Supporting Information Appendix

The carbonic anhydrase CAH1 is an essential component of the carbon-concentrating mechanism in *Nannochloropsis oceanica*

Christopher W. Gee, Krishna K. Niyogi

SI Materials and Methods

Cells and culture conditions. *Nannochloropsis oceanica* CCMP1779 was obtained from The Provasoli-Guillard National Center for Culture of Marine Phytoplankton (https://ncma.bigelow.org). For all experiments, media consisted of artificial seawater (final solute = 21.2 g L^{-1}) with nutrient enrichment based on full-strength *f* medium (final concentrations of macronutrients: NaH_2PO_4 = 83 µM, NH_4Cl = 2 mM, micronutrients as described in (50)). Media was buffered by 10 mM Tris-HCl pH 8.1. Growth chambers (Percival E41L2, Perry, IA, USA) were set to 28°C with continuous light from fluorescent lamps (Philips Alto II F25T8) at 100 µmol photons m^{-2} s^{-1}. High CO_2 conditions were maintained at 3% CO_2 by a gas mixer and compressed CO_2 tank (Alliance Gas Products, Berkeley, CA, USA); the low CO_2 condition was ambient air (0.04% CO_2). Liquid cultures were shaken at ~100 rpm in sterilized borosilicate Erlenmeyer flasks or polystyrene tissue culture plates (Genesee Scientific, USA). For spot growth assays and strain maintenance, cell patches were streaked onto solid media with 0.9% bactoagar in polystyrene petri dishes.

Because the *cah1* mutant requires elevated CO_2 conditions for growth, cultures were kept at 3% CO_2 to grow to mid-log phase ($1x10^6$ – $2x10^7$ cells mL^{-1}) in preparation for most experiments. Cell density was measured with a Coulter counter (Beckman Coulter Multisizer 3). For measurements of O_2 evolution, immunoblot analysis, and fluorescence quantification and microscopy, liquid cultures were grown at 3% CO_2 and then either kept at 3% CO_2 (control) or transferred to ambient air for 24 h before the assay, unless stated otherwise.

Cloning of the homologous recombination, complementation, and cell marker constructs. To generate the mutations in the alpha carbonic anhydrase gene NannoCCMP1779_6698-mRNA-1, which we named "*CARBONIC ANHYDRASE 1*" (*CAH1*), we utilized this strain's ability to perform homologous recombination (29). All DNA sequences from CCMP1779 were obtained from the version 1 genome assembly (https://bmb.natsci.msu.edu/faculty/christoph-benning/nannochloropsis-oceanica-ccmp1779/) (23). 1-kb flanking homology regions were first amplified from genomic DNA by polymerase chain reaction (Phusion polymerase, Thermo Fisher) with appropriate *attB* sites in the primers for Multisite Gateway cloning (Thermo Fisher) into pDONOR 221 vectors. The 5' homology region includes the 1-kb region 5' of the ATG start codon, and the 1-kb 3' region begins 250 bp 3' of the ATG. The hygromycin resistance gene driven by the promoter region of the *LIPID DROPLET SURFACE PROTEIN* (*LDSP*) (NannoCCMP1779_4188) gene was generously provided by the lab of Christoph Benning (Michigan State University, USA) (23). This hygromycin resistance cassette was cloned into a pENTR vector resulting in flanking *attR* sites (23). These three entry vectors were cloned into pDEST-R1-R2 with the following final order: 5' homologous region (1 kb), hygromycin resistance cassette (2.1 kb), 3' homologous region (1 kb) (Fig. S1).

Complementation constructs were cloned into a pDONOR 221-derived vector containing a kanamycin resistance gene for bacterial selection and pUC origin of replication. For selection in *Nannochloropsis oceanica*, a zeocin resistance cassette was formed by a 0.9-kb promoter region of the native *EUKARYOTIC INITIATION FACTOR* (*eIF3*) gene (NannoCCMP1779_11214-mRNA-1), the *Sh ble* zeocin resistance gene, and 0.3-kb terminator from *60S RIBOSOMAL PROTEIN L21* (*RPL21*) (NannoCCMP1779_9668). This vector also contained a terminator region from the *ADP RIBOSYLATION FACTOR* (*ARF*) (NannoCCMP1779_4318) gene at the 3' end of the cloning site. Cloning to assemble the following complementation constructs was accomplished via the Gibson method (51). For both the FLAG and Venus-tagged complementation lines, the 1-kb promoter region immediately 5' of the ATG of *CAH1* was used to drive the coding sequence of *CAH1* obtained by PCR from cDNA. At the C-terminus of the coding sequence, either the coding sequence for the Venus fluorescent protein (54) or a 3x tandem FLAG tag with 9x glycine linker was inserted with a new stop codon at the 3' end. For the Venus construct, the *ARF* terminator was used, and in the FLAG construct, this was replaced with the 0.7-kb terminator region from the native *CAH1* locus.

For the cytoplasmic fluorescent protein marker, the zeocin resistance gene was replaced with *AphVIII*, which confers resistance to paromomycin, and a 1-kb promoter from *YIDC* driving expression of mCerulean (37) was cloned 5' of the *ARF* terminator. The epiplastid ER luminal mCerulean marker was derived from this construct by replacing the *YIDC* promoter with the 1-kb promoter region of the ATPase β subunit gene (9984-mRNA), adding the targeting peptide from *PROTEIN DISULFIDE ISOMERASE* (*PDI*) (10287-mRNA) (20) in-frame to the N-terminus of mCerulean and additionally adding a KDEL ER retention signal to the C-terminus. To generate the His to Ala mutation of the active site (H177 in *N. oceanica* CAH1, corresponding to H119 in *H. sapiens* CA-II), mutagenic primers were used in conjunction with Gibson assembly. The *cah1* mutant line #5 was used as the recipient strain for all complementation experiments.

Transformation was performed as described previously (29). Briefly, after PCR amplification from the plasmid templates and purification, 1 µg of DNA was delivered by electroporation (BIO-RAD Gene Pulser II; 0.2 mm cuvette, 2,200 V, 50 µF, 500 Ω) to cells washed in 384 mM sorbitol. After recovery in liquid *f*

medium overnight, cells were plated on agar medium containing either hygromycin B (50 µg mL^{-1}), paromomycin (30 µg mL^{-1}), or zeocin (2 µg mL^{-1}) to select for resistant transformants. Primers flanking the expected insertion site were used to establish correct recombination (F 5'-GAGGACGAAAGCATGAAGGC-3', R 5'- GACACATGCGTTGGAATTCTCG-3').

Growth assays. For spot growth assays, the cell density of liquid starter cultures was normalized to $2x10^7$ cells mL^{-1} across samples in a given experiment. 1:10 serial dilutions were made, and 5 µL of each suspension was deposited on duplicate agar plates resulting in 10^5, 10^4, 10^3, and 10^2 cells in a series of spots. One duplicate plate was placed at 3% CO_2 and the other in ambient air for 7 days of growth before imaging with a digital camera.

DIC affinity curves and half-saturation determination. 15 mL samples of liquid cell cultures were collected by centrifugation and re-suspended in fresh media (without added DIC) to volumes appropriate to equalize cell density to $5x10^7$ cells mL^{-1}. 2 mL of these suspensions were placed in a sealed chamber with a Clark-type electrode (Hansatech Instruments, UK) to measure O_2 concentrations. Cells were mixed by a micro stir bar at 100 rpm. DIC-dependent O_2 evolution was measured similarly to (26, 52). Each run was bubbled for 5-10 min with N_2, sealed, then allowed to fix residual DIC and reach the DIC compensation point (defined as near-zero O_2 evolution in the light) before the beginning of the experiment. Concentrated stocks of DIC made from $NaHCO_3$ were added to the reaction sequentially to yield the desired [DIC] and O_2 evolution rate recorded. All genotypes or treatments were measured in one day as a replicate, and the experiment repeated on sequential days for triplicates or quadruplicates. Pigments were extracted in ethanol and quantified by absorbance spectroscopy (53) to normalize the O_2 evolution rates by chlorophyll. The resulting normalized DIC-dependent photosynthesis curves were then processed by a Python (http://www.python.org/) script that used the scipy 'curve_fit' function to fit a Michaelis-Menten equation to the data and solve for V_{max} and $K_{0.5}$(DIC). Due to the relatively small sample sizes (n = 3 or 4), individual data points are shown rather than error bars (55). Ethoxyzolamide (EZ) (Sigma-Aldrich) stocks were made in DMSO at 1,000x concentrations. For photosynthesis inhibition experiments, cells were incubated for 1 h in standard medium with a final concentration of 100 µM EZ or an equivalent volume of DMSO as the mock treatment.

Chlorophyll fluorescence by PAM fluorometry. Cells were grown in 6-well tissue culture plates at 3%, and then acclimated for 24 h in ambient air. Cells were collected in 2 mL microcentrifuge tubes (8,000g x 5 min) and resuspended in fresh *f* medium with a final concentration of either 2 mM or 40 mM DIC added as $NaHCO_3$, capped, and placed in the dark for 30 min prior to the beginning of the experiment. Chlorophyll fluorescence was analyzed by pulse amplitude modulation (PAM) fluorometry using a DUAL-DR detector and Dual-PAM-100 analyzer (both Heinz Walz, Germany). Red actinic light of 100 µmol photons m^{-2} s^{-1} was used and measurements taken after 5 min of illumination when parameters had stabilized. To assess if the two genotypes responded differently in the assay [DIC], a two-factor ANOVA analysis performed using JASP Version 0.7.5.5 to determine a p-value for the interaction term Genotype * [DIC].

Multiple sequence alignment of alpha-carbonic anhydrases. The predicted amino acid sequence of CAH1 was used as a query for blastp using the default settings (https://blast.ncbi.nlm.nih.gov). Protein sequence was downloaded from the BLAST results for manually selected organisms to show the highly-conserved zinc binding domain across divergent taxa. The open-source bioinformatics suite UGENE (56) was used to perform the alignment using the Multiple Sequence Comparison by Log-Expectation (MUSCLE) application with default settings. Alignments were prepared for figures using Jalview v2 (57).

Prediction of intracellular targeting signals. TargetP and SignalP (http://www.cbs.dtu.dk/services/TargetP/) (35, 58) were used to predict intracellular localization of CAH1 and the signal peptide cleavage site, respectively. To corroborate these predictions, a method specifically designed for heterokont protein targeting called HECTAR (HEterokont subCellular TARgeting) was employed (http://webtools.sb-roscoff.fr/) (36).

Fluorescence microscopy. Liquid cell cultures were grown to mid-log phase (~1-2 x 10^7 cells mL^{-1}) at 3% CO_2 and transferred to ambient air for 24 h before imaging. Cells were embedded by mixing directly on the slide 5µL of a cell suspension with 5µL of a 1% low-melting point agarose solution in *f* medium that was boiled and cooled to less than 40℃ before mixing. Slides were imaged with a Zeiss AxioImager M2 with a 100x oil objective and 1.6x optovar. Differential interference contrast (DIC), Cy5 (chlorophyll autofluorescence), CFP (cyan fluorescent protein), and YFP (yellow fluorescent protein) filter sets were used, and the same settings were used for all image acquisition. Linear image leveling (uniform across all images of a given channel), cropping, and montage assembly was accomplished with ImageJ (59).

Electron microscopy. Cells were harvested from mid-log phase liquid cultures by centrifugation and subjected to high pressure freezing for cryoimmobilization (Balzers HPM010), freeze substitution and embedding in LR white resin (60). Samples were sectioned, affixed to TEM grids and imaged with a FEI Tecnai 12 transmission electron microscope.

Immunoblot Analysis. 15 mL of ~$2x10^7$ cells mL^{-1} were collected by centrifugation and resuspended in 200 µL of solubilization buffer (0.25 M Tris-HCl, pH 6.8, 3.5% (w/v) SDS, 10% (v/v) glycerol, 6% (w/v) urea) and disrupted by bead beating (MP Bio Fast Prep-24; 2x 60 s at 6.5 m s^{-1}, lysing matrix D). Samples were centrifuged at 21,000g for 10 min to pellet debris and supernatant was transferred to clean microcentrifuge tubes. Chlorophyll was extracted and quantified in ethanol (53) and used to normalize samples by chlorophyll concentration prior to incubation with 0.1 M DTT and heated to 95℃ for 60 s. Equal volume of normalized sample was loaded into Mini-PROTEAN

TGX Any kD (BIO-RAD) pre-cast protein electrophoresis gels for SDS-PAGE, which was followed by transfer (Hoefer; Semi-Phor semi-dry transfer unit) to 0.45 μm PVDF membranes (Amersham; Hybond). After blocking with skim milk in PBS-T, membrane sections were probed with either an anti-FLAG antibody (Sigma-Aldrich; rabbit host, F7425) or an antibody recognizing the β subunit of ATP synthase (Agrisera; rabbit host, AS05 085). For this blot, we consistently observed a band near 75 kDa and one near the 50 kDa marker; we chose the ~54 kDa band as this corresponds to the expected size for these proteins. After PBS-T washes, membranes were probed with a horseradish peroxidase-linked anti-rabbit IgG secondary antibody (GE Healthcare, UK; NA934V) and signal detected with SuperSignal West Femto enhanced chemiluminescent (ECL) substrate (Thermo Scientific) and Fluorochem HDII (Alpha Innotech, USA) imager.

Fluorescent reporter induction. Bulk culture fluorescence was measured with an absorbance/fluorescence plate reader (Tecan; Infinite M1000 Pro) directly in the clear polystyrene 24-well plates containing the cultures. Venus fluorescence was quantified (excitation = 515 nm, emission = 535 nm; 5nm bandwidth; bottom read, 50 flashes at 400 Hz) and normalized to absorbance at 750 nm to account for differences in cell density. Cultures were well-mixed immediately prior to measurements by sealing wells with parafilm and shaking, followed by 4000 g x 5 s to collect liquid back in the wells. The experiment was initiated when one plate was transferred to ambient air. During the induction time courses, to prevent cultures from overgrowing, 1/3 of the volume of each well was replaced with fresh media after measurements at 13 h and 31 h post transfer.

SI Appendix References

54. Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K (2001) A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnology* 20:87-90.
55. Krzywinski M, Altman N (2013) Points of Significance: Error bars. *Nat Methods* 10(10):921–922.
56. Okonechnikov K, et al. (2012) Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics* 28(8):1166–1167.
57. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189–1191.
58. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8(10):785–6.
59. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671–675.
60. McDonald KL (2014) Out with the old and in with the new: Rapid specimen preparation procedures for electron microscopy of sectioned biological material. *Protoplasma* 251(2):429–448.
61. Iverson TM, Alber BE, Kisker C, Ferry JG, Rees DC (2000) A closer look at the active site of gamma-class carbonic anhydrases: high-resolution crystallographic studies of the carbonic anhydrase from *Methanosarcina thermophila*. *Biochemistry* 39(31):9222–9231.
62. Tchernov DAN, Livne A, Kaplan A, Sukenik A, Journal I (2008) The kinetic properties of ribulose-1,5-bisphosphate carboxylase/oxygenase may explain the high apparent photosynthetic affinity of *Nannochloropsis* sp. to ambient inorganic carbon. *Isr J Plant Sci* 56(1):37–44.

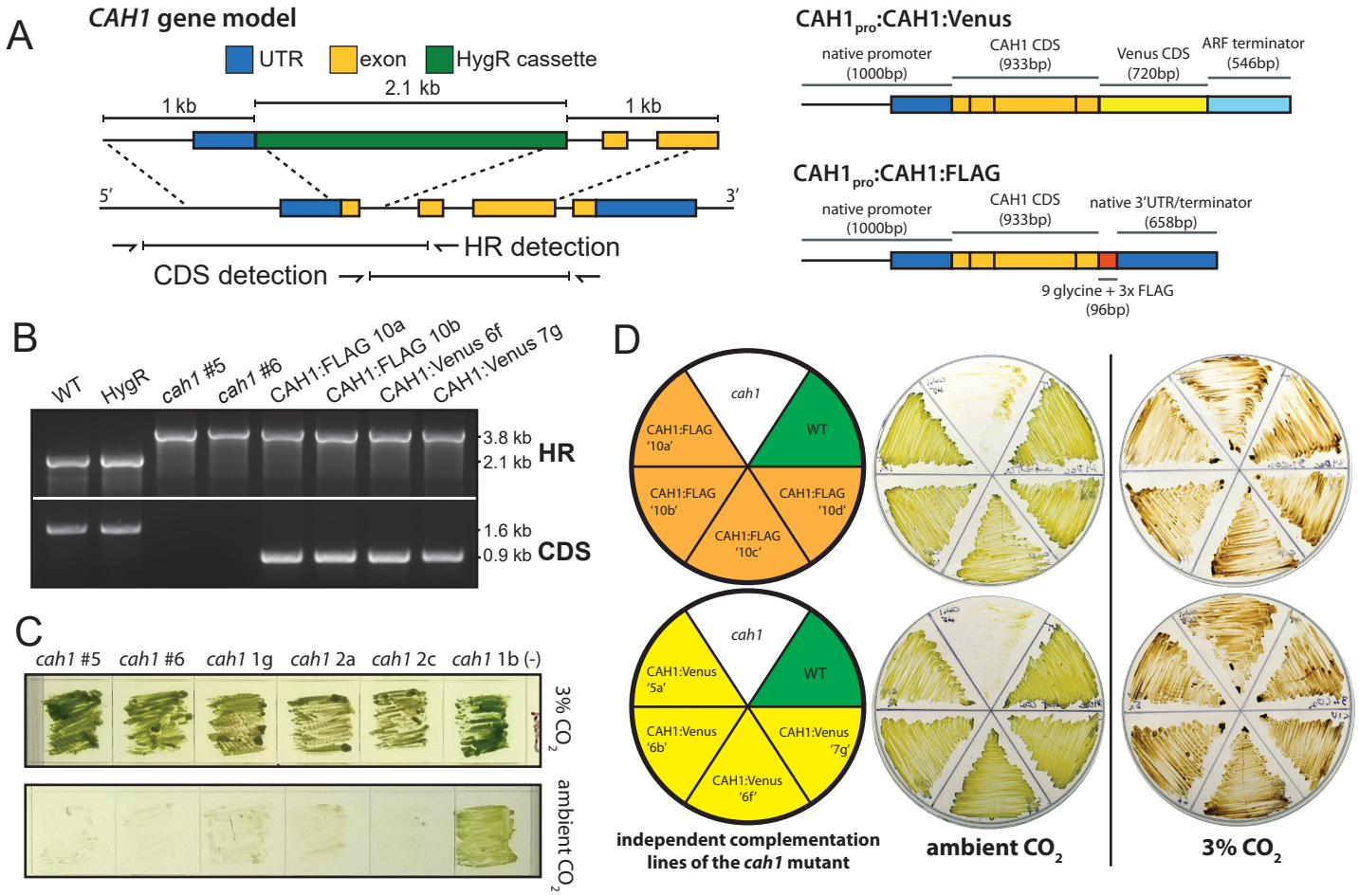

Fig. S1. Disruption of the *CAH1* locus by homologous recombination. (*A*) Gene model with expected recombination event. 1 kb homology regions flanking a hygromycin resistance cassette direct recombination to replace 250 bp that includes the first exon. Genotyping primer binding sites are shown; for homologous recombination (HR), primers sites flank the expected recombination event to screen for amplicon size. To genotype complementation strains (rightmost 4 lanes), the forward primer binds within the replaced region, leading to presence/absence of amplicon for *cah1*, and a size change for WT/complementation line because the coding sequence (CDS) was used. Schematics are shown for the complementation constructs. (*B*) Genotyping PCR results. The 2.8 kb band corresponds to the disrupted CAH1 locus, and the 2.1 kb band corresponds to the WT product. For the CDS PCR, the 1.6 kb band corresponds to the WT genomic sequence, and 0.9 kb is the expected product from amplifying the CAH1:FLAG or CAH1:Venus constructs (lacking introns). (*C*) Additional *cah1* lines showing the CO_2-dependent growth defect. Line 1b(-) was kept as a negative control line that was hygromycin resistant but failed the genotyping PCR, showing that the *cah1* phenotype was not linked to expression of the hygromycin resistance cassette. (*D*) Additional complementation lines showing rescue of normal WT growth at ambient CO_2.

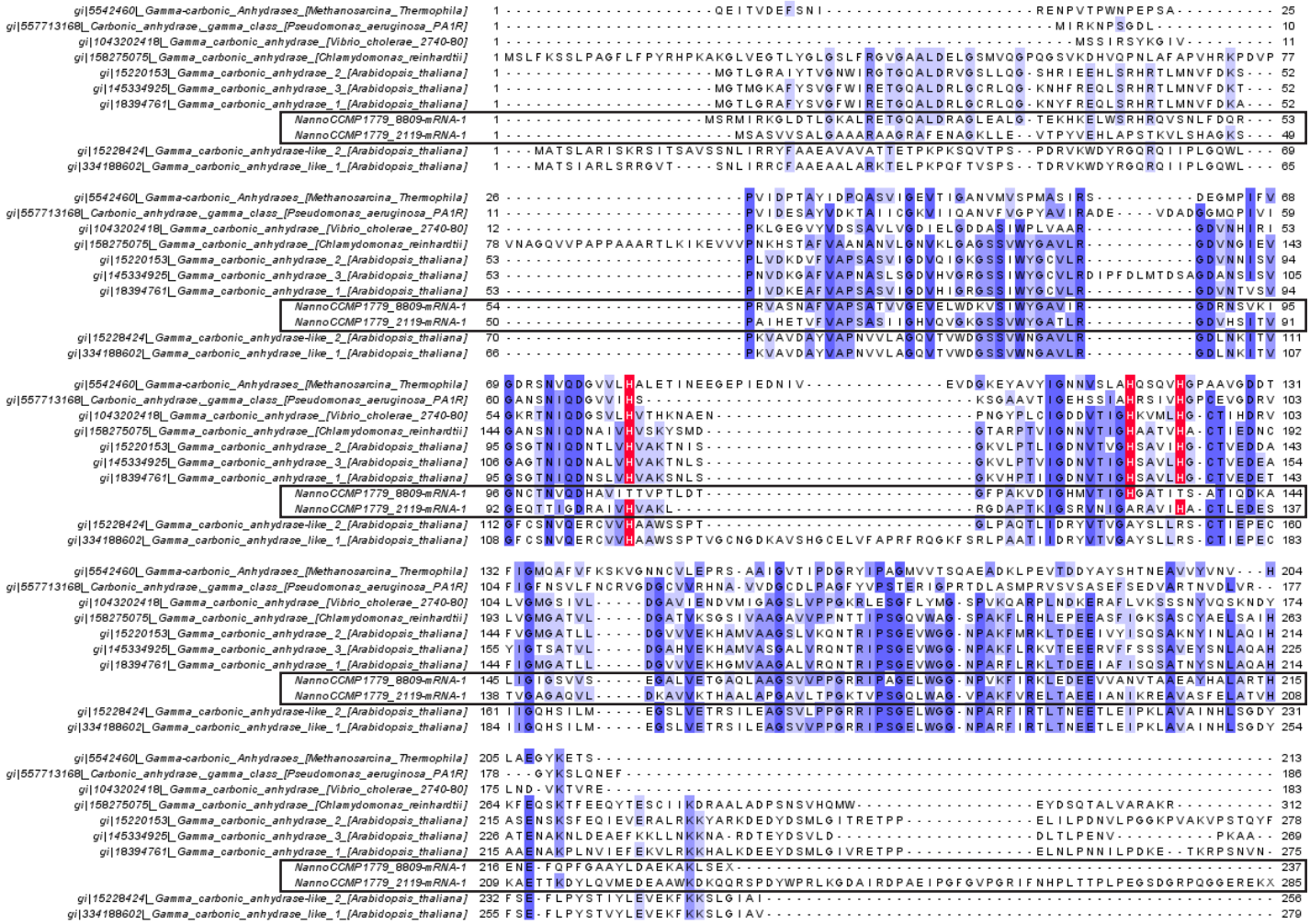

Fig. S2. Full-length multiple sequence alignment of gamma-type carbonic anhydrases. The predicted proteins from *N. oceanica* CCMP1779 are missing one or more conserved histidine residues. Primary sequences for gamma-type carbonic anhydrases were obtained from the online NCBI Protein database (GI numbers used as identifies) and aligned with the two candidates from *N. oceanica* using MUSCLE. Conservation is shown as blue highlighting, critical histidine residues corresponding to H81, H117, and H122 of the prototypical enzyme from *Methanosarcina thermophila* (61) are highlighted in red, and the *N. oceanica* candidates are marked by boxes.

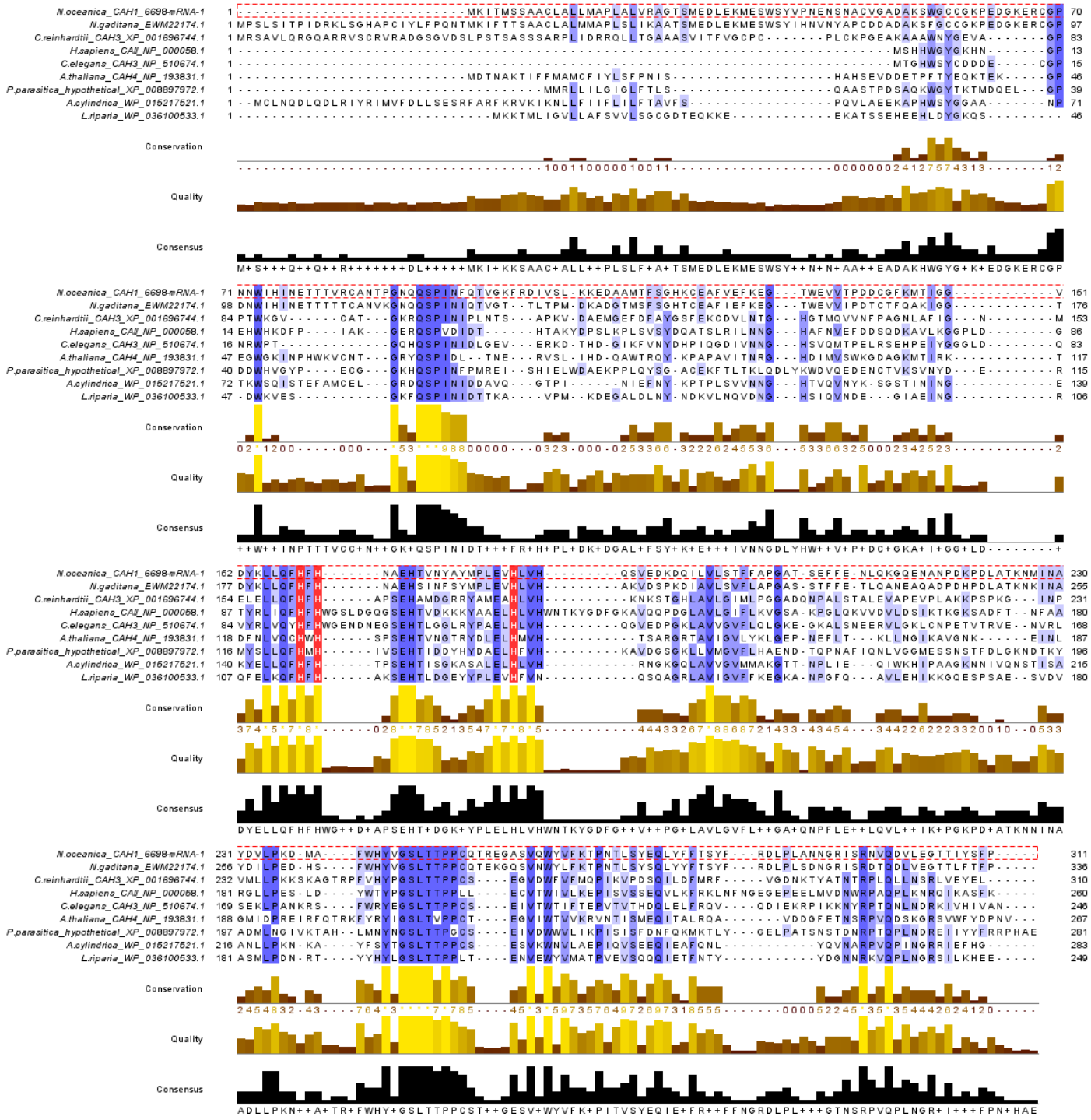

Fig. S3. Full-length multiple sequence alignment of alpha-type carbonic anhydrases. Primary sequence for alpha-type carbonic anhydrases from a diverse set of taxa were obtained from the NCBI Protein database (identifier given at the end of each name) and aligned with MUSCLE. Critical histidine residues corresponding to the *H. sapiens* CA-II H94, H96, and H119 are highlighted in red, and other conserved residues are shown in blue. *N. oceanica CAH1* (6698-mRNA) is marked by the red dashed boxes.

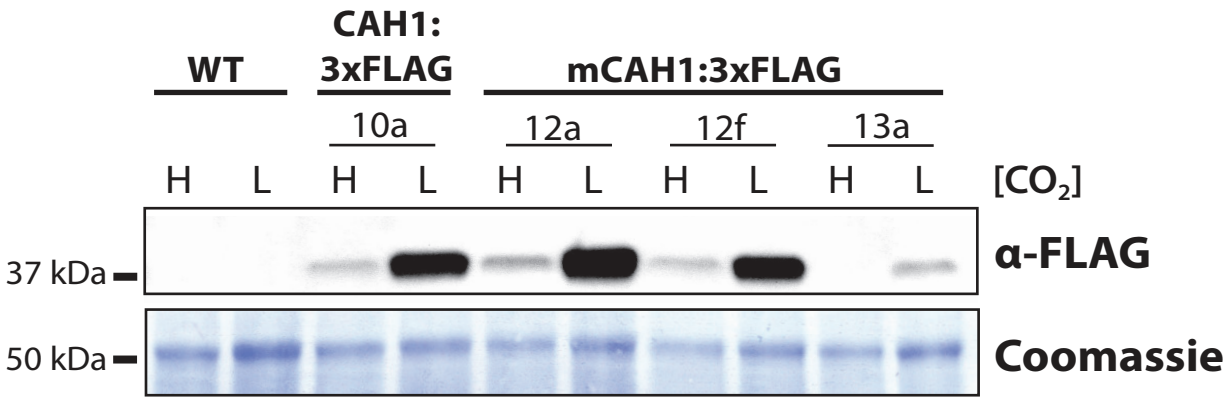

Fig. S4. Immunoblot analysis of site-directed mutant (H177A) mCAH1:FLAG fusion protein. To confirm that the mutagenized complementation protein was being expressed, total cell lysates were obtained from cells either acclimated to ambient CO_2 or kept at 3% CO_2. Both the control CAH1:3xFLAG and mCAH1:3xFLAG constructs are driven by the native *CAH1* promoter region. H and L labels indicate whether cells were kept at high (3%) or low (0.04%) CO_2, respectively, for 24 h prior to protein extraction. Expected protein size = 38 kDa. The ~55 kDa band shown in the image of the Coomassie blue stained membrane corresponds to the large subunit of Rubisco (62).

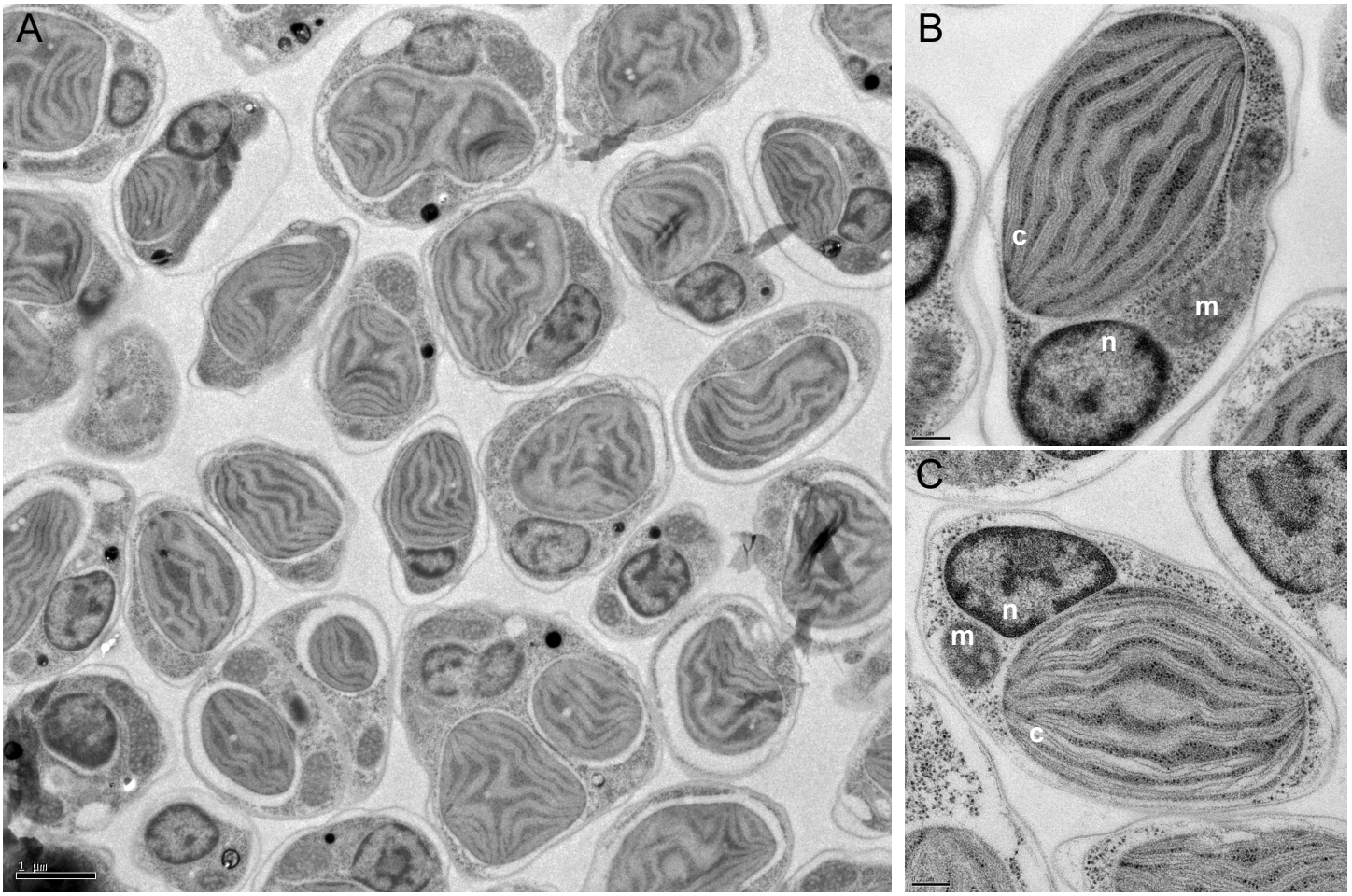

Fig. S5. Transmission electron microscopy of *N. oceanica* cells. Cells were preserved with high-pressure freezing, fixed, and imaged by TEM. (*A*) Scale bar equals 1 µm. (*B*) and (*C*) Scale bar equals 0.2 µm. n = nucleus, c = chloroplast, m = mitochondrion.

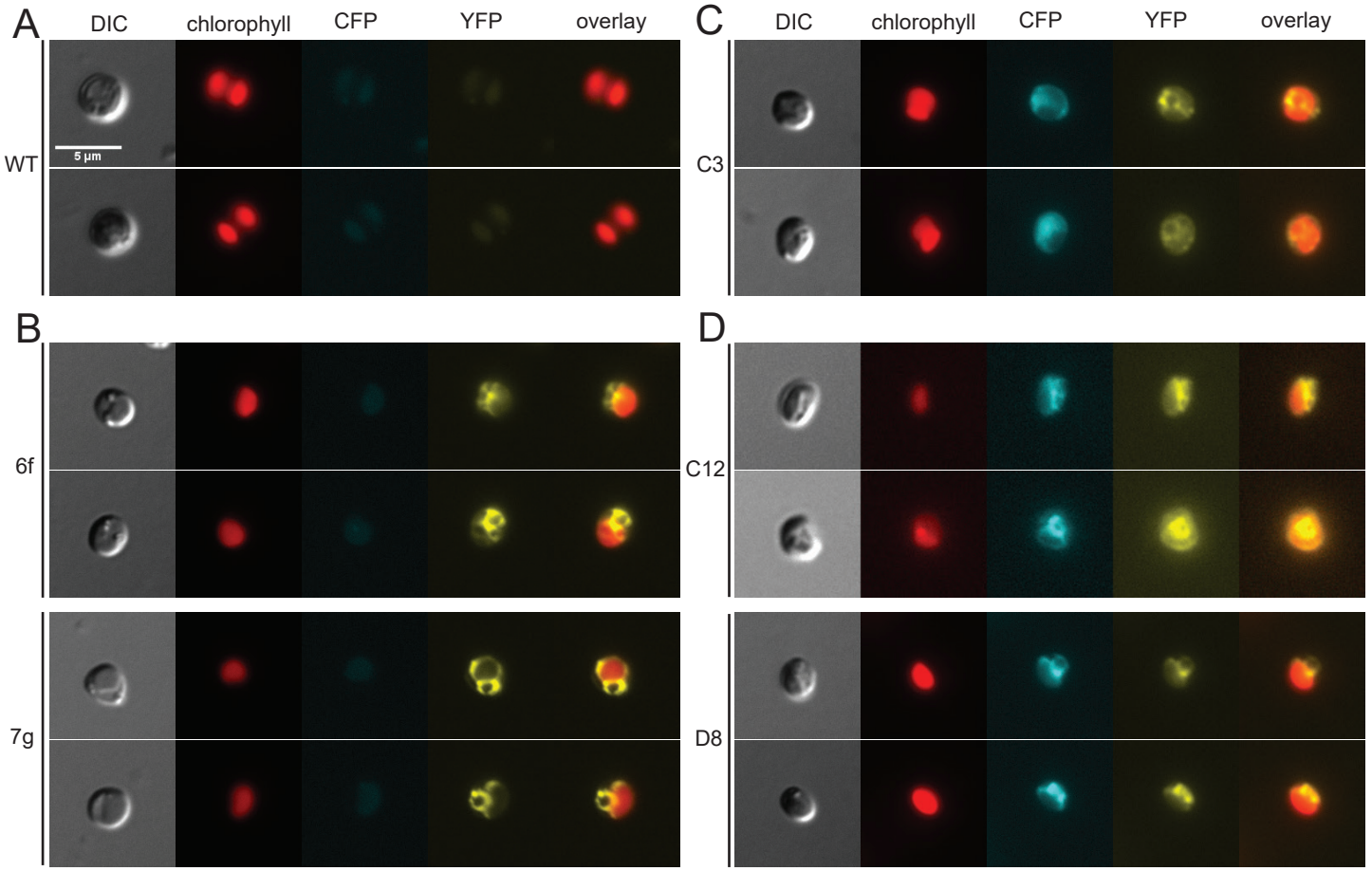

Fig. S6. Subcellular localization of CAH1:Venus fusion protein- additional lines and images. (*A*) Wild type (WT) cells shown for background autofluorescence. (*B*) The *cah1* mutant was transformed with a CAH1:Venus fusion construct (see complementation in Fig.1) to produce lines 6f and 7g. (*C*) The CAH1:Venus line 6f was further transformed with a cytoplasmic mCerulean (CFP) marker to produce line C3. (*D*) CAH1:Venus line 6f was also transformed with mCerulean targeted to the ER lumen with the signal peptide of *PROTEIN DISULFIDE ISOMERASE* (10287-mRNA) and a C-terminal KDEL retention signal to produce lines C12 and D8. The scale bar (5 µm) in (*A*) applies to all images, and all images of a given channel were acquired and processed in the same manner. Overlay refers to YFP and chlorophyll channels. Chlorophyll autofluorescence was imaged using a Cy5 filter set.

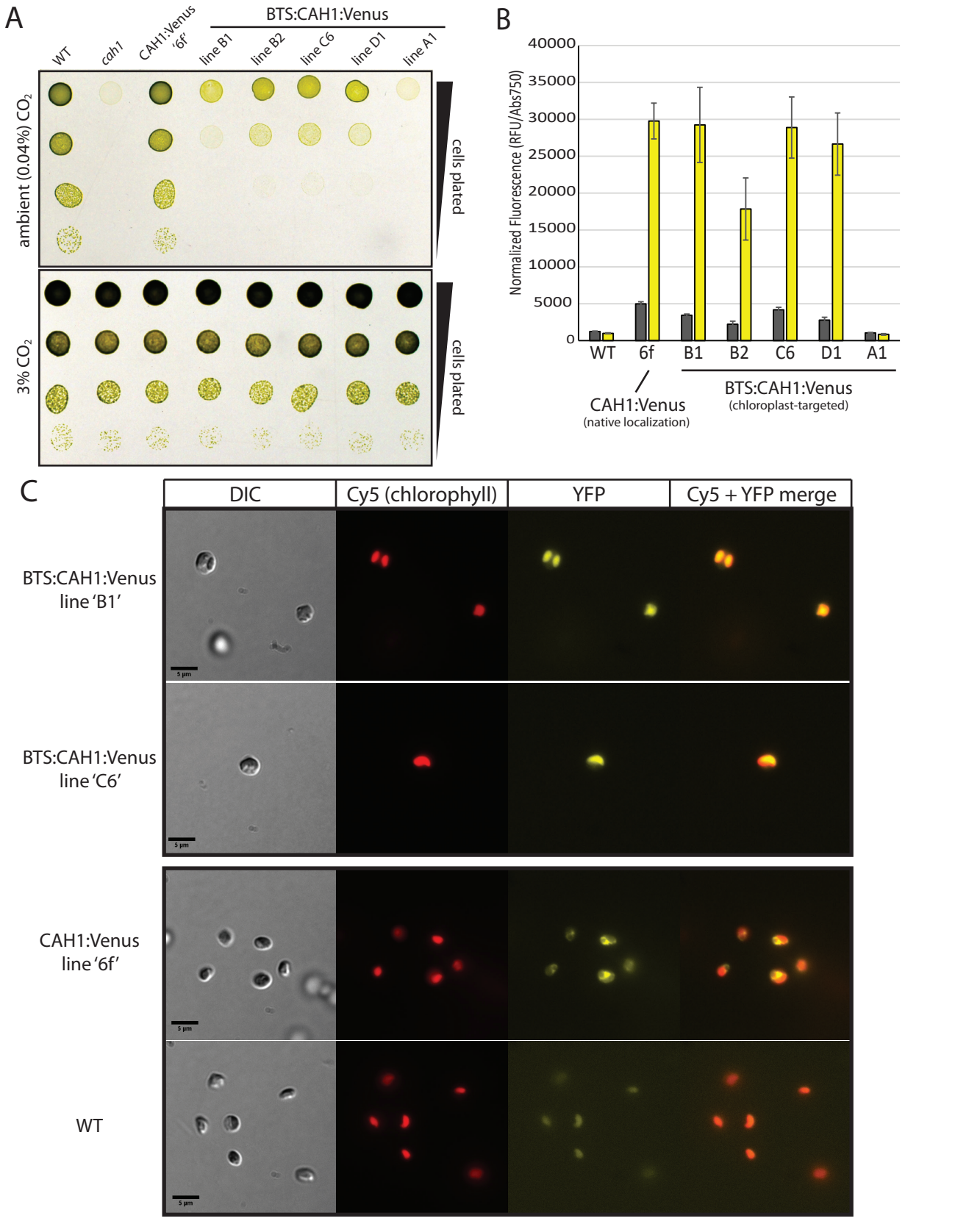

Fig. S7. Mis-targeting CAH1 to the chloroplast stroma only partially complements the *cah1* mutant. The bipartite targeting sequence (BTS) from a chloroplast protein (*VCP1*, 8367-mRNA) was cloned in-frame with the CDS of CAH1:Venus fusion and transformed into the *cah1* mutant. (*A*) Spot growth assay of wild type (WT), *cah1* mutant, *cah1* mutant + CAH1:Venus control complementation, and five independent BTS:CAH1:Venus lines are shown. BTS:CAH1:Venus lines showed intermediate growth greater than the *cah1* mutant, but clearly less than WT or the control complementation. (*B*) To confirm expression, Venus fluorescence was quantified by microplate reader. Line A1 shows very little growth and little expression. Dark gray bars correspond to the control plate kept at 3% CO_2 and the yellow bar to a duplicate plate transferred to ambient (0.04%) CO_2 for 24 h. Error bars denote s.d. (n = 3). (*C*) Subcellular localization of BTS:CAH1:Venus visualized by fluorescence microscopy. Note that the background autofluorescence from the chloroplasts in wild-type cells in the Venus channel overlaps entirely with the chlorophyll signal, which is distinctly dimmer than the BTS:CAH1:Venus lines and differently localized than the CAH1:Venus example.

Table S1 BLAST results for carbonic anhydrase families against the *Nannochloropsis oceanica* CCMP1779 genome

		blastp			tblastn			
Query ID	Query Description	CCMP1779 Hit ID	Score	E-Value	CCMP1779 Hit Locus	CCMP1779 Hit ID	Score	E-Value
EDP00852.1	alpha carbonic anhydrase 3 [Chlamydomonas reinhardtii]	NannoCCMP1779_6698-mRNA-1 ("CAH1")	69	2.0E-12	nanno_840:3775..3338	NannoCCMP1779_6698-mRNA-1 ("CAH1")	57	2.0E-08
		NannoCCMP1779_4225-mRNA-1	27	5.2	nanno_1082:4249..4064	NannoCCMP1779_2742-mRNA-1	32	0.68
		NannoCCMP1779_11833-mRNA-1	27	6.7	nanno_5174:4628..4488	NannoCCMP1779_8366-mRNA-1	31	0.88
EDO96058.1	mitochondrial carbonic anhydrase 4, beta type [Chlamydomonas reinhardtii]	NannoCCMP1779_11263-mRNA-1	87	4.00E-18	nanno_4923:16715..16966	NannoCCMP1779_11263-mRNA-1	48	2.00E-13
		NannoCCMP1779_11083-mRNA-1	29	0.84	nanno_1382:2391..2248	NannoCCMP1779_7446-mRNA-1*	35	0.064
		NannoCCMP1779_4842-mRNA-1	28	1.9	nanno_604:7938..8351	NannoCCMP1779_11538-mRNA-1	31	0.71
		NannoCCMP1779_4442-mRNA-1 (partial) †						
2FGY_B	Beta/epsilon Carbonic Anhydrase [Halothiobacillus Neapolitanus] (Csosca)	NannoCCMP1779_1779-mRNA-1	29	2.1	nanno_2846:2190..2005	NannoCCMP1779_5259-mRNA-1	32	1.1
		NannoCCMP1779_11006-mRNA-1	27	7.9	nanno_228:57121..56912	NannoCCMP1779_517-mRNA-1	31	2.4
		NannoCCMP1779_8418-mRNA-1	27	7.9	nanno_4735:250..95	NannoCCMP1779_7957-mRNA-1	30	5.3
1QRM_A	Gamma-carbonic anhydrase [Methanosarcina thermophila]	NannoCCMP1779_8809-mRNA-1	50	3.00E-07	nanno_300:667..593	NannoCCMP1779_3997-mRNA-1	28	5.5
		NannoCCMP1779_2119-mRNA-1	47	4.00E-06	nanno_102:599..525	NannoCCMP1779_3871-mRNA-1	28	5.5
		NannoCCMP1779_6938-mRNA-1	30	0.47	nanno_564:92797..92639	NannoCCMP1779_4394-mRNA-1	27	7.1
		NannoCCMP1779_6720-mRNA-1	28	1				
BAO52719.1	delta carbonic anhydrase [Thalassiosira pseudonana CCMP1335]	NannoCCMP1779_8982-mRNA-1	27	7.2	*no hits found*	*n/a*	*n/a*	*n/a*
3BOJ_A	zeta cadmium-binding carbonic anhydrase CDCA [Thalassiosira weisflogii]	NannoCCMP1779_8530-mRNA-1	27	3.9	nanno_956:3596..3715	NannoCCMP1779_5047-mRNA-1	29	2.5
		NannoCCMP1779_1777-mRNA-1	27	3.9	nanno_1046:70780..70682	NannoCCMP1779_10085-mRNA-1	28	3.2
		NannoCCMP1779_1277-mRNA-1	26	5.1				
ABG38184.1	LciB [Chlamydomonas reinhardtii]	NannoCCMP1779_6781-mRNA-1 **	35	0.023	nanno_196:7269..7412	NannoCCMP1779_2140-mRNA-1	29	5.5
		NannoCCMP1779_4632-mRNA-1	28	4.8	nanno_870:3895..4269	NannoCCMP1779_9748-mRNA-1	29	7.2
		NannoCCMP1779_8079-mRNA-1	27	8.2	nanno_786:17138..16989	NannoCCMP1779_11636-mRNA-1	28	9.4
BAV00142.1	Pt43233 (LCIB homolog) [Phaeodactylum tricortunum]	NannoCCMP1779_10772-mRNA-1	30	1.1	nanno_1063:24101..24337	NannoCCMP1779_6238-mRNA-1***	30	0.11
		NannoCCMP1779_2725-mRNA-1	29	2.6	nanno_587:80416..80252	NannoCCMP1779_9320-mRNA-1	32	1.3
		NannoCCMP1779_6341-mRNA-1	28	4.4	nanno_82:1215..1355	NannoCCMP1779_11568-mRNA-1	31	2.3

*Pfam prediction = CobN/Magnesium chelatase, top hit to blastp against NCBI non-redundant protein = ADI46571.1 magnesium chelatase subunit H, partial [Vaucheria litorea].

**Pfam prediction = methyltransferase PF02636 (e-value = 3.2e-63)

***No Pfam prediction, only blastp hit e-value <1 in NCBI non-redundant database were *N. gaditana* genes "hypothetical protein Naga_100346g1 (4E-46) and hypothetical protein Naga_103298g1 (7E-6)"

† Partial sequence identified in Wei *et al.*, 2016. Plant Journal. Full sequence unknown due to contig edge.

Table S1. Top blastp and tblastn hits for different carbonic anhydrase homologs in *N. oceanica* CCMP1779. The protein sequence for each query (NCBI accession number shown for identification) was obtained from the NCBI protein database, and blastp/tblastn was used to search for homologs of the predicted proteins in *N. oceanica*. Queries were selected to find homologs of the carbonic anhydrase families including the alpha, beta, epsilon, gamma, delta, zeta and theta types. Hits with very low E-values are highlighted. Several other hits with E-values < 1 (asterisks) were investigated further using Pfam domain prediction to clarify the potential function of the candidate. NannoCCMP1779_4442-mRNA-1 was identified elsewhere (33) as a beta-type carbonic anhydrase, but was not identified in our blast searches, possibly because the gene is fragmented in the current v1 assembly.