Supplementary Files:

Supplemental File 1. Information on important strains and reagents used in the study. XLSX file: 12 KB.

Supplemental File 2. Transposon insertion information and essentiality determinations broken down by gene. Data is from two technical replicates of the library mapping experiment. P values are calculated using a one-tailed binomial test as defined in the methods. P values are provided both before and after a Bonferroni correction. Numbers of transposon insertions seen for each gene in each replicate are also provided. CSV file: 505 KB.

Supplemental File 3. Fitness effects and HCR phenotypes broken down by gene. Data is from two replicates of the competitive growth assay. CSV file: 268 KB.

Supplemental File 4. FASTA file containing the genes used to generate Supplemental Figure 4a. FASTA file: 471 KB.

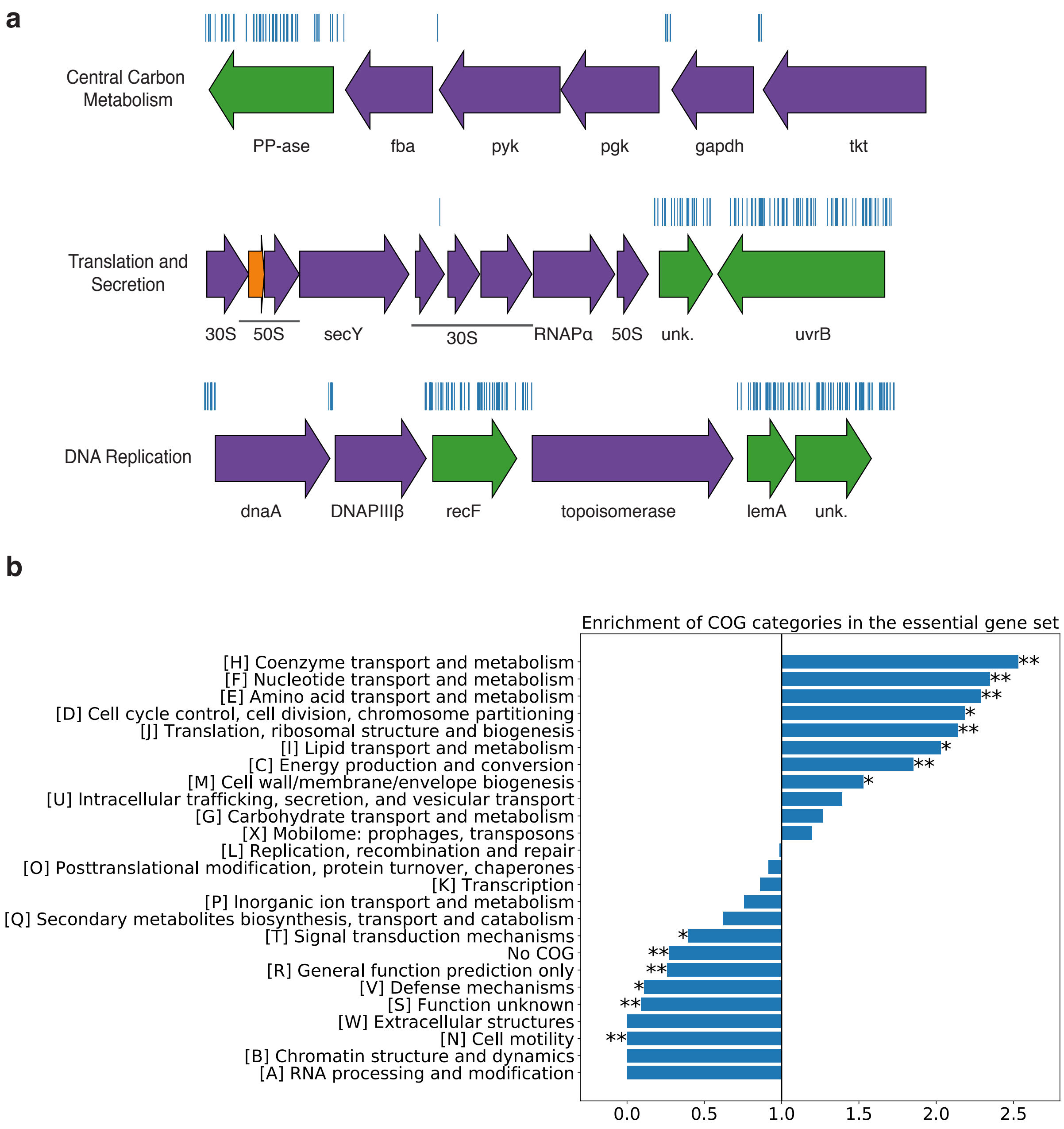
Supplemental File 5. FASTA file containing the genes used to generate figure 5a. FASTA File: 886 KB.

Supplementary Tables:

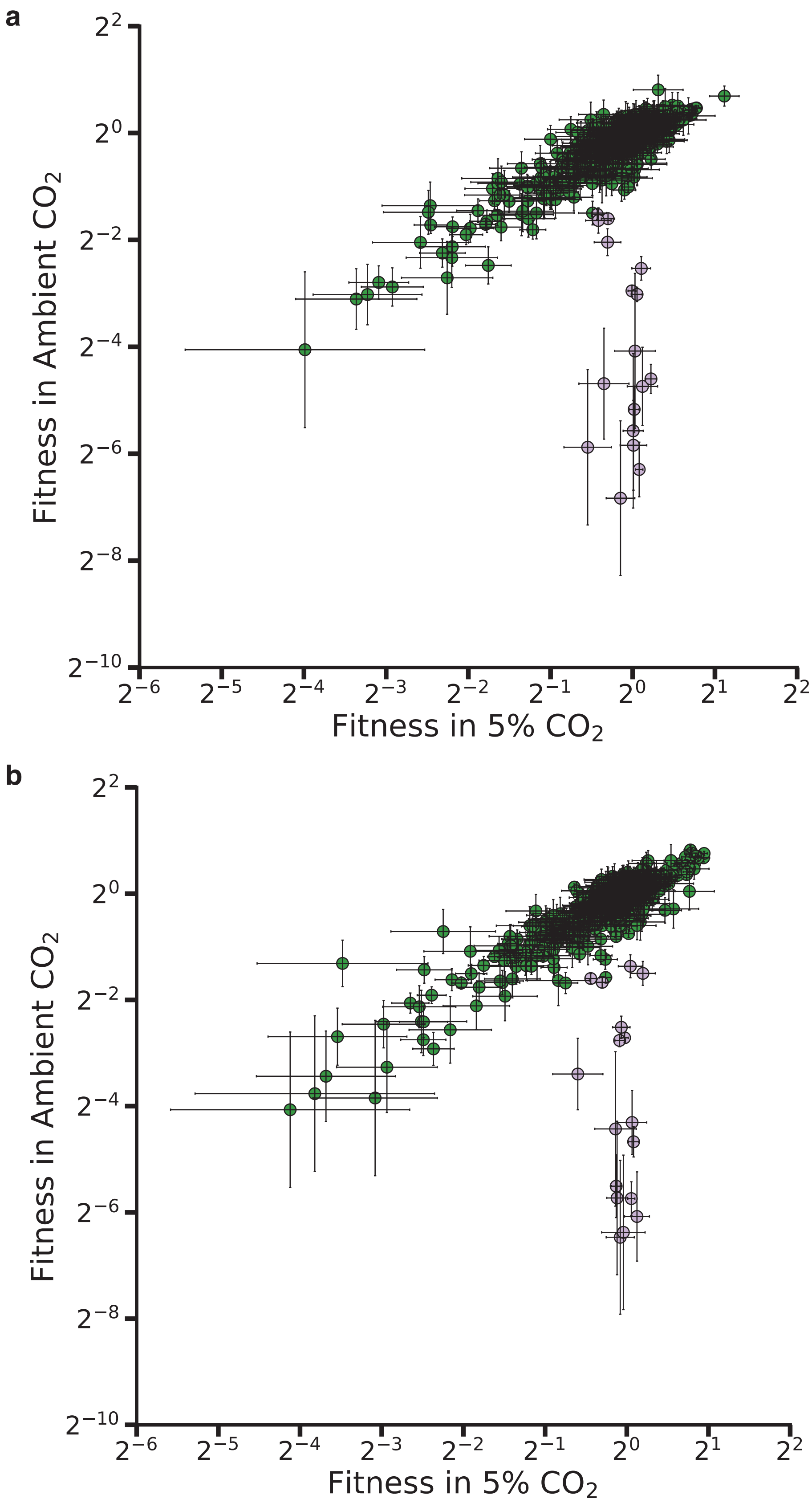
Locus Id	NCBI Accession number	NCBI gi number	Gene description	Has HCR phenotype
HNEAP_RS01030	WP_012823110.1	502585319	DabA2	TRUE
HNEAP_RS01035	WP_012823111.1	502585320	DabB2	TRUE
HNEAP_RS01040	WP_012823112.1	502585321	LysR	TRUE
HNEAP_RS04565	WP_012823782.1	502586009	Csos1D	FALSE
HNEAP_RS04570	WP_012823783.1	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

Supplemental Table 1 Genes from HCR operons. This table includes genes from the HCR operons with their phenotype and identifying information. “unk.” indicates a hypothetical protein.

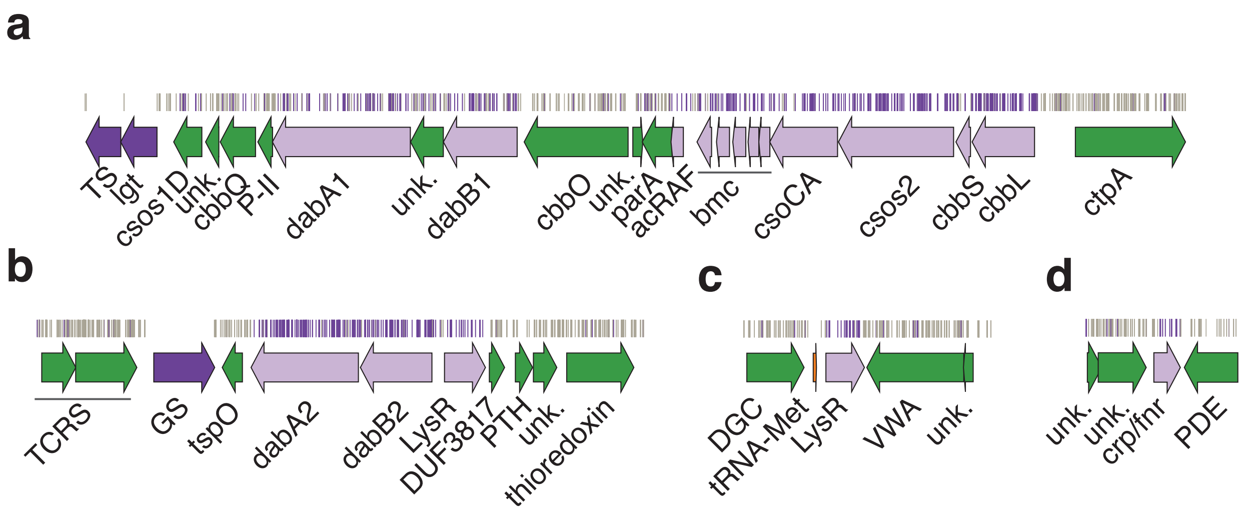
Supplementary Figures:



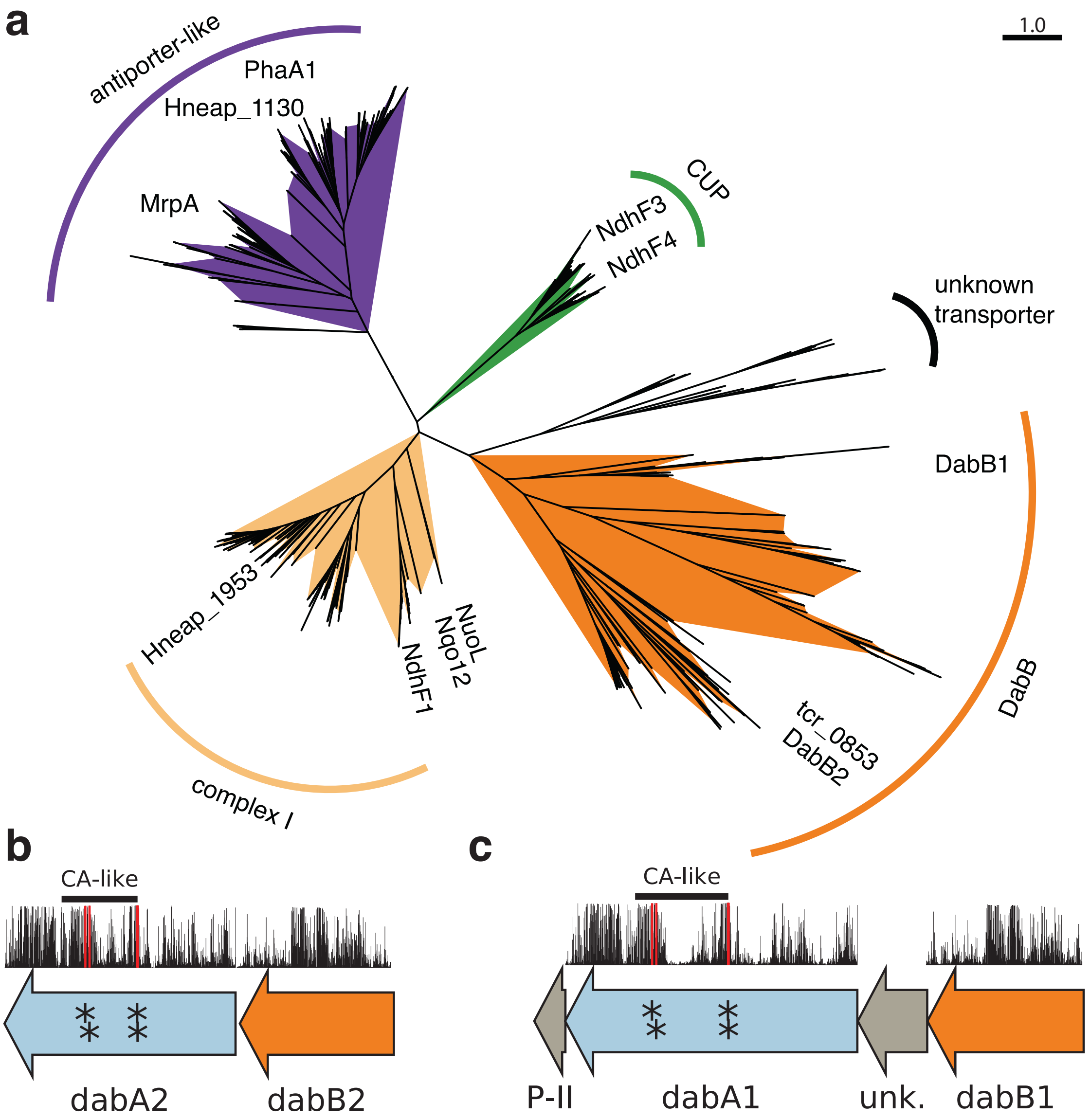
Supplemental Figure 1 The essential gene set is enriched for COGs associated with essential cellular processes. a. Representative essential genes and nonessential genes in the *Hnea* genome. The blue track indicates the presence of an insertion. Genes in purple were called essential and genes in green 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. 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). **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 $p < 5X10^{-4}$. Statistics are as follows, No COG: $n_{genome} = 734$, $n_{essential} = 46$, $p = 4.2*10^{-43}$; A: $n_{genome} = 1$, $n_{essential} = 0$, $p =19.3$; B: $n_{genome} = 1$, $n_{essential} = 0$, $p =19.3$; C: $n_{genome} = 118.0$, $n_{essential} = 50$, $p = 9.6*10^{-6}$; D: $n_{genome} = 28$, $n_{essential} = 14$, $p = 9.5*10^{-3}$; E: $n_{genome} = 128$, $n_{essential} = 67$, $p = 8.6*10^{-11}$; F: $n_{genome} = 54$, $n_{essential} = 29$, $p = 3.7*10^{-6}$; G: $n_{genome} = 62$, $n_{essential} = 18$, $p =2.4$; H: $n_{genome} = 95$, $n_{essential} = 55$, $p < 4.2*10^{-43}$; I: $n_{genome} = 56$, $n_{essential} = 26$, $p = 5.6*10^{-4}$; J: $n_{genome} = 186$, $n_{essential} = 91$, $p < 4.2*10^{-43}$; K: $n_{genome} = 71$, $n_{essential} = 14$, $p =7.9$; L: $n_{genome} = 84$, $n_{essential} = 19$, $p =13.5$; M: $n_{genome} = 143$, $n_{essential} = 50$, $p = 6*10^{-3}$; N: $n_{genome} = 57$, $n_{essential} = 0$, $p = 7.6*10^{-6}$; O: $n_{genome} = 86$, $n_{essential} = 18$, $p =9.7$; P: $n_{genome} = 98$, $n_{essential} = 17$, $p =2.8$; Q: $n_{genome} = 28$, $n_{essential} = 4$, $p =4.9$; R: $n_{genome} = 101$, $n_{essential} = 6$, $p = 7.6*10^{-5}$; S: $n_{genome} = 96$, $n_{essential} = 2$, $p = 9.8*10^{-8}$; T: $n_{genome} = 77$, $n_{essential} = 7$, $p = 3.1*10^{-2}$; U: $n_{genome} = 44$, $n_{essential} = 14$, $p =1.5$; V: $n_{genome} = 39$, $n_{essential} = 1$, $p = 1.2*10^{-2}$; W: $n_{genome} = 10$, $n_{essential} = 0$, $p =1.8$; X: $n_{genome} = 11$, $n_{essential} = 3$, $p =5.7$. Some listed p values are above 1, this is an outcome of the Bonferroni correction for multiple hypothesis testing.


a
Fitness in Ambient CO_2
2^2
2^0
2^{-2}
2^{-4}
2^{-6}
2^{-8}
2^{-10}
2^{-6}
2^{-5}
2^{-4}
2^{-3}
2^{-2}
2^{-1}
2^0
2^1
2^2
Fitness in 5% CO_2
b
Fitness in Ambient CO_2
2^2
2^0
2^{-2}
2^{-4}
2^{-6}
2^{-8}
2^{-10}
2^{-6}
2^{-5}
2^{-4}
2^{-3}
2^{-2}
2^{-1}
2^0
2^1
2^2
Fitness in 5% CO_2

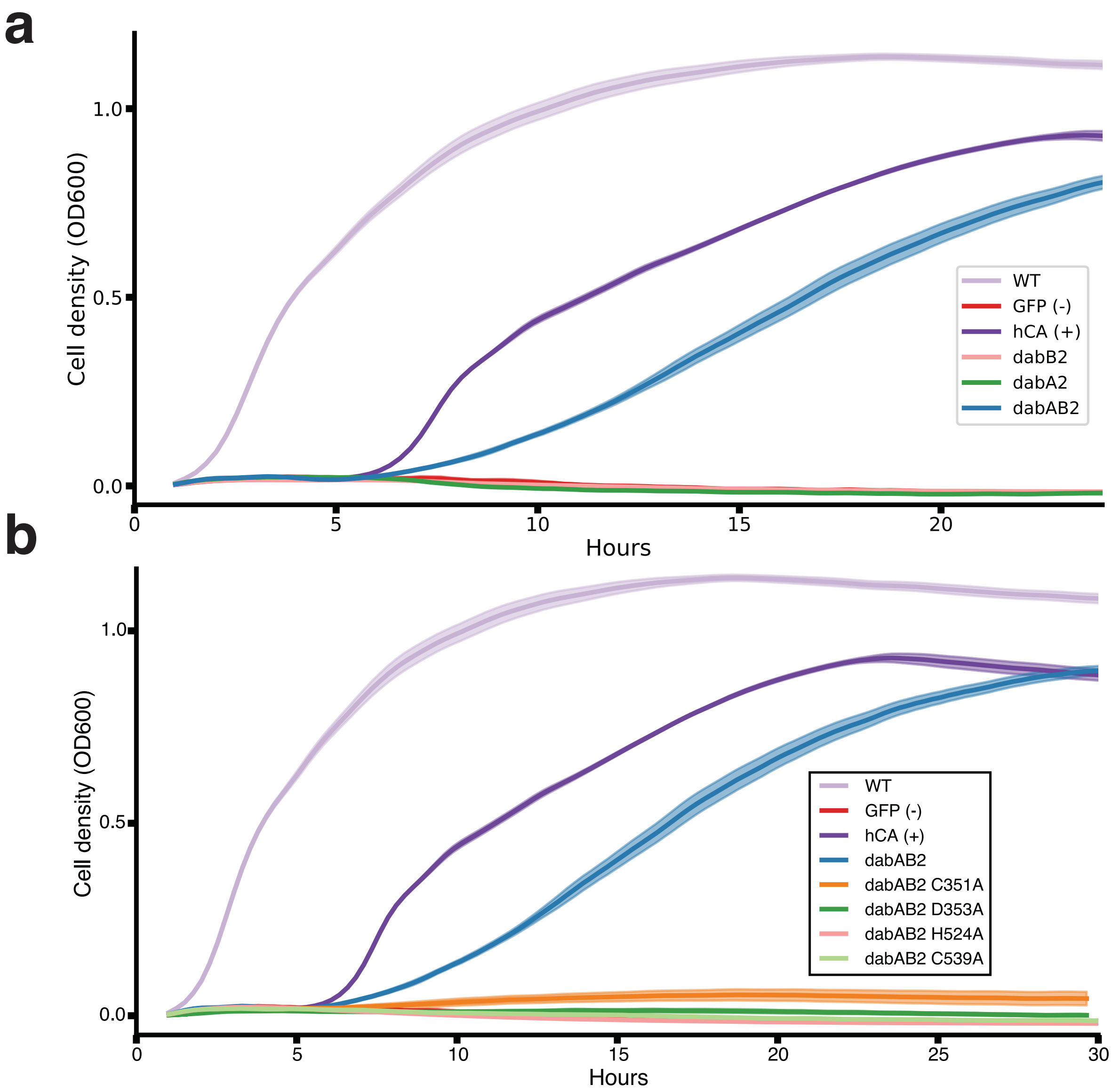
Supplemental Figure 2 Gene fitnesses measurements for each replicates. Fitness effects of gene knockouts in 5% CO_2 as compared to ambient CO_2.The effects of single transposon insertions into a gene are averaged to produce the gene-level fitness value plotted. Points represent means of insertions in the middle 80% of the gene. Error bars represent one standard error of the mean. Sample sizes, means, and standard errors are recorded in supplemental file 3: Fitness effects and HCR phenotype by gene. We define HCR mutants as those displaying a twofold fitness defect in ambient CO_2 relative to 5% CO_2. HCR genes are colored light purple. Panel **a** contains data from the first replicate experiment and panel **b** contains data from the second replicate experiment.



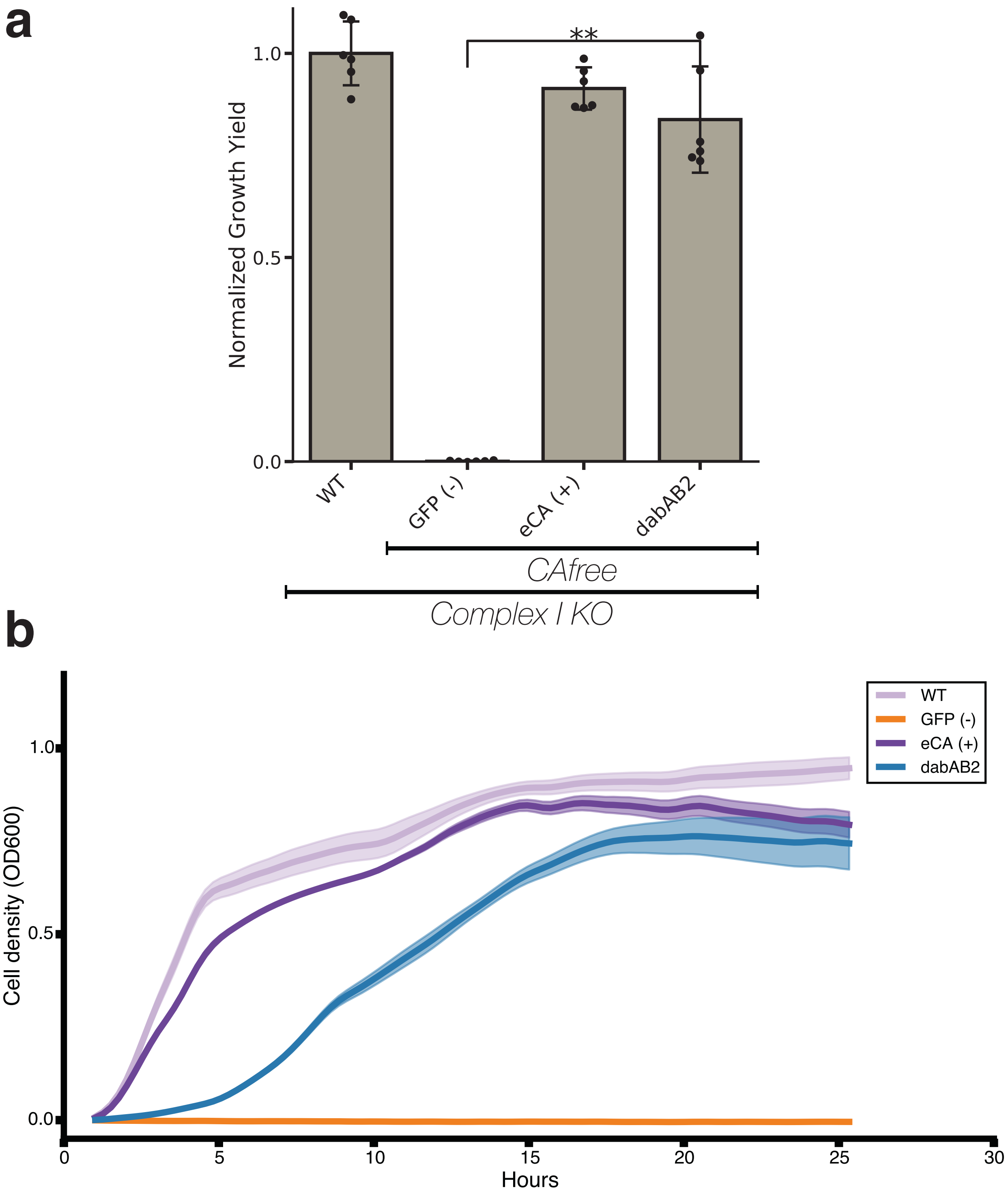
Supplemental Figure 3 Genomic context of *Hnea* HCR genes identified in our genome-wide screen. Panels **a-d** show regions of the *Hnea* genome containing genes annotated as HCR. Essential genes are in dark purple, HCR genes are in light purple, and other genes are in green. The top tracks show the 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 (HNEAP_RS04660-HNEAP_RS04620) and a second CCM-associated operon. This second operon contains acRAF (HNEAP_RS04615), a FormIC associated cbbOQ-type Rubisco activase (HNEAP_RS04575 and HNEAP_RS04600), parA (HNEAP_RS04610), P-II (HNEAP_RS04580) and dabAB1 (dabA1: HNEAP_RS04585 and dabB1: HNEAP_RS04620). **b.** The DAB2 operon and surrounding genomic context (lysR: HNEAP_RS01040, dabA2: HNEAP_RS01030, and dabB2: HNEAP_RS01035). **c.** The genomic context of a lysR-type transcriptional regulator (HNEAP_RS05490) that shows an HCR phenotype. **d.** Genomic context of a crp/fnr-type transcriptional regulator that displays an HCR phenotype (HNEAP_RS07320). Accession numbers and gi numbers for selected genes can be found in Supplemental Table 1. Abbreviations for Supplemental Figure 3: 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).



Supplemental Figure 4 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 an approximate maximum likelihood 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. Scale bar indicates one substitution per site. The tree contains 566 sequences. These sequences can be found in supplemental file 4. **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.

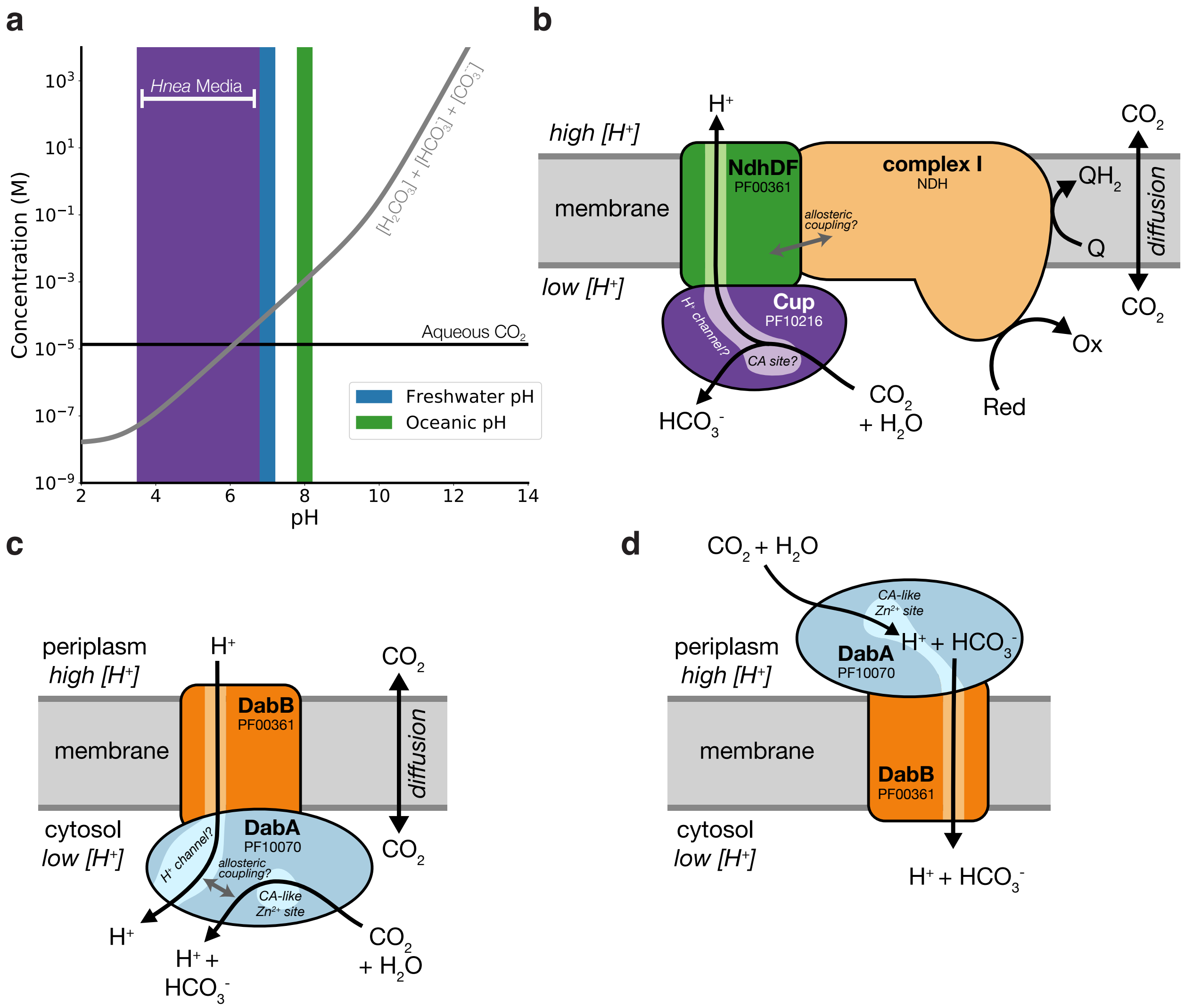


Supplemental Figure 5. Expression of DabAB2 rescues growth of CAfree *E. coli* in ambient CO_2. a. These growth curves were used to generate the growth yield values in Figure 3b. Mean OD600 is plotted +/- standard error for four biologically replicate cultures. Wild-type *E. coli* (BW25113) and CAfree 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 biologically 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.

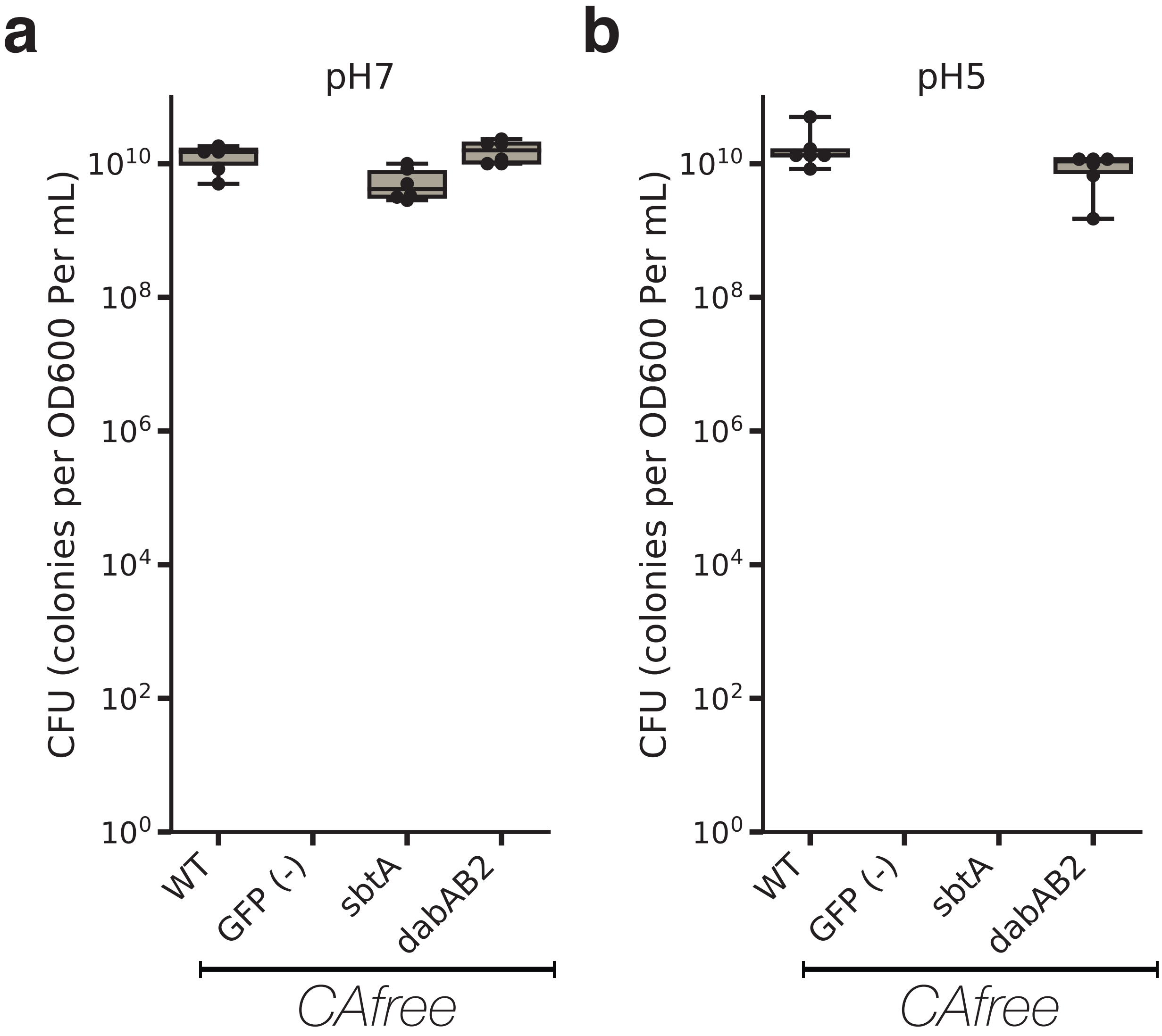


Supplemental Figure 6. DAB2 function is not dependant on complex 1. a. DAB2 is still able to rescue growth of CAfree cells in the absence of Complex I (Δ(*nuoA*-*nuoN*)). dabAB2 rescues better than GFP (t=15.7, p=2.37*10^{-8}). "**" denotes that means are significantly different with Bonferroni corrected p < 5X10^{-4} according to a two-tailed t-test. Bar heights represent means and error bars represent standard deviation of six biologically replicate cultures. Consistent results were observed in an independent growth experiment.

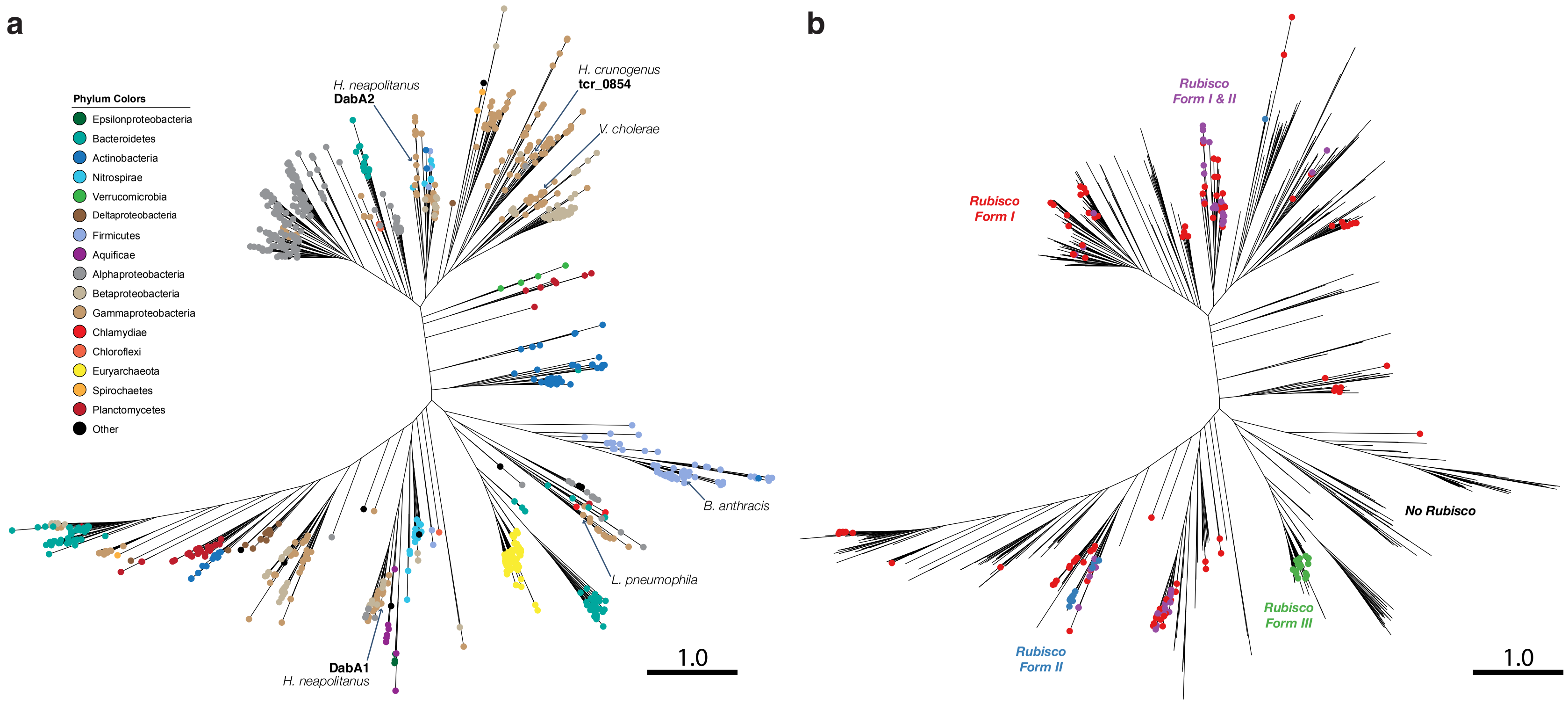
b. These growth curves were used to generate the growth yield values in Supplemental Figure 6a. Mean OD600 is plotted +/- standard error of six biologically replicate cultures. Consistent results were observed in an independent growth experiment. All strains are Complex I knockout strains. DAB2 is still able to rescue growth of CAfree cells in the absence of Complex I.



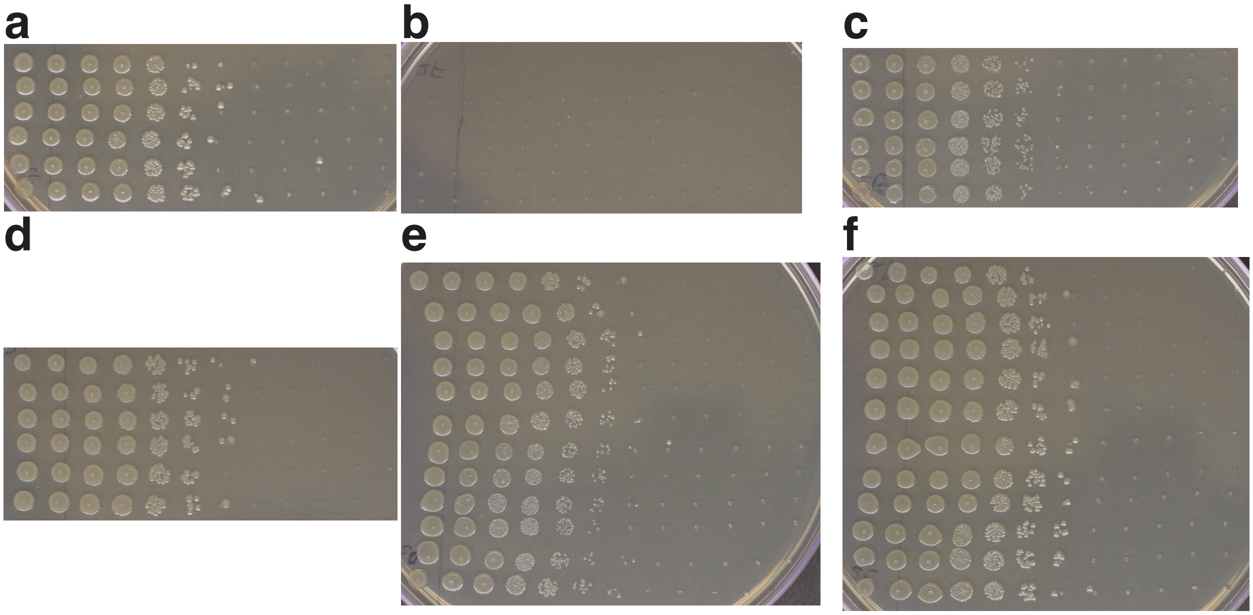
Supplemental Figure 7. Comparison of models of vectorial CA activity for DABs and the Cyanobacterial CUP systems. a. Equilibrium concentrations of dissolved inorganic carbon as a function of pH. In this plot we assume the growth medium is in Henry's law equilibrium with present-day atmosphere (400 PPM 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.



Supplemental Figure 8. 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 technical 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
1.0
b
Rubisco
Form I & II
Rubisco
Form I
No Rubisco
Rubisco
Form III
Rubisco
Form II
1.0

Supplemental Figure 9. 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. For both panels, scale bars indicate one substitution per site. For the trees from both panels, the 878 sequences used to generate the tree can be found in Supplemental File 5.



Supplemental Figure 10. 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.