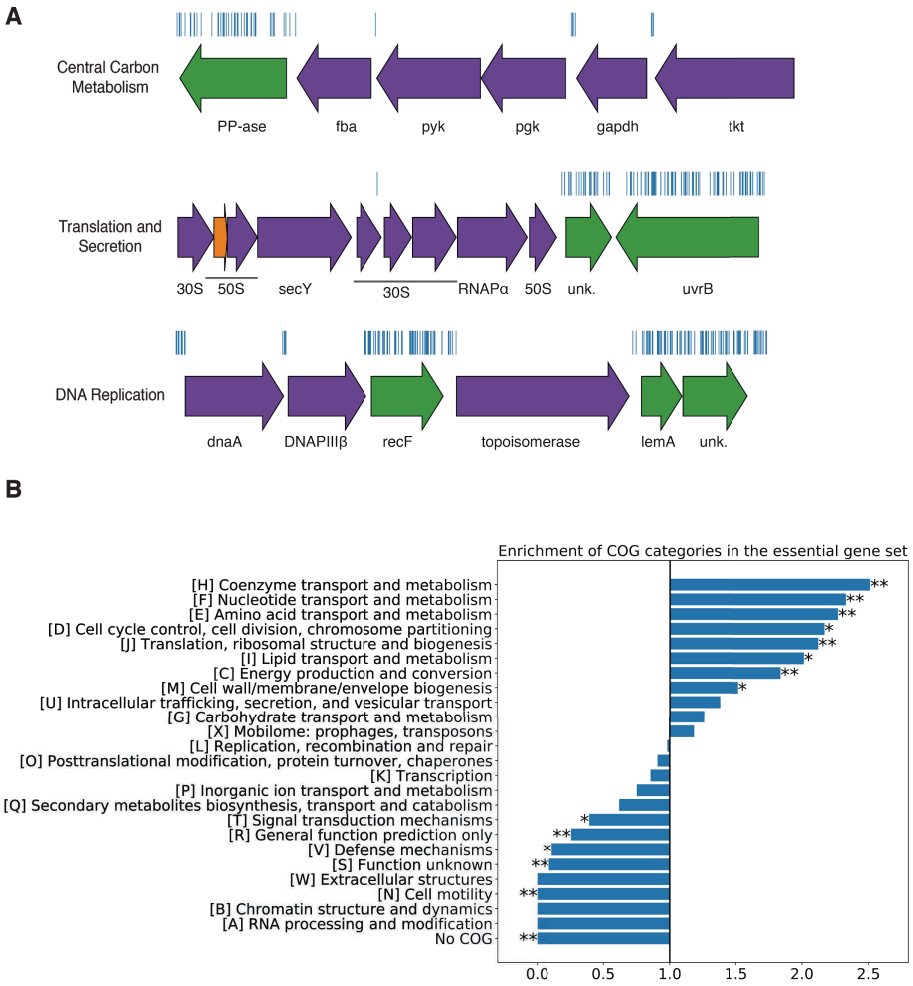

Figure S1 The essential gene set is enriched for COGs associated with essential cellular processes. A. Representative essential genes and nonessential genes in the *Hnea* genome. The blue track indicates the presence of an insertion. Genes in purple were called essential and genes in green are nonessential. 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 encodi 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. **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 $P < 5X10^{-4}$. Exact p values are as follows for each category, No COG:$7*10^{-68}$, C:$1.1*10^{-5}$, D:$1*10^{-2}$, E:$<7*10^{-68}$, F:$4.3*10^{-6}$, H:$<7*10^{-68}$, I:$6.3*10^{-4}$, J:$<7*10^{-68}$, 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 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

translocase subunit SecY (SecY), DNA-directed RNA polymerase subunit alpha (RNAPα), hypothetical protein (unk.), excinuclease ABC subunit UvrB (UvrB), chromosomal replication initiator protein dnaA (dnaA), DNA polymerase III subunit beta (DNAPIIIβ), DNA replication and repair protein recF (recF), DNA topoisomerase (ATP-hydrolyzing) subunit B (topoisomerase), lemA family protein (LemA).

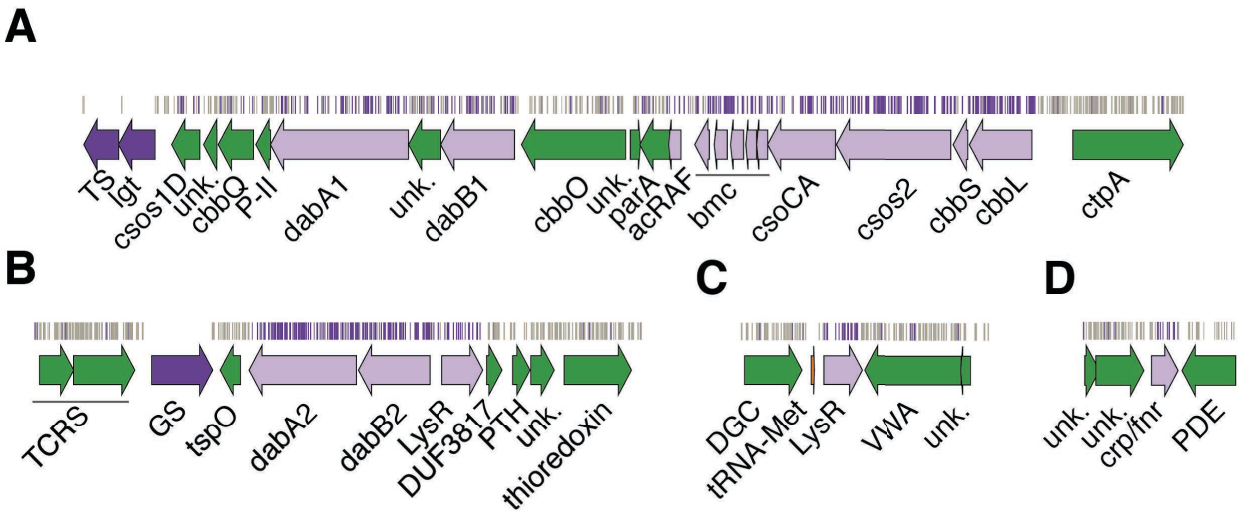

Figure S2 Genomic context of *Hnea* HCR genes identified in our genome-wide screen. Panels **A-D** show regions of the *Hnea* genome containing genes annotated as HCR. Essential genes are in dark purple, HCR genes are in light purple, and other genes are in green. The top tracks show the presence of an insertion in that location. Insertions are colored colored grey unless they display a twofold or greater fitness defect in ambient CO_2, in which case they are colored purple. **A.** The gene cluster containing the carboxysome operon and a second CCM-associated operon. This second operon contains acRAF, a FormIC associated cbbOQ-type Rubisco activase and dabAB1. **B.** The DAB2 operon and surrounding genomic context. **C.** The genomic context of a lysR-type transcriptional regulator that shows an HCR phenotype. **D** Genomic context of a crp/fnr-type transcriptional regulator that displays an HCR phenotype. Abbreviations for Figure S2: thymidylate synthase (TS), prolipoprotein diacylglyceryl transferase (lgt), Rubisco activase Rubisco activase subunits (cbbOQ), 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).

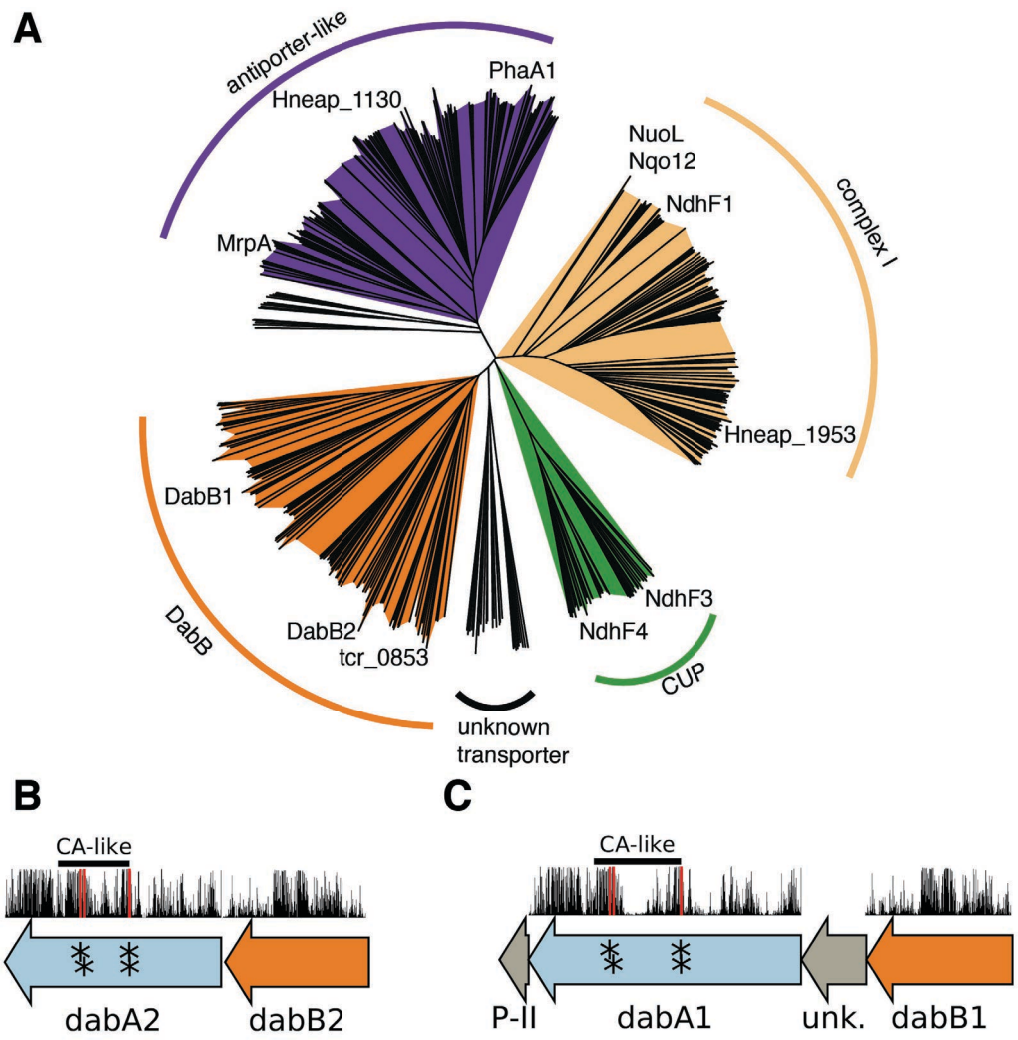

Figure S3 PF0361 contains multiple subfamilies, but some regions of DAB subunits are highly conserved. A. PF0361 is a large and diverse protein family containing multiple subgroups with different documented activities. These subfamilies include Mrp-family cation antiporters, proton translocating subunits of complex I, membrane subunits of CUP (CO_2 uptake protein) complexes, and DabB proteins. These subfamilies are highly diverged and perform a variety of activities. This means that it is not possible to draw conclusions about the mechanism of DAB complexes just from their homology to PF0361. This panel contains a nearest neighbor tree of PF0361 genes. Clades were colored according to the presence of genes with known functions. The purple clade contains the *Bacillus subtilis* and *Staphylococcus aureus* MrpA cation antiporter subunits and the *Sinorhizobium meliloti* antiporter PhaA1. The light orange clade contains the known cation translocating subunits of complex I: nuoL from *Escherichia coli*, Nqo12 from *Thermus thermophilus*, and NdhF1 from both *Synechococcus elongatus* PCC7942 and *Thermosynechococcus elongatus* BP-1. The green clade contains CUP-associated membrane subunits ndhF3 from both *Synechococcus elongatus* PCC7942 and *Thermosynechococcus elongatus* BP-1 and ndhF4 from from the same two species. The dark orange clade includes DabB1-2 and tcr_0853 from *Thiomicrospira crunogena*. We note that the clade containing DabB1-2 is distinct from that containing known complex I subunits or to mrp-family antiporters. This tree is consistent with our model, where DabB

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

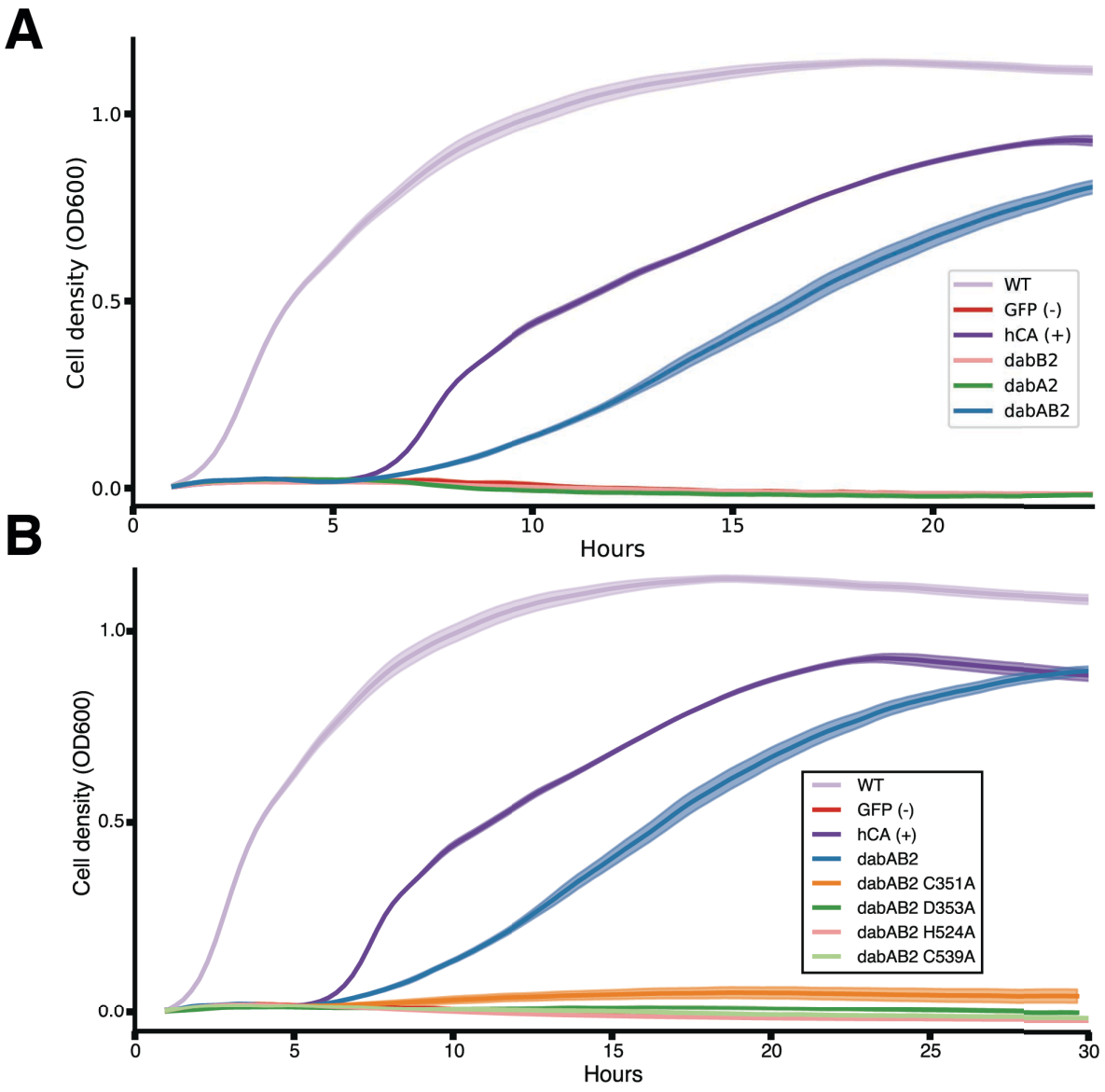

Figure S4. Expression of DabAB2 rescues growth of CAfree *E. coli* in ambient CO_2. A. These growth curves were used to generate the growth yield values in Figure 3B. Mean OD600 is plotted +/- standard error for four replicate cultures. Wild-type *E. coli* (BW25113) and CAfree strains expressing either dabAB2 or human carbonic anhydrase II (hCA) grow in ambient CO_2 while CAfree expressing GFP, dabB2 alone, or dabA2 alone fail to grow. **B.** These growth curves were used to generate the growth yield values in Figure 4B. Mean OD600 is plotted +/- standard error of four replicate cultures. Wild type cells and CAfree expressing either DabAB2 or human carbonic anhydrase II (hCA) grow robustly. CAfree cells expressing putative active site mutants of DabAB2 (C351, D353, H524, or C539) grow as poorly as the negative control – CAfree expressing superfolder GFP in the same plasmid backbone.

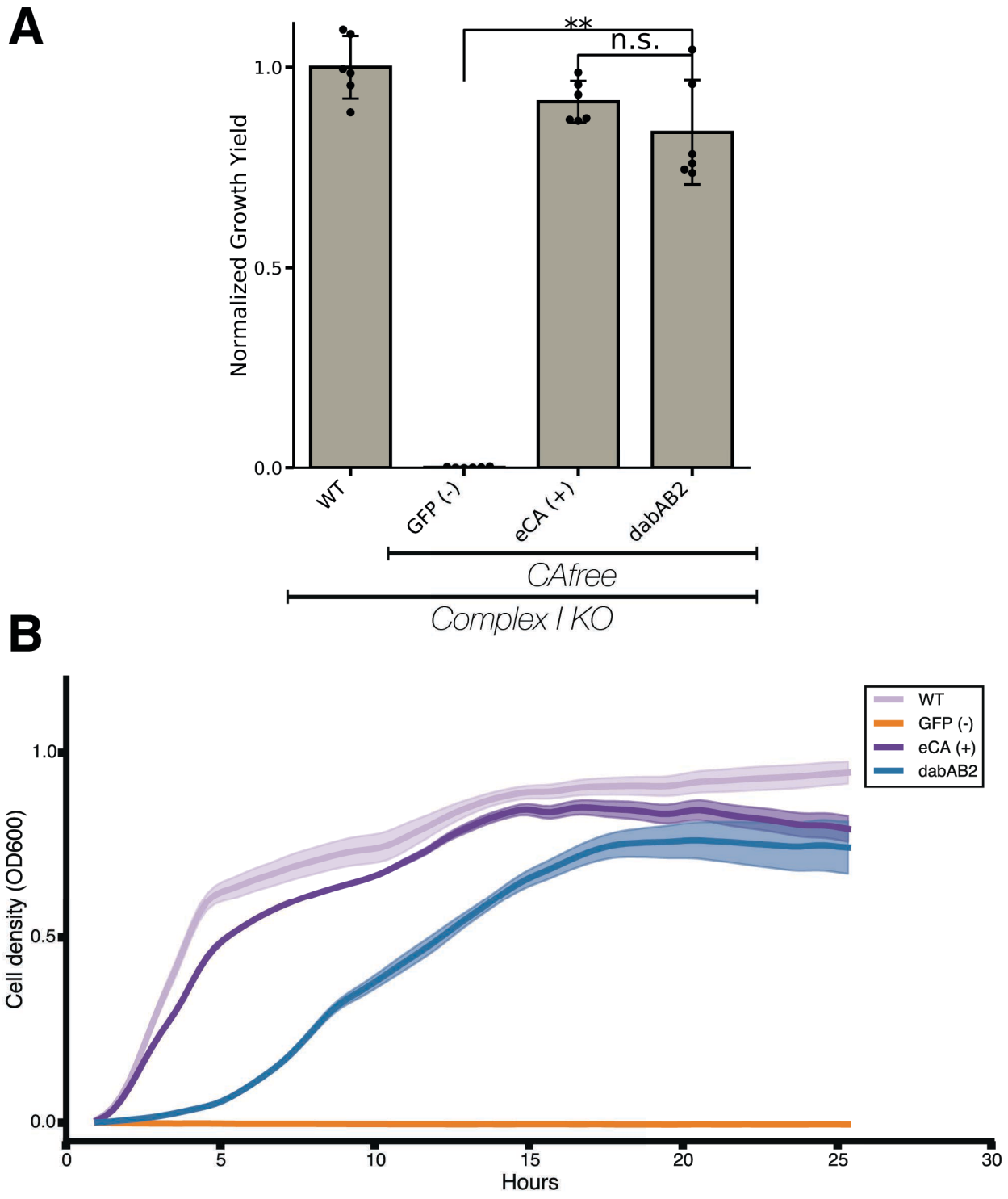

Figure S5. DAB2 function is not dependant on complex 1. A. DAB2 is still able to rescue growth of CAfree cells in the absence of Complex I (Δ(*nuoA*-*nuoN*)). dabAB2 rescues better than GFP ($t=15.7$, $p=2.37*10^{-8}$). Error bars represent standard deviation of six replicate cultures. "n.s." denotes means do not differ significantly, "*" denotes that means differ with bonferroni corrected $P < 0.05$ by a two-tailed T-test, and "**" denotes $P < 5X10^{-4}$. **B.** These growth curves were used to generate the growth yield values

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

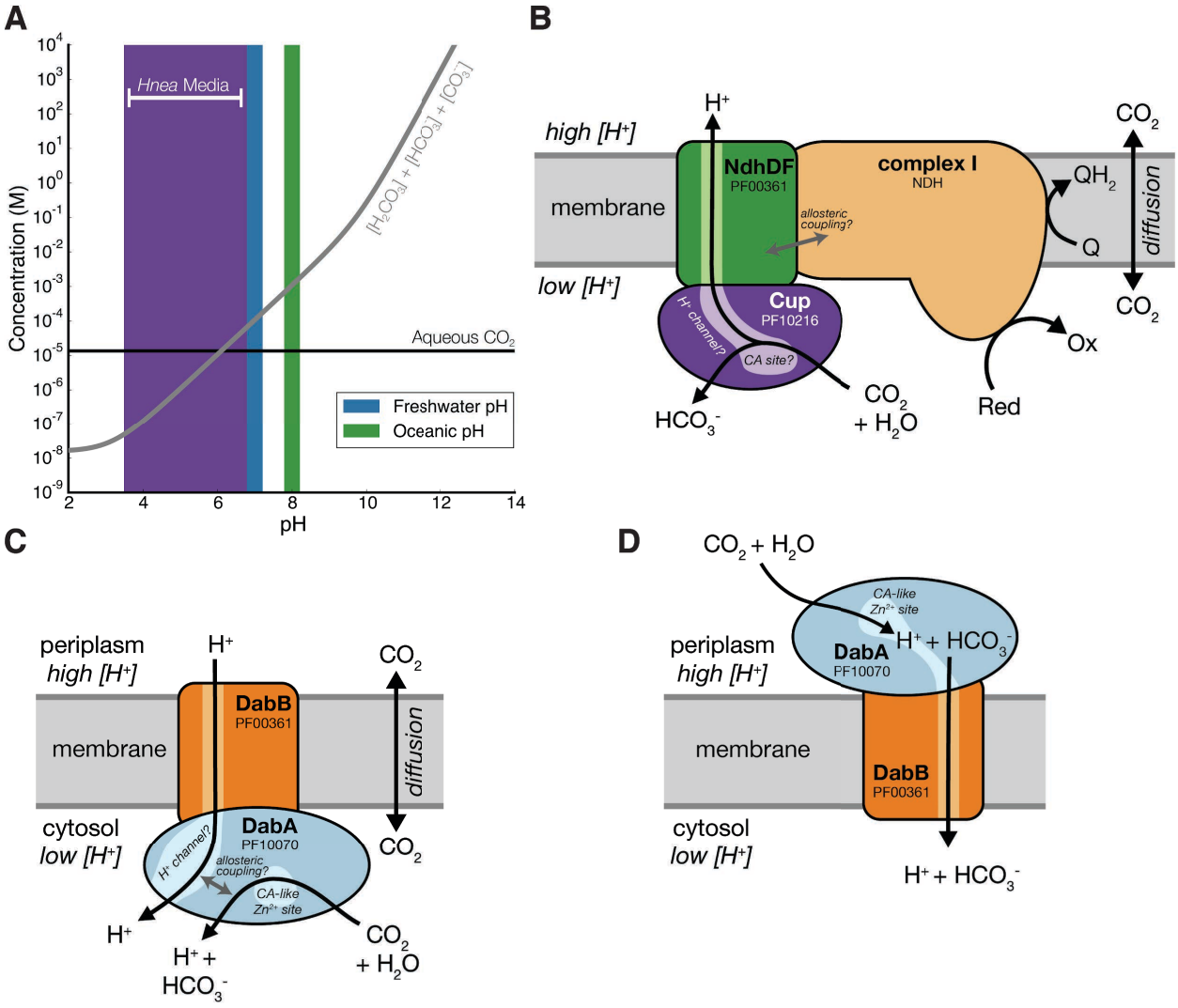

Figure S6. Comparison of models of vectorial CA activity for DABs and the Cyanobacterial CUP systems. A. Equilibrium concentrations of dissolved inorganic carbon as a function of pH. In this plot we assume the growth medium is in Henry's law equilibrium with present-day atmosphere (400 PPM CO_2) at 25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

cyanobacterial Ci uptake systems. Cup-type systems are believed to couple electron transfer to vectorial CA activity and, potentially, outward-directed proton pumping. This model is based on the observation that Cup systems displace the two distal H^+-pumping subunits of the cyanobacterial complex I and replace them with related subunits that bind CupA/B (illustrated in green as NdhD/F). **C.** As our data are consistent with DAB2 functioning as a standalone complex (i.e. DabAB do not appear to bind or require the *E. coli* complex I), we propose a different model for DAB function where energy for unidirectional hydration of CO2 is drawn from the movement of cations along their electrochemical gradient (right panel above). **D.** An alternative model for DAB activity is that DabA is localized to the periplasm and DabB is functioning as a H^+ : HCO_3^- symporter. In this model DabA CA activity is made vectorial by removal of products. Energy is provided in the form of the PMF driving H^+ (and therefore HCO_3^-) uptake. This model is not preferred because no secretion signals were observed in the DabA sequence. Moreover, the *Acidimicrobium ferrooxidans* genome contains an apparent DabA:DabB fusion protein. The predicted architecture the fusion would place DabA in the cytoplasm.

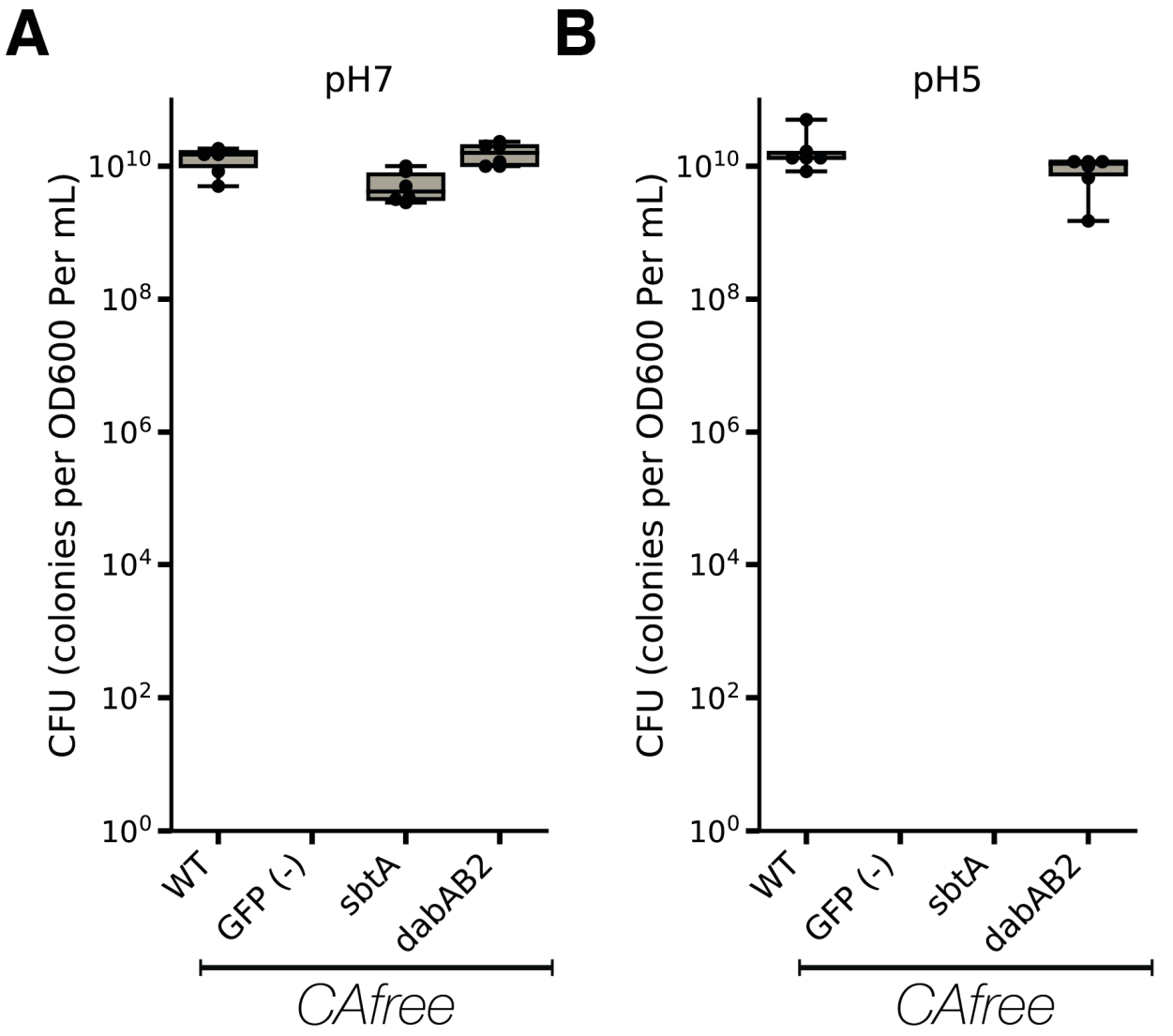

Figure S7. pH independence of *dabAB2* rescue of CAfree Colony forming units per OD600 per ml were measured on LB plates with induction in air at both pH 7 (**A.**) and 5 (**B.**). dabAB2 rescued growth at both pH7 and pH 5, sbtA only rescued growth at pH 7. Whiskers represent the range of the data,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

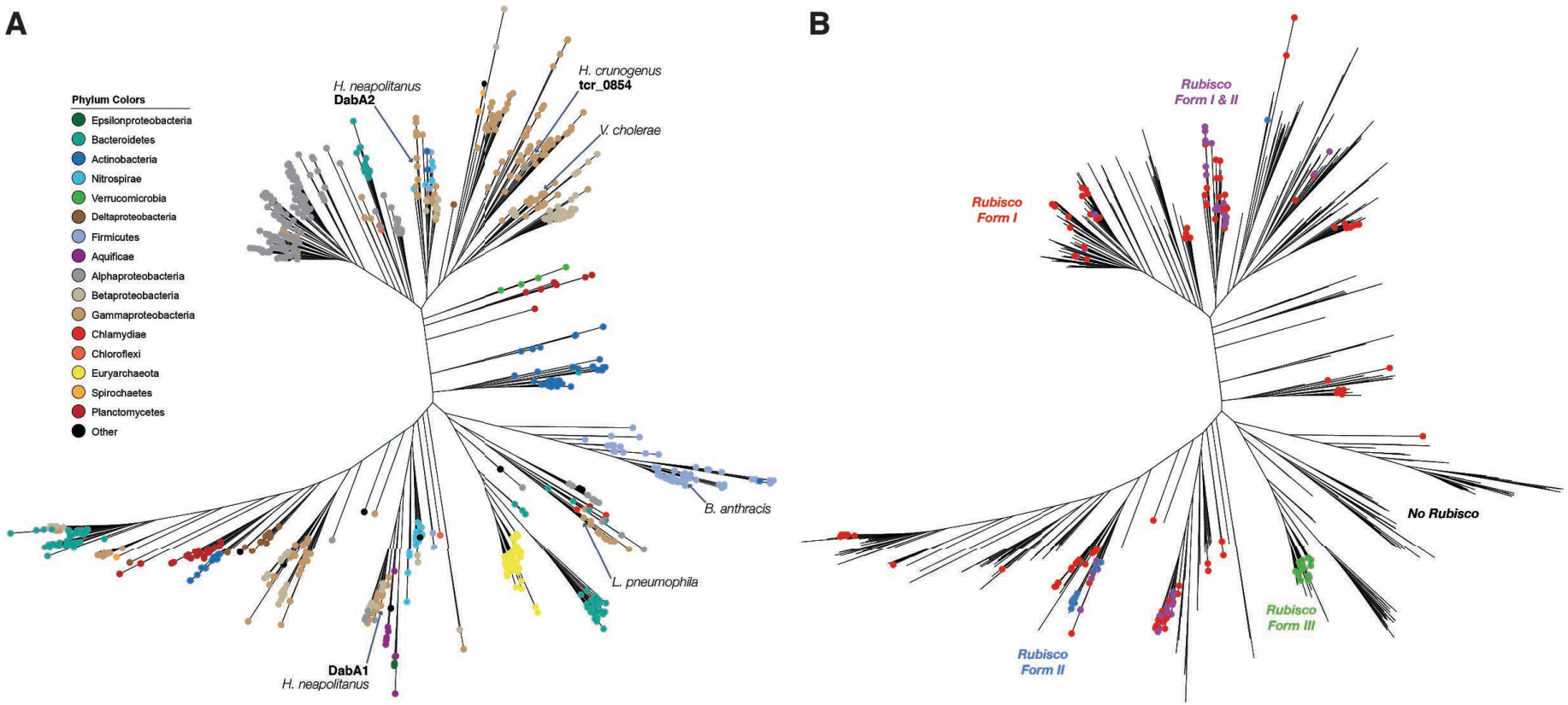
A
Phylum Colors
Epsilonproteobacteria
Bacteroidetes
Actinobacteria
Nitrospirae
Verrucomicrobia
Deltaproteobacteria
Firmicutes
Aquificae
Alphaproteobacteria
Betaproteobacteria
Gammaproteobacteria
Chlamydiae
Chloroflexi
Euryarchaeota
Spirochaetes
Planctomycetes
Other
H. neapolitanus
DabA2
H. crunogenus
tcr_0854
V. cholerae
B. anthracis
L. pneumophila
DabA1
H. neapolitanus
B
Rubisco
Form I & II
Rubisco
Form I
No Rubisco
Rubisco
Form III
Rubisco
Form II

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

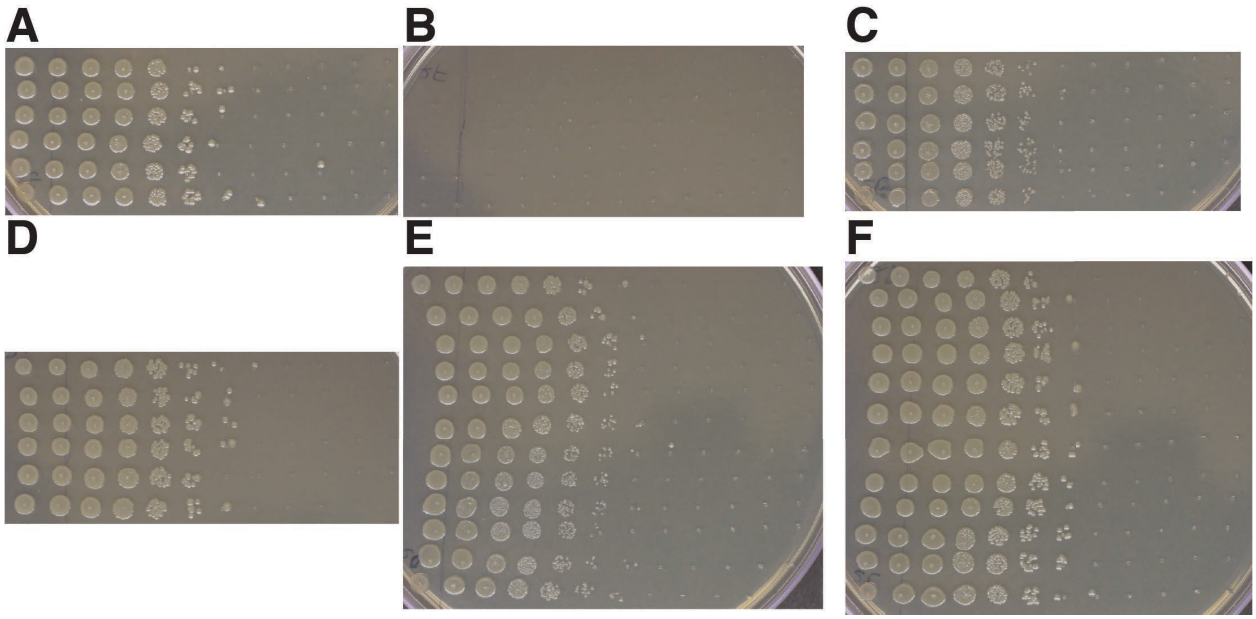

Figure S9. Plates used for determining CFU counts for Figure 5B. A. Wt positive control. **B.** CAfree sfGFP negative control does not rescue. **C.** CAfree hCA positive control rescues growth. **D.** CAfree DAB2 rescues growth. **E.** baDAB from *Bacillus anthracis* rescues growth of CAfree. **F.**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

represents 3ul of an OD 0.2 culture grown at 10% CO_2 each subsequent spot is 3 ul of a 1:10 dilution of the previous spot.