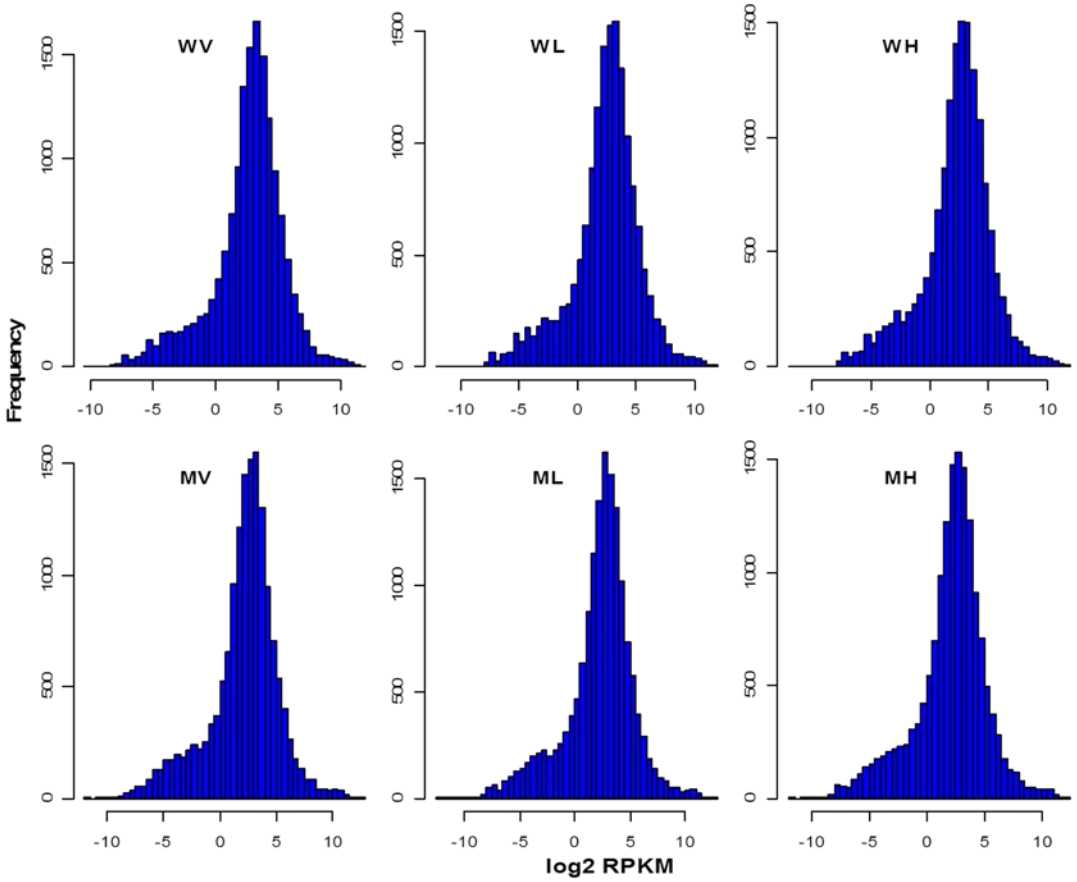
A



B

Percentiles	WV	WL	WH	MV	ML	MH	average
5th	0.052	0.046	0.045	0.036	0.036	0.036	0.050
median	7.559	6.393	6.085	5.685	6.114	5.412	6.602
95th	90.893	89.327	89.106	79.273	83.608	82.291	85.495

Supplemental Figure 1. Gene expression level distributions for each treatment condition.

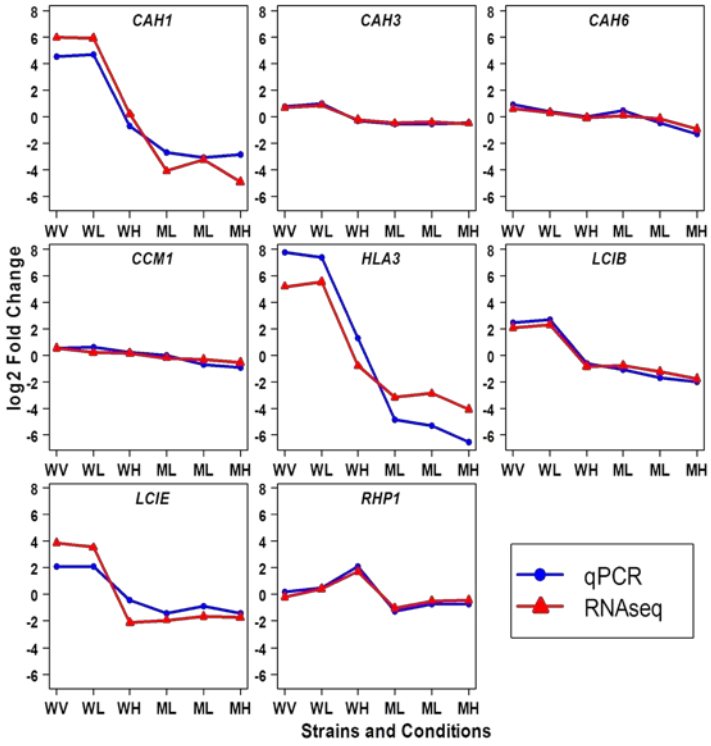
(A) The expression RPKM values are averaged between replicates under each treatment condition, then transformed by logarithm base 2. The title of each histogram indicates each strain and CO_2 treatment condition: WV = wild type under VL-CO_2 induction; WL = wild type under L-CO_2 induction; WH = wild type under H-CO_2 induction; MV = *cia5* under VL-CO_2 induction; ML = *cia5* under L-CO_2 induction; MH = *cia5* under H-CO_2 induction. The shapes of these distributions are very similar among all 6 conditions.

(B) This table summarizes the untransformed RPKMs' 5^{th}, 50^{th}, and 95^{th} percentiles for each treatment condition and the average across conditions; these values are relatively consistent among all 6 treatment conditions.

A

Protein ID	Name	Correlation
522126	CAH1	0.98555
526413	CAH3	0.994444
512520	CAH6	0.970274
518901	CCM1	0.921299
518934	HLA3	0.980099
510298	LCIB	0.995218
522129	LCIE	0.968522
523557	RHP1	0.981095

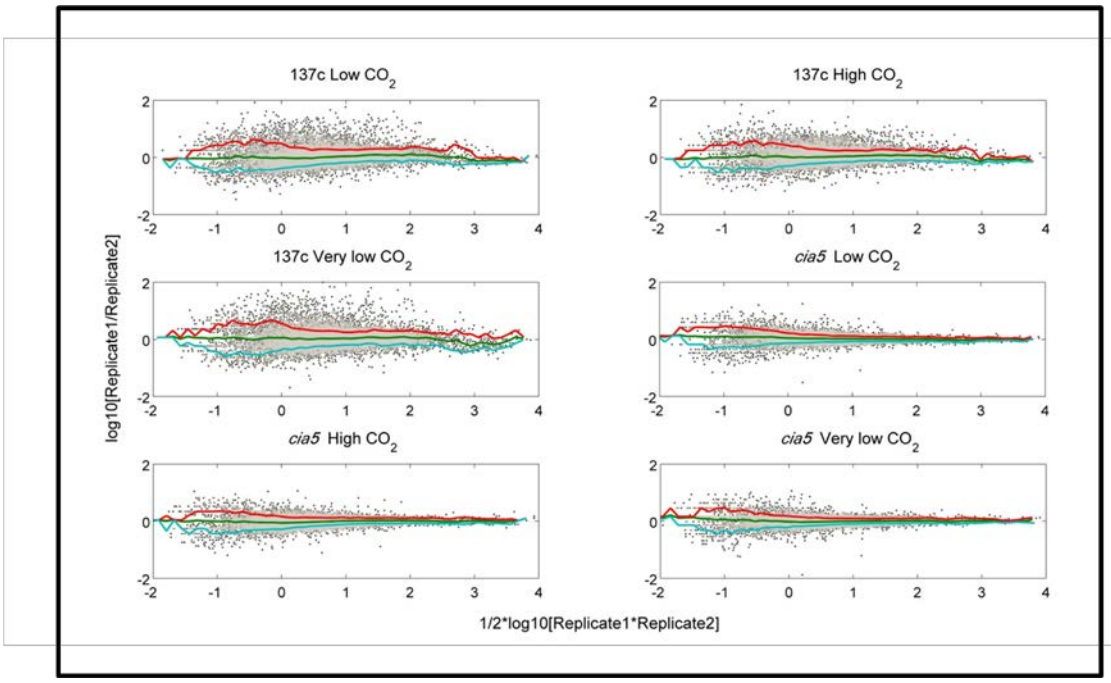
B



Supplemental Figure 2. Validation of RNA-Seq by qPCR.

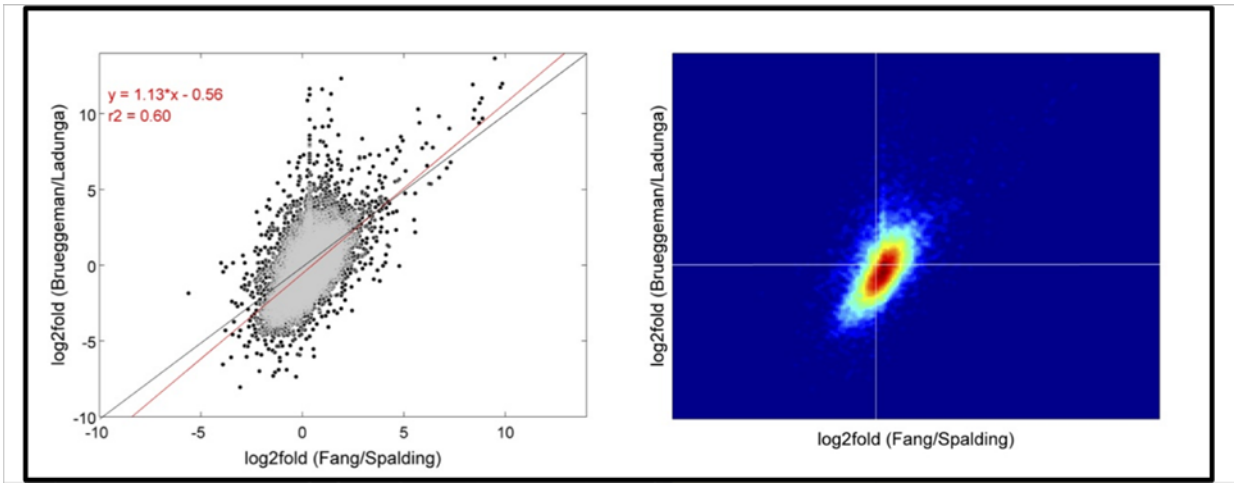
(A) Correlation coefficients between RNA-Seq and qPCR results for each of 8 genes. The correlation coefficients were calculated based on the $\log_2$ Fold Change of each individual condition relative to the overall mean across six conditions from the normalized RNA-Seq data and relative fold change values from the normalized qPCR data.

(B) Relative $\log_2$ Fold Change plot for selected genes. Horizontal axis indicates each strain and CO_2 induction condition: WV = wild type under VL-CO_2 induction; WL = wild type under L-CO_2 induction; WH = wild type under H-CO_2 induction; MV = *cia5* under VL-CO_2 induction; ML = *cia5* under L-CO_2 induction; MH = *cia5* under H-CO_2 induction. Red lines and blue lines separately represent RNA-Seq and qPCR relative $\log_2$ fold, and the closeness of the two lines visually illustrates the agreement between the 2 techniques.



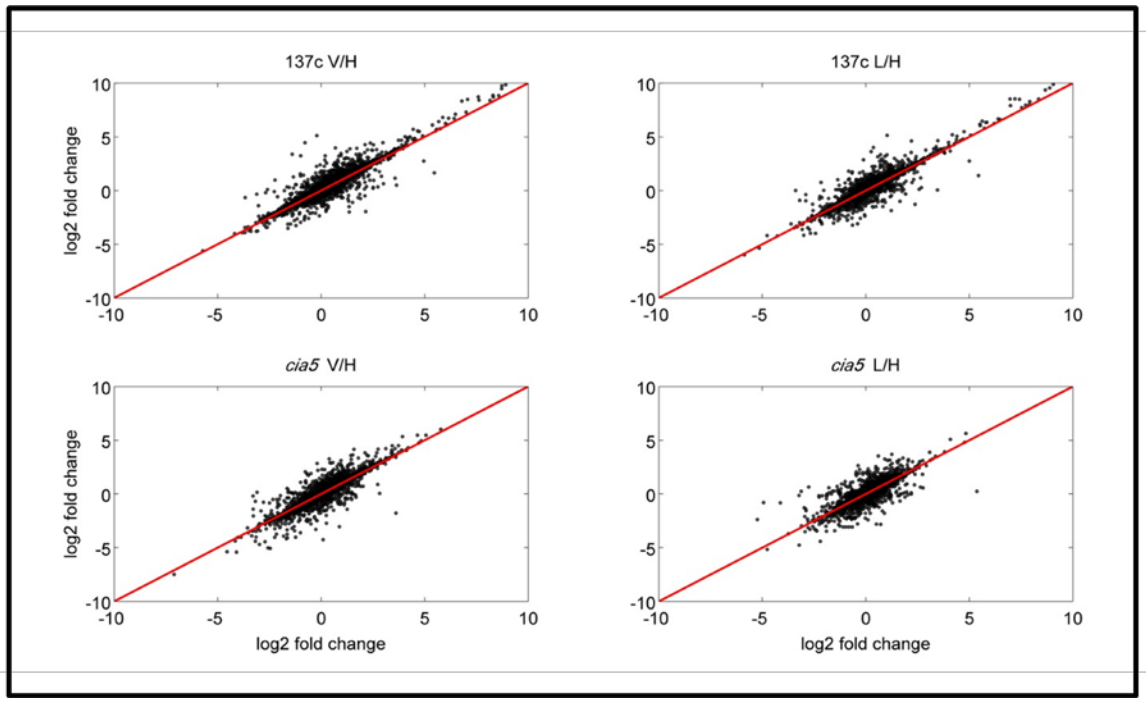
Supplemental Figure 3. Mean-difference scatter plots for biological replicates.

For each experiment, the log-log plot represents the fold change (y-axis) as a function of the geometric mean (x-axis), for each pair of replicates of the same experiment. Quantile line plots in running windows on the x-axis represent the 90^{th} (red), mean (green) and 10^{th} (cyan) quantile of the fold changes. In all cases it can be seen that the mean fold change is around 0 (green line) and a majority of genes show high correlation between different replicates, across most of the dynamic range of the mean expression.



Supplemental Figure 4. Comparison of log2 fold change estimates between different datasets.

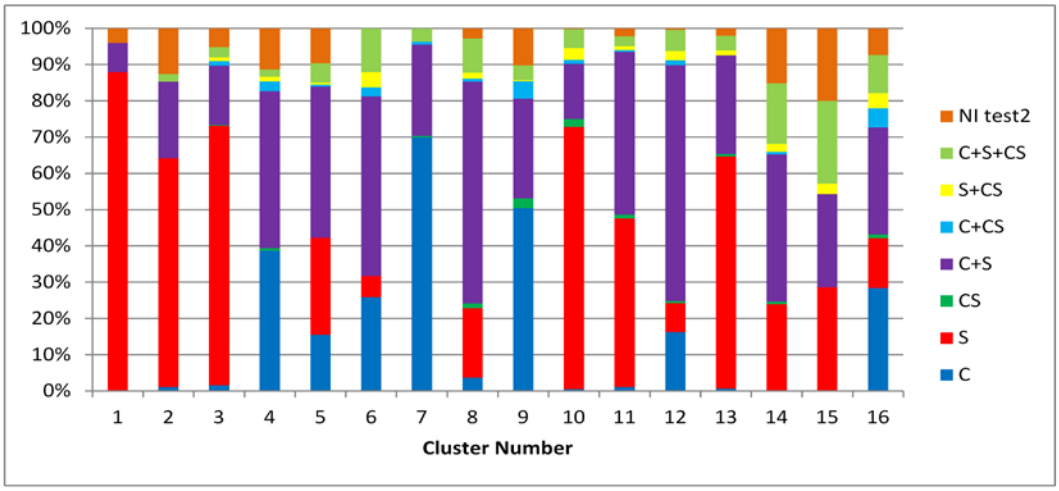
Left: $\log_2$ fold changes of 3 hours after CO_2 deprivation from our companion paper (Brueggeman et al. 2012) are plotted against our fold changes estimates of very low versus high CO_2. A perfect linear relationship is represented with a black line, and the results from a linear fit are highlighted in red. Right: a density histogram of the same data is plotted to show that a majority of fold change estimates are in agreement between both datasets.



Supplemental Figure 5. Comparison of log_2 fold change estimates between different analysis pipelines.

Each panel compares log_2 fold change estimates as presented in the manuscript (x axis) to those obtained from an alternative, simplified pipeline on trimmed sequences (see Supplementary Data Methods). For each strain (137c, *cia5*), shown are log_2 fold changes of very low vs. high CO_2 (V/H) and low vs. high CO_2 (L/H).

A



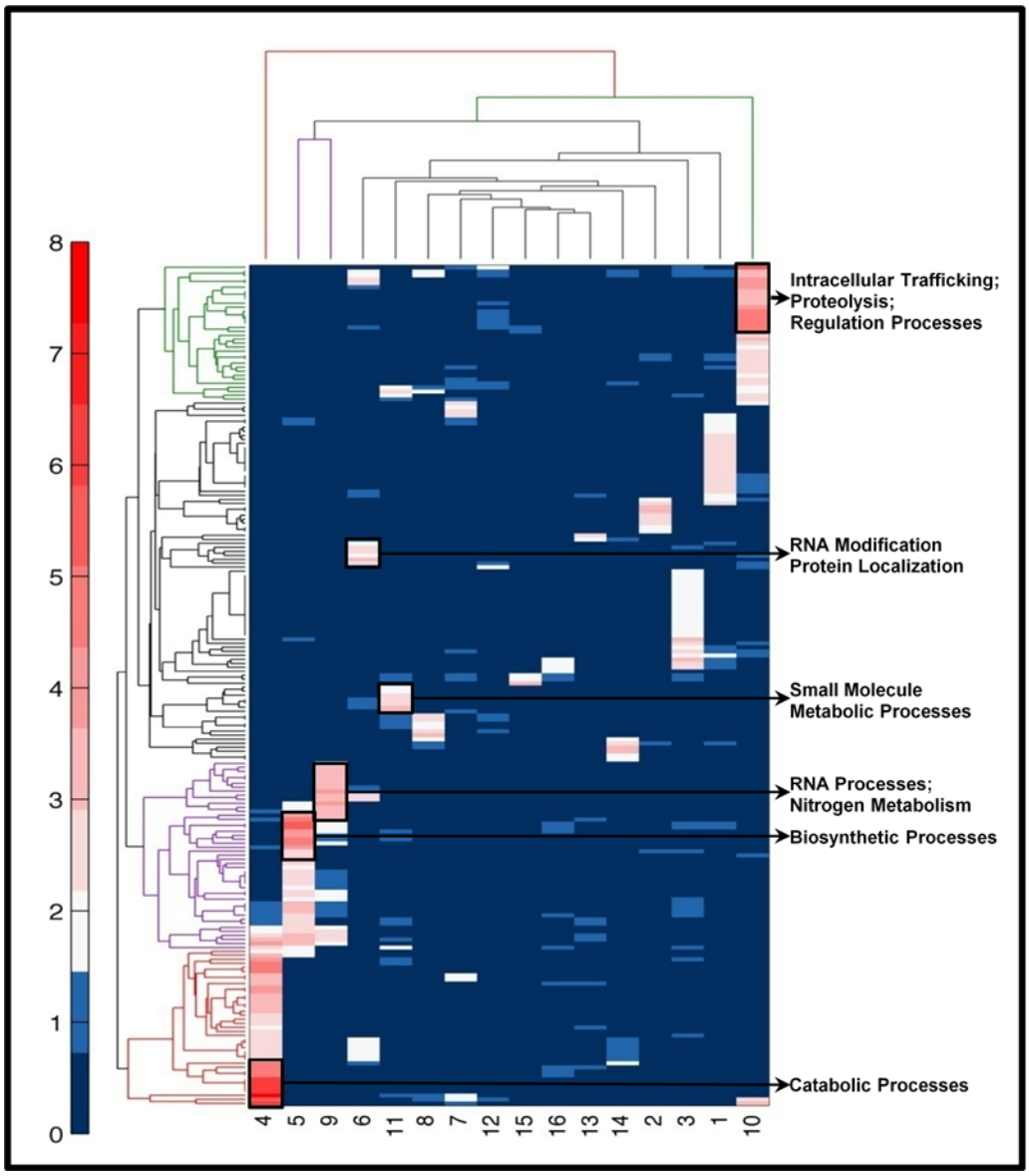
B

Significant effects	Total genes	For each Cluster															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
C	633	0	1	6	58	29	62	170	13	187	2	2	75	1	0	0	27
S	1184	109	60	277	0	50	14	0	69	0	333	85	37	94	33	10	13
CS	36	0	0	1	1	0	0	1	5	10	10	2	3	1	1	0	1
C+S	1324	10	20	63	65	78	119	61	220	102	70	82	301	40	56	9	28
C+CS	57	0	0	5	4	1	6	2	3	18	5	1	6	0	1	0	5
S+CS	63	0	0	4	2	1	10	0	6	1	15	2	12	2	3	1	4
C+S+CS	216	0	2	11	3	10	29	9	34	15	24	5	27	6	23	8	10
NI test2	165	5	12	20	17	18	0	0	10	38	1	4	2	3	21	7	7
total C	2230	10	23	85	130	118	216	242	270	322	101	90	409	47	80	17	70
total S	2787	119	82	355	70	139	172	70	329	118	442	174	377	142	115	28	55
total CS	372	0	2	21	10	12	45	12	48	44	54	10	48	9	28	9	20
total	3678	124	95	387	150	187	240	243	360	371	460	183	463	147	138	35	95

Supplemental Figure 6. Distribution of C/S impact test results by cluster.

(A) C/S impact test results for genes identified by the overall test as DE genes and clustered in 16 clusters. Cluster number is indicated on the horizontal axis, and the vertical axis indicates the percentage sum for significant individual effects, where significant = means q-value <0.025. Different colors indicate specific individual effects or combinations as: C = significant CO_2 effect only; S = significant strain effect only; CS = significant strain and CO_2 interaction effect only; C+S = significant CO_2 and strain effect only; C+CS, significant CO_2 and interaction effect only; S+CS = significant strain and interaction effect only; C+S+CS = all 3 effects are significant; NI test2 = no significant effects in the C/S impact test but identified as a DE gene in the general test.

(B) Summary of the quantitative details for genes in the C/S impact test. The first column lists all combinations of significant individual effects; Total C, Total S or Total CS = all genes with indicated effect, including genes having either or both of the other individual effects. Totals are shown for all genes, as well as totals for each cluster.



Supplemental Figure 7. Heat map for GO category hits based on the Algal Functional Annotation Tool.

The heat map summarizes the Gene Ontology (GO) analysis results in the category of Biological Processes. GO terms and gene clusters were subjected to hierarchical clustering so that gene clusters with common significant (p <0.01) ontology terms are placed close to each other in the tree for clearer illustration. Color schemes are indicated by the left vertical bar, where the numbers show the scale of negative logarithm of *p-values*. As a guide, darker in red means higher statistical significance for GO terms enriched in each cluster. Missing GO terms in any given cluster were assigned a p-value of 1. The almost complete absence of common GO hits between different clusters verifies the functional specificity of our gene clusters. Some highly enriched functional categories for specific clusters are highlighted as examples. Full details and enrichment *p-values* are provided in Supplemental Data Set 3 and discussed in the text.

Supplemental Table 1. Alignment Statistics for the transcriptome sequencing experiment.

Condition and replicate	Total Reads	Read Length	Uniquely Aligned (%)	Uniquely Aligned to AU5 Models	Uniquely Aligned to AU5 Models (%)
H-137c #1	14619355	75	92.5	10896815	74.54%
H-137c #2	13479946	80	93.0	10661929	79.09%
L-137c #1	13777581	75	92.6	10159785	73.74%
L-137c #2	12671440	80	92.9	10023000	79.10%
VL-137c #1	12228767	75	90.3	9749032	79.72%
VL-137c #2	12040923	80	93.1	7982253	66.29%
H-*cia5* #1	14659855	75	91.0	11199868	76.40%
H-*cia5* #2	15759589	83	93.0	12191567	77.36%
L-*cia5* #1	13574874	75	92.0	9464649	69.72%
L-*cia5* #2	18051524	83	92.6	15124925	83.79%
VL-*cia5* #1	15234796	75	91.7	11410522	74.90%
VL-*cia5* #2	19956363	83	93.0	15364398	76.99%

"Condition and replicate" column lists all RNA samples sequenced in this article: "H", "L", and "VL" are the CO_2 conditions; "137c" and "cia5" are the two strains we used in this experiment; "#1" or "#2" indicate the first or second biological replicate. "AU5 model" is the Augustus 5.0 gene model.

Supplemental Table 2. List of qPCR primers.

Augustus 5.0 Protein ID	Gene Name	Primer pair sequences
522126	*CAH1*	5' TCCTGGACGGGAAGGGTT 3' 5' CGATGCGGTTGGTCTGGTT 3'
526413	*CAH3*	5' AACCTGGCGTTCATTGGC 3' 5' CCTTGGGCGAGGGCTT 3'
512520	*CAH6*	5' TCTGGAGTATGCCGTGCTT 3' 5' TTGGCGCTCATGCTGTT 3'
518901	*CIA5/CCM1*	5' GGTCACGATGCGTCATTAGCG 3' 5' CAAGTGGTCCCTGTGATGCTCC 3'
518934	*HLA3*	5' CTCCGAGCGTCGTCTTTGTT 3' 5' TCGGCGTTCAGCTCCTCA 3'
510298	*LCIB*	5' TCACTGGTGACAACACCATCGC 3' 5' TGTTGAACGAGGAGCCGAAGATG 3'
522129	*LCIE*	5' AGCTACGTGGTGGTGAACGG 3' 5' TCATCATGTACTTGCGAGGGAT 3'
523557	*RHP1*	5' TTCGGAGCCTACTACGGATTG 3' 5' GCCTTCTTGGCATCGGTC 3'
514942	*CBLP*	5' ATGTGCTGTCCGTGGCTTTC 3' 5' CAGACCTTGACCATCTTGTCCC 3'