Response to Reviewers
Desmarais et al. 2019

Reviewer comments in black / author response in-line in magenta
Significant changes are marked in magenta on corresponding manuscript file

Reviewer #1 (Remarks to the Author):

The key results of this work are the identification of genes in the Gammaproteobacterium Halothiobacillus neapolitanus necessary for growth under low inorganic carbon conditions, with an emphasis on genes encoding an inorganic carbon transporter. A library of mutants was created via barcoded transposons, from which 1) essential genes were elucidated (e.g., no, or fewer than expected, mutants carrying ablated versions of these genes found in the library) and 2) the relative fitness of mutants was compared when grown under low (400 ppm CO2) or high (5% CO2 headspace) inorganic carbon conditions. Genes conferring an advantage under low CO2 conditions included those encoding the components of carboxysomes, as expected, as these microcompartments have been identified as being essential to the growth of this and other autotrophic bacteria under low CO2 conditions. Also included are regulatory poteins (lysR (2), crp/fnr), as well as two putative transporters (dabAB1 and dabAB2).

Homologs of these transporter-encoding genes were found to be present in many phyla of Bacteria as well as some Archaea. Evidence for inorganic carbon uptake by these transporters included rescue of a carbonic anhydrase-free strain of E. coli from a requirement for high CO2 concentrations for growth, as well as measurement of inorganic carbon uptake by these E. coli via silicone oil centrifugation. Homologs of these genes from pathogens Vibrio cholera and Bacillus anthracis were also expressed heterologously in E. coli, and rescued the carbonic anhydrase-free construct from requiring high-CO2 concentrations for growth. The proteins encoded by dabAB2 were shown to physically interact. The non-membrane-spanning partner was shown to bind zinc, and its shared ancestry with beta carbonic anhydrase was suggested by structural homology modeling. Similarities with the zinc-binding active site of beta carbonic anhydrase were used to guide the substitutions of single amino acids; when these putative active site amino acids were changed, the complex could no longer rescue carbonic anhydrase-free strain of E. coli, though the altered proteins could still bind zinc.

We thank the reviewer for their insightful commentary and critical feedback, particularly the suggestions related to setting the historical context for our work. Please find our comments and a description of the revisions below.

This is careful, thorough, and interesting work; however, this type of transporter has been previously described by others studying the autotrophic Gammaproteobacterium, Hydrogenovibrio crunogenus. Inorganic carbon uptake by H. crunogenus has been shown to be

energy dependent (Dobrinski, 2005; cited by this manuscript for silicone oil centrifugation protocol). The manuscript cites Mangiapia 2017 (reference 43) as showing that these homologs are 'involved in Ci transport in proteobacteria' (lines 87-8), but much more is known than mere 'involvement'. In Mangiapia 2017, a transposon-mediated mutant library of over 5,000 mutant strains of H. crunogenus, and site-directed mutants, were screened for a high-CO2 requiring phenotype. High-CO2 requiring mutants included strains in which the two genes encoding the transporter were interrupted with transposons; these mutants were unable to grow with ambient air headspace, and were unable to accumulate elevated concentrations of inorganic carbon. More importantly, in a subsequent paper (Appl. Environ. Microbiol. Jan 2019, 85 (3) e02096-18), when these genes were expressed simultaneously and heterologously in E. coli, they conferred an ability to E. coli to transport inorganic carbon. In short, inorganic carbon transport by these types of transporters has already been demonstrated convincingly elsewhere.

This was an oversight on our part and we have expanded the commentary regarding these works as follows:

In the intro: "Scott and colleagues have recently identified and validated homologs of these genes as a C_i import system in hydrothermal vent chemolithoautotrophs".

In the results: "Importantly, operons of this type were recently demonstrated to be capable of C_i uptake in the hydrothermal vent chemolithoautotroph *Hydrogenovibrio crunogenus*".

In the discussion: "DAB1 and DAB2 are homologous to C_i pumps from hydrothermal vent chemolithoautotrophs recently discovered by Scott and colleagues".

Also, in Mangiapia 2017, it was noted that homologs of these genes are widespread among many phyla of Bacteria and Archaea. In Mangiapia 2017, a phylogenetic analysis was also presented, with a long list of organisms in which they were found.

This is a good point and we have added citations to the text in two places, accordingly.

In the discussion before talking about the phylogeny we cite Mangiapia et al. 2017 after the following sentence: "DABs are present in a wide variety of bacteria and archaea".

We also added a citation to Mangiapia et al. 2017 in the results to show that "DabAs were found in a wide variety of prokaryotes including bacteria and archaea (Figure 5A and S9), as is consistent with previous work"

Holthuijzen is cited at line 58 for the energetic input needed to generate elevated concentrations of intracellular inorganic carbon, but they also showed that inorganic carbon uptake in Halothiobacillus neapolitanus was sensitive to disruption of membrane potential, and that CO2 was the likely substrate transported by this species. Since these results are consistent with this study, it would be good to cite them for these advances.

Yes, this is helpful connection to our work and we have included a reference and mention of the Holthujizen et al. results in several locations.

Once in the introduction: “Older physiological measurements suggest that *Hnea* possesses an energized C_i uptake system, but the molecular identity of this activity is unknown”.

Twice in the results: “This is consistent with previous observations that C_i uptake in *Hnea* is powered by a membrane gradient” and “This is consistent with previous observations that CO_2 is the likely substrate of *Hnea* C_i uptake”.

Again in the discussion: “A previous physiological study suggested that *Hnea* C_i uptake is coupled to the membrane electrochemical potential and uses CO_2 as a substrate, but the protein(s) responsible for this activity were unknown”.

Having mentioned these things, it is important to note that there are many new and novel parts of this work that are of highest merit. This is the first use of cutting edge techniques (barcoded transposons, and the fitness assay) to identify, in comprehensive fashion, components of a CO2 concentrating mechanism in an autotrophic microorganism. Other very interesting advances include demonstration of inorganic carbon uptake activity by transporters from H. neapolitanus, probable inorganic carbon uptake by homologs from Vibrio cholera and Bacillus anthracis, strong evidence for protein-protein interaction between the two putative subunits of these transporters, zinc-binding, homology with beta carbonic anhydrase, and more elaborate phylogenetic analyses of the two genes encoding the two subunits of these transporters.

However, it would be appropriate to rephrase several parts of the manuscript, and potentially the title, to make it clearer that these systems, and some of these analyses, have previously been described/published elsewhere, by others.

We have altered the title accordingly. It now reads: “DABs are inorganic carbon pumps found throughout prokaryotic phyla”

Description of inorganic carbon transporting systems is of immense interest, as microorganisms are currently being engineered to catalyze carbon-neutral pathways of industrial importance, and CO2-concentrating mechanism components are being engineered into crop plants to enhance crop yields.

The data, their presentation, and methodology are sufficiently detailed and transparent to enable reproducing the results. Minor points on these items are as follows:

Line 82: Two conditions were used—ambient lab air (~400 ppm CO2) and 5% CO2, and differential fitness was determined by comparing the two populations after a period of time. Might other CCM components ‘pop up’, or display different fitness coefficients, if two other CO2 concentrations were compared? E.g., 5% CO2 vs 100 or 50 ppm CO2? The multiple inorganic

carbon transport systems in cyanobacteria are differentially regulated by a variety of CO2 concentrations (e.g., the ABC transporter is used when the CO2 concentrations are particularly low). It does not seem that more experiments are needed for this particular manuscript, but it seems worthy of mentioning this possibility.

Performing the experiment at reduced CO_2 concentrations would be extremely interesting, as the Reviewer intuits. E.g. it is possible that there are additional genes in the CCM which we were not able to capture in our screen because they don't become essential for CCM function until lower than ambient CO_2 concentrations are experienced or they are rescued by genetic redundancy that would fail to rescue at below ambient CO_2 concentrations. We suspect we are able to catch such redundant proteins because of the nature of quantitative phenotypes (growth defects that don't fully stop growth), and we did in fact observe phenotypes for two independant DAB operons, both of which are functional C_i transporters (We did not show the data for DAB1 for the sake of brevity). However, DAB1 was at the edge of our detection capabilities, and it is possible that there are additional transporters or other CCM genes which had phenotypes below our detection limits due to the effects listed above.

Unfortunately, we do not have the physical setup to lower CO_2 concentrations below ambient levels in a fashion that is suitable for barseq, which makes performing the experiment in a timely fashion challenging. The Reviewer may be interested to know that we have, however, tried intermediate CO_2 concentrations between 400 PPM and 5% and did not observe any new genes appear in these conditions. We did not include these data for the sake of brevity.

This remains an excellent point, however, and we have included the following text in the discussion to highlight how additional genes could have escaped our screen by having phenotypes at lower CO_2 concentrations or genetic redundancy:

"Though it is possible that genetic redundancy, conditional phenotypes, or impairment only at sub-ambient CO_2 permit some genes to escape notice, these data suggest that the proteobacterial CCM is composed of < 30 functionally distinct components."

Line 82: HCR phenotype. HCR is high CO2 requiring. This isn't exactly what is measured with the fitness assay. Some of the strains with a fitness defect of 2 or greater are still growing under low CO2 conditions, so, strictly speaking, they are not HCR. This is clarified in the caption of figure 2; it would be good to broaden the definition of HCR earlier in the manuscript, or use a different terminology

We agree that this is an important point to clarify. We have therefore formally defined HCR in the Results as follows:

“Mutants in a particular gene were designated as HCR if the average effect of a knockout in that gene was a twofold (or greater) growth defect in ambient CO_2 as compared to 5% in two replicate experiments.”

Line 87: what does the ‘D’ in ‘DABs’ stand for? This is not clarified here or elsewhere in the manuscript?

It is a recursive acronym in the style of, for example, the imaging software FIJI, which stands for “FIJI is just imageJ.” We have adjusted the text to read “...DAB operon for ‘**D**ABs **A**ccumulate **B**icarbonate’” to ensure that it is clear that the “D” stands for DABs.

Line 117: ‘contains’ should be ‘contain’

This is now fixed.

Line 168: ‘conserved helical protein’-? Should this be ‘conserved hypothetical protein’?

“Helical” was a mistaken addition, it has been removed.

Line 219: insert ‘not’ between H2CO3 and HCO3-?

This is now fixed.

Line 243: The model is consistent with the data, but at this point ‘more likely’ seems too strong without further experiments showing this to be the case

We agree and have replaced ‘more likely’ with “perhaps”.

Line 255: ‘operon as a i gene’--?

That was a typo, thank you for catching. Additionally, the whole sentence has been restructured to improve clarity, and now reads “843 (96%) of the identified *dabA* sequences were either within three genes of, or fused to, a *dabB*”

Line 279 and elsewhere (e.g., line 592): ‘chemotroph’ should be replaced with ‘chemolithotroph’ or ‘chemolithoautotroph’; ‘chemotrophs’ include E. coli and basically everything else that isn’t a phototroph

‘Chemotroph’ has been replaced with ‘chemolithoautotroph’ throughout the text.

Line 427: ‘grown overnight in atmosphere’—400 ppm CO2?

Yes, this needed clarification; ‘Atmosphere’ was referring to ambient conditions (400 ppm CO_2) and this has been fixed in the text.

Line 454: ‘re-aligned MAFFT and..’ --?

There was a ‘with’ missing from this line, it has been included.

Line 573: ‘chemoautotrophic’ should probably be replaced with ‘chemolithoautotrophic’ to keep the terminology consistent

‘Chemoautotrophic’ has been replaced with ‘chemolithoautotrophic’ where appropriate.

Line 597: “When the effects of single transposon insertions into a gene are mutually consistent, those effects are averaged to produce the gene-level fitness value plotted”… Shouldn’t error bars be included in this graph? Maybe they are but they are very tiny? Why are only mutually consistent data included?

This comment brings up two important points, which have been addressed as follows. First, there was a typo in the figure legend stating that mutually consistent data was averaged. What that sentence should have said was genes were only labeled HCR if their phenotypes were consistent across replicates, but that the data for the graph was the average of fitness effects in a gene in one replicate. This sentence has been fixed to reflect this and now reads “Data is from one of two replicates of BarSeq. The effects of single transposon insertions into a gene are averaged to produce the gene-level fitness value plotted. We define HCR mutants as those displaying a twofold fitness defect in ambient CO_2 relative to 5% CO_2 in both replicates.” In order to be fully transparent, the data from the other BarSeq replicate has been included in supplemental figure S2 and in supplemental file 3: fitness effects and HCR phenotype by gene.

Second, in order to address the question of error bars, we have included a new supplemental figure (S2) containing the fitness graph plotted independently for each replicate experiment and with error bars indicating the standard error included. These graphs are included in the supplemental because there are ~1700 points on the graph making it difficult to visualize the bulk of the data points along with their error bars. Additionally, we have included the standard error of the average for every gene in both replicates in supplemental file 3: fitness effects and HCR phenotype by gene. We have included the sentence “Data from both replicates and the associated standard errors are shown in Figure S2 and in supplemental file 3.” to direct readers to these additions.

Line 625 (caption for figure 3) and caption for figures 4 and 5: ‘T-test’ should be ‘t-test’, ‘P’ should be ‘p’

This is a good point. Capitalization for T-test and P have been corrected to read t-test and p here and in the other figure legends.

Figure 4: In figure 4C, His524 is labeled, while in figure 4E, His534 is labeled. Should both of these be the same?

This was a typo and all has been corrected to read H524 throughout.

Other than the question with regard to figure 2, statistical tests are appropriate, and all error bars are clearly described.

Reviewer #2 (Remarks to the Author):

In this significant and well-written work, Desmarais et al. constructed a genome-saturating mutant library of H. neapolitanus genes using barcoded random insertions. By comparing the mutant growth in ambient and 5% CO2, they were able to identify essential and CCM-related genes. This screen identified known CCM-related genes as well as novel genes like DABs.

They further characterized the role of DABs and found that DabA2 and DabB2 can mediate inorganic carbon transport in E.coli and form a heterodimeric complex. According to phylogenetic analysis, DabA contains a carbonic-anhydrase-like domain, but the authors were not able to observe carbonic anhydrase activity in vitro. Finally they showed that the DAB operon exists in a wide range of prokaryotes including the pathogen V. Cholera. Based on these results, the authors propose a model where DabAB directionally converts CO2 into HCO3-.

The work will be very important for the field because it identifies the parts list of the proteobacterial alpha-CCM, which will enable its further characterization and will enable its transfer into crops to enhance yields. Furthermore, the work identifies and characterizes the DABs as new inorganic carbon transporters.

I think the work is acceptable for publication as-is, pending the authors double-checking the following potential typos (and check through the rest of the manuscript for any others I may have missed):

We thank Reviewer 2 for their feedback and positive comments!

Line 218-219: I did not understand this sentence – is there a typo (missing “not”)? If not, more explanation is needed: “Since DabAB2 rescue is pH-independent in this range, its substrate is likely CO2 and H2CO3, HCO3-, or CO3-2.”

Yes, indeed, there was a missing ‘not’ which has been included.

Line 255: There appears to be a typo: “operon as a i gene”

This typo has been fixed. Additionally, the whole sentence has been restructured to improve clarity, and now reads “843 (96%) of the identified *dabA* sequences were either within three genes of, or fused to, a *dabB*.”

Reviewer #3 (Remarks to the Author):

The manuscript by Desmarais et al. is a truly excellent piece of work that is very novel and would have wide appeal to a range of microbiologists and general biologists. For many years the sulphur bacterium Halothiobacillus neapolitanus (Hnea) has been a model system for analysis of carboxysomes (polyhedral structures that encapsulate Rubisco), part of the CO2 concentrating mechanism (CCM) present in these acidic-adapted bacteria. Although it has understood for some time that the carboxysomal Rubisco in these species would need to be bathed in >10 mM inorganic carbon (Ci) to function, nobody was able to identify the transporters responsible for this essential function of Ci uptake. The authors set out to thoroughly screen Hnea for CCM genes using transposon mutagenesis. The data shown in Fig-1 (and Supp files) represent an extraordinary amount of valuable work that could have occupied an entire publication, but from this comprehensive approach they identified two Dab operons
potentially involved in Ci uptake. The work went on then to extensively characterise Dab2 (2 genes coding for a membrane protein and partner extrinsic protein) and prove that when expressed that they affect accumulation of Ci in E. coli cells (fig-2), probably using CO2 as substrate (Fig-S7) and an energized-conversion to bicarbonate. Furthermore, they were able to identify a putative Zinc-binding domain in DabA2 required for vectorial carbonic anhydrase activity (Fig-4).

Dab transporters, or energized facilitators, are certainly a novel group of transporters that are widespread in bacteria, but not cyanobacteria. The authors note that Dab2 may have utility in strategies to add CCM components to crop plant for the objective of raising photosynthetic efficiency and yield. Interesting they also show that several pathogenic bacteria possess Dab homologs. Furthermore, the authors showed that dab operons from V. cholerae and B. anthracis are active in E.coli (data not shown?). This latter aspect is likely to receive more attention in future. The speculative model for the operation of Dab shown in Fig-6 may not be entirely correct, but it will stimulate future research to obtain a detailed functional model.

The manuscript is well written and represents a very interesting and complete set of data.

We thank the reviewer for their positive sentiment. We agree with their point that our model is, at this point, somewhat speculative. We have therefore added “speculative” to the Figure 6 title.

Minor items:
- Phylogenetic tress really need to have a divergence scale bar include to be of much use. Fig 5a, S3b and S8ab need to have these scale bars added for completeness.

This is an important point. Scale bars of 1 substitution per site have been added to all figures containing phylogenetic trees.

- In Fig-2 it would be very helpful to readers to have specific gene identifiers added for DabAB genes and proteins. From Integrated Microbial Genomes (img.jgi.doe.gov) or Genbank accession numbers.

This is a good point. In order to save space in the figure 2 legend, we have included a supplemental table (Table S1) with the Locus ID, NCBI accession number, NCBI gi number, Gene description, and HCR phenotype call for the 22 genes from the HCR operons and the 3 transcriptional regulators. Additionally, We have included NCBI accession number and NCBI gi number for every gene in the *Hnea* genome in "**Supplemental File 2.** Transposon insertion information and essentiality determination by gene" and "**Supplemental File 3.** Fitness effects and HCR phenotype by gene."

- Ln267: alpha and beta linages are based on what? Form 1a verses 1b linages? Please fix.
We have deleted this section for the sake of brevity.

- Fig-6: it might be best here to refer to the model as a "speculative model" since a considerable body of further work may be required to test or reformulate the mechanism for Dab2.

This is a good point, we have changed Figure 6 to be titled "A speculative model of the unidirectional energy-coupled CA activity of DAB complexes".