
Article:
Mackinder, Luke orcid.org/0000-0003-1440-3233, Chen, Chris, Leib, Ryan et al. (6 more authors) (2017) A spatial interactome reveals the protein organization of the algal CO2 concentrating mechanism. Cell. e14. pp. 133-147. ISSN 1097-4172

https://doi.org/10.1016/j.cell.2017.08.044

A Spatial Interactome Reveals the Protein Organization of the Algal CO_2 Concentrating Mechanism

Authors:

Luke C.M. Mackinder[1,2], Chris Chen[1,3], Ryan D. Leib[4], Weronika Patena[1], Sean R. Blum[5], Matthew Rodman[3], Silvia Ramundo[6], Christopher M. Adams[4], and Martin C. Jonikas[1,7,*]

[1]Department of Plant Biology, Carnegie Institution for Science, Stanford, CA 94305, USA

[2]Department of Biology, University of York, York, YO10 5DD, UK

[3]Department of Biology, Stanford University, Stanford, CA 94305, USA

[4]Stanford University Mass Spectrometry, Stanford University, Stanford, CA, USA

[5]Department of Biomolecular Engineering, UC Santa Cruz, Santa Cruz, CA 95064, USA

[6]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94158, USA

[7]Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA

[*] Correspondence: mjonikas@princeton.edu

Highlights :

- Localizations and physical interactions of candidate CCM proteins were determined
- The data reveal three previously un-described pyrenoid layers and 89 pyrenoid proteins
- Plasma membrane inorganic carbon transporters LCI1 and HLA3 form a complex
- Carbonic anhydrase 6 localizes to the flagella, changing the model of the CCM

SUMMARY

Approximately one-third of global CO_2 fixation is performed by eukaryotic algae. Nearly all algae enhance their carbon assimilation by operating a CO_2 concentrating mechanism (CCM), built around an organelle called the pyrenoid, whose protein composition is largely unknown. Here, we developed new tools in the model alga *Chlamydomonas reinhardtii* to determine the localizations of 135 candidate CCM proteins, and physical interactors of 38 of these proteins. Our data reveal the identity of 89 pyrenoid proteins, including novel Rubisco-interacting proteins, photosystem I assembly factor candidates and inorganic carbon flux components. We identify three previously un-described protein layers of the pyrenoid: a plate-like layer, a mesh layer and a punctate layer. We find that the carbonic anhydrase CAH6 is in the flagella, not in the stroma that surrounds the pyrenoid as in current models. Together, these results provide an overview of proteins operating in the eukaryotic algal CCM, a key process that drives global carbon fixation.

Keywords: CO_2-concentrating mechanism, CCM, carbon-fixation, *Chlamydomonas reinhardtii*, photosynthesis, pyrenoid, Rubisco, high-throughput fluorescence protein tagging, affinity purification mass spectrometry

INTRODUCTION

Over the past three billion years, the carbon-fixing enzyme Rubisco drew down atmospheric concentrations of CO_2 to trace levels (Dismukes et al., 2001), in effect starving itself of its substrate. In parallel, the oxygenic reactions of photosynthesis have caused the appearance of abundant O_2, which competes with CO_2 for the active site of Rubisco and results in a loss of fixed CO_2 via photorespiration (Bauwe et al., 2010). To overcome these challenges of CO_2 assimilation in today's atmosphere, many photosynthetic organisms increase CO_2 levels in the vicinity of Rubisco by operating CO_2 concentrating mechanisms (CCMs). Such mechanisms increase the CO_2:O_2 ratio at the active site of Rubisco, enhancing CO_2 fixation and decreasing photorespiration. CCMs are found in nearly all marine photoautotrophs, including cyanobacteria and eukaryotic algae (Reinfelder, 2011), which together account for approximately 50% of global carbon fixation (Behrenfeld et al., 2001; Field et al., 1998).

In the alpha and beta cyanobacterial CCMs, inorganic carbon in the form of bicarbonate (HCO_3^-) is pumped into the cytosol to a high concentration (Price and Badger, 1989). This HCO_3^- is then converted into CO_2 in specialized icosahedral compartments called carboxysomes, which are packed with Rubisco (Shively et al., 1973). The components of the cyanobacterial CCMs have largely been identified, facilitated in part by the organization of the genes encoding them into operons (Price et al., 2008). Knowledge of these components has enabled the detailed characterization of the structure and assembly pathway of the beta carboxysome (Cameron et al., 2013; Rae et al., 2013).

Analogous to the cyanobacterial CCM, the eukaryotic green algal CCM concentrates HCO_3^- in a microcompartment containing tightly-packed Rubisco, called the pyrenoid. The pyrenoid is located in the chloroplast, surrounded by a starch sheath and traversed by membrane tubules that are continuous with the surrounding photosynthetic thylakoid membranes (Engel et al., 2015). Associated with the pyrenoid tubules is a carbonic anhydrase that converts HCO_3^- to CO_2 for fixation by Rubisco (Karlsson et al., 1998; Sinetova et al., 2012).

66 The mechanism of delivery of HCO_3^- to the pyrenoid thylakoids remains unknown. In contrast to
67 the prokaryotic CCM, the protein composition of the eukaryotic algal CCM and the structural
68 organization of the pyrenoid remain largely uncharacterized.
69 We reasoned that we could make rapid advances in our understanding of the algal CCM
70 by systematically determining the localizations and physical interactions of a large number of
71 candidate proteins. High-throughput protein localization and physical interaction studies have
72 rapidly advanced our understanding of cellular structure and processes in yeast (Huh et al.,
73 2003; Krogan et al., 2006), *Drosophila melanogaster* (Guruharsha et al., 2011), *Caenorhabditis*
74 *elegans* (Sarov et al., 2012) and mammalian cell lines (Sowa et al., 2009). They have also been
75 implemented in higher plants (Tian et al., 2004), with significant success using transient
76 expression in *Arabidopsis thaliana* cell cultures (Koroleva et al., 2005). By comparison, due to
77 poorly understood challenges with nuclear transgene expression, the throughput of protein
78 localization and identification of physical interactions in algae has lagged far behind. Indeed, to
79 our knowledge, the largest number of endogenous proteins localized in algae by a single study
80 is 11 (Kobayashi et al., 2016).
81 In this study, we developed a high-throughput fluorescence protein tagging and affinity
82 purification mass spectrometry (AP-MS) pipeline for the model green alga *Chlamydomonas*
83 *reinhardtii* (Figure 1A). With this pipeline, we determined the localizations of 135 candidate CCM
84 proteins and the physical interactions of 38 core CCM components. Our microscopy data
85 reveals an unexpected localization for the carbonic anhydrase CAH6, identifies three previously
86 undescribed pyrenoid protein layers, and suggests that the pyrenoid shows size selectivity for
87 stromal proteins. The AP-MS data produce a spatially resolved protein-protein interaction map
88 of the CCM and pyrenoid, identifying novel protein complexes including a complex between
89 inorganic carbon transporters LCI1 and HLA3, and suggesting CCM functions for multiple
90 proteins. These results transform our basic knowledge of the eukaryotic CCM and advance the

prospects of transferring this system into higher plants to improve crop production (Atkinson et al., 2016; Long et al., 2015).

RESULTS AND DISCUSSION

We Developed a High-Throughput Pipeline for Systematic Localization of Proteins in *Chlamydomonas*

To allow the parallel cloning of hundreds of genes, we designed an expression cassette that enabled high-throughput seamless cloning via Gibson assembly (Gibson et al., 2009). Open reading frames (ORFs) were amplified by PCR from genomic DNA and cloned in frame with a C-terminal Venus YFP and a 3xFLAG epitope, driven by the strong PsaD promoter. These constructs were transformed into wild-type *Chlamydomonas*, where they inserted into random locations in the genome (Figure 1B). To allow dual tagging of different proteins in the same cell, we developed a second expression vector with an mCherry fluorophore and a hygromycin selection marker (Figure S1A). Potential caveats of our system include loss of the endogenous transcriptional regulation of the protein, including information encoded in the promoter, terminator and genomic locus. Additionally, the C-terminal protein tag could obscure subcellular targeting signals or disrupt functional domains.

A common challenge with the use of fluorescent proteins in *Chlamydomonas* is that only a small percentage of antibiotic-resistant transformants successfully express the tagged protein. To enable the screening of hundreds of *Chlamydomonas* lines per tagged construct, we screened colonies for fluorescence directly on the transformation plates using a fluorescence scanner (Figure 1B).

Our Data Reveal Guidelines for Protein Localization in *Chlamydomonas*

Given the notorious difficulties with expressing tagged genes in *Chlamydomonas* (Fuhrmann et al., 1999; Neupert et al., 2009; Rasala et al., 2012), we started with the

understanding that only a fraction of our target proteins will be localized. We reasoned that the fraction of proteins amenable to this approach would likely provide new insights if a sufficiently large number of proteins could be localized. We sought to tag genes from three sources: 1) genes currently thought to be involved in the CCM, including previously characterized CCM genes (See review: Wang et al., 2015); 2) candidate CCM genes, including those identified from both transcriptomic studies (Brueggeman et al., 2012; Fang et al., 2012; Yamano et al., 2008) and a proteomic analysis of the pyrenoid (Mackinder et al., 2016); and 3) organelle markers, using homologs of conserved, well-characterized yeast and plant marker proteins (Figure 1B and Table S1). We were able to determine the localizations of 146 out of the 624 target genes (23%).

We sought to leverage the large scale of this study to uncover factors that may contribute to cloning and tagging success in *Chlamydomonas*. We successfully cloned 298 of the 624 target genes (48%). Unsurprisingly, our cloning success rate decreased with gene size (Figure S1B), likely due to the 68% GC content of coding regions and the high repeat content of genes, which makes PCR amplification challenging. Intriguingly, cloning success was higher for genes with high expression levels (Figures S1C and D; $P = 4 \times 10^{-13}$, Mann Whitney U test), suggesting that intrinsic properties of a gene that influence endogenous expression may also affect polymerase activity.

After multiple attempts to transform the 298 successfully cloned genes into *Chlamydomonas*, we acquired protein localization data for 146 of them (49%). We found that the two main factors correlated with our ability to obtain localization data were: 1) high endogenous gene expression level (Figures S1E and F; $P = 6 \times 10^{-14}$, Mann Whitney U test) and 2) absence of upstream in-frame ATGs (Figure S1G; Cross, 2016). The failure to obtain localization data for genes with in-frame uATGs is most likely due to absence of the correct translational start site in the cloned construct, resulting in a truncated protein that can be functionally impaired, structurally unstable or lacking essential organelle targeting sequence(s).

These data suggest that transcript abundance is predictive for localization success and that future protein expression studies will benefit substantially from improved annotation of *Chlamydomonas* translation start sites.

146 Tagged Proteins Show 29 Distinct Localization Patterns

To aid in the classification of unknown proteins to subcellular regions, we tagged a series of conserved, well-characterized organelle and cellular structure proteins. This included proteins whose homologs localized to the ER (BIP1), Golgi (ARL6), mitochondria (MITC10), chloroplast envelope (TIC20), chloroplast stroma (FDX1), thylakoid lumen (CYC6), thylakoid membrane (PSAH), actin (IDA5), plasma membrane (ACA3), tonoplast (ATPvE) and cytosol (Venus only; Table S1). We then employed a decision tree (Figure 2A) to classify visually the localization of 135 additional proteins into 29 distinct subcellular regions, representing nearly all of the known organelles and cellular structures of *Chlamydomonas* (Figure 2B).

We observed a diverse range of cytosolic localizations, with subtle differences between localization patterns (Figure S2A). Some diffuse cytosolic proteins were excluded from the nuclear region. Another subset of cytosolic proteins had access to the nucleus, with signal either throughout the nucleus or in a circular intranuclear region. Several cytosolic proteins were enriched in the perinuclear zone. In addition, a subset of cytosolic proteins were difficult to localize either due to weak signals, or because of distinct nonhomogeneous patterns, and were classed as “Other”.

Interestingly, 12 proteins were not confined to one organelle but were seen in multiple compartments (Figure 2C and Table S2). Of the 10 proteins found in flagella, seven were also found in the cytosol, consistent with the known exchange of many flagellar components with the cytosol (Rosenbaum and Witman, 2002). Six proteins were found both in the chloroplast and in the cytosol, and two of these proteins additionally showed flagellar localization. If these multiple localizations are not artefacts of our expression system, they may represent proteins that

function in multiple compartments or are involved in inter-organelle signalling. The protein localizations from our study are available on a searchable website, https://sites.google.com/site/chlamyspatialinteractome/. These localization data and the availability of strains to the community will provide a useful resource for future gene characterization studies.

Localization Assignments Agree With Previous Studies for 39/41 Proteins

To evaluate the accuracy of our method, we compared our results with published localizations of individual proteins. Our data shared 25 proteins with the validated "training" set of chloroplast, mitochondria and secretory pathway genes from Tardif et al. (2012). Nearly all (24/25) matched our localization data, with the only exception being ACP2 (Cre13.g577100). Whereas we saw ACP2 in the chloroplast (Figure 2D), it was identified as mitochondrial in one of three proteomic studies of isolated *Chlamydomonas* mitochondria (Tardif et al., 2012). However, it was not detected in the mitochondria in another study (Atteia et al., 2009), and, in a third proteomics study, it was seen in approximately equal abundances in isolated chloroplasts and mitochondria (Terashima et al., 2010). Finally, ACP2 is predicted to be chloroplast-localized by PredAlgo (Tardif et al., 2012). Overall, the ambiguity in the published data leave open the possibility that our localization data may in fact be correct. Beyond the reference set of Tardif et al. (2012), we compared our data with previously published localization of CCM components, and found that 15 of 16 localizations matched. The strong overlap with previously known localizations indicates that our dataset is of high quality (>95% accurate) and that C-terminal tagging of *Chlamydomonas* proteins results in minimal localization artefacts.

CAH6 Localizes to the Flagella

Carbonic anhydrases, which catalyse the reversible reaction of HCO_3^- to CO_2, play a critical role in CCMs (Badger, 2003). Our successful localization of nine *Chlamydomonas* carbonic

anhydrases show that they are found in a diverse range of cellular locations (Figure S2B). In all current models of the CCM (Moroney et al., 2011; Wang et al., 2015), the carbonic anhydrase CAH6 is in the chloroplast stroma, where it has been proposed to convert CO_2 back to HCO_3^-.

Surprisingly, in our study, CAH6 localized to the flagella in two independent transformation lines (Figure 2D and S2B), and produced no detectable signal in the chloroplast. To exclude the possibility that our observation is due to an artefact (e.g. due to the C-terminal Venus tag), we analysed the localization of CAH6 in existing proteomic datasets. CAH6 is present in the flagellar proteome (Pazour et al., 2005) and has been shown to be an abundant intraflagellar transport (IFT) cargo (Engel et al., 2012), providing independent validation of CAH6 localization to the flagella. Additionally, CAH6 is absent from both the chloroplast proteome (Terashima et al., 2010) and the mitochondrial proteome (Atteia et al., 2009), further suggesting that levels in the chloroplast are low or non-existent.

Previous evidence for CAH6 in the stroma came from immunogold labeling experiments, in which Mitra et al. (2004) found a 4.7 fold enrichment of gold particles associated with chloroplast starch relative to control pre-immune serum. This could be an artefact due to cross-reactivity of the immunized serum with another epitope. Alternatively, CAH6 may be an abundant flagellar protein, but present at very low levels in the chloroplast.

The apparent absence of carbonic anhydrase in the stroma may be a requirement of the *Chlamydomonas* CCM. A stromal carbonic anhydrase could risk short-circuiting the CCM by promoting the release of CO_2 from HCO_3^- in areas that are not in close proximity to Rubisco. In fact, it has been shown that the expression of carbonic anhydrase in the cyanobacterial cytosol, the likely functional equivalent of the chloroplast stroma, results in the disruption of the cyanobacterial CCM (Price and Badger, 1989).

Instead of directly participating in the CCM, CAH6 could be involved in inorganic carbon sensing by *Chlamydomonas* flagella, which are found at the leading edge of swimming cells. *Chlamydomonas* was recently shown to demonstrate chemotaxis towards HCO_3^- (Choi et al.,

2016) and their flagella have been found to contain mechanosensors (Fujiu et al., 2011). In other eukaryotes, flagella are known to contain sensors for a range of signals (Zimmerman and Yoder, 2015). Carbonic anhydrases have previously been implicated in inorganic carbon sensing in multiple biological processes including stomatal opening in higher plants (Hu et al., 2010), sour taste receptors in the tongue (Chandrashekar et al., 2009) and regulation of cilia beat frequency in lung epithelial cells (Tresguerres et al., 2010). Therefore, our evidence for the localization of CAH6 to the flagella has substantial implications for our understanding of the algal CCM.

PredAlgo is the Best Protein Localization Predictor for *Chlamydomonas*

The excellent agreement of our localization data with previous studies provided an opportunity to test the accuracy of the two main localization prediction algorithms used for *Chlamydomonas* proteins, PredAlgo (Tardif et al., 2012) and TargetP (Emanuelsson et al., 2000). For proteins that we observed in the chloroplast, PredAlgo predicted a chloroplast localization for 90% of them, whereas TargetP only predicted a chloroplast localization for 31% (Figure 2E). For mitochondrial proteins, the accuracy dropped to 31% for PredAlgo and 15% for TargetP. For secretory pathway proteins, the accuracy was 38% for PredAlgo and 24% for TargetP. The higher accuracy of PredAlgo is likely because the algorithm was trained using *Chlamydomonas* proteins, whereas TargetP was trained using higher plant proteins. These results highlight that PredAlgo is the best localization predictor for *Chlamydomonas* proteins, but its accuracy drops off significantly when proteins localize to compartments other than the chloroplast.

We Assigned 82 Proteins to 13 Sub-Chloroplast Locations

As expected, our study was highly enriched for proteins localized to the chloroplast. 56% (82/146) of our proteins localized to the chloroplast, compared with approximately 19% of all genes in the genome predicted to encode chloroplastic proteins. We assigned these 82 proteins

247 to 13 sub-chloroplast locations (Table S1; Figures 2A and 3A). Chloroplast envelope proteins
248 showed three subcategories of localization: 1) envelope homogeneous (signal observed evenly
249 throughout the chloroplast envelope); 2) envelope non-homogenous and; 3) envelope plus
250 chloroplast homogenous (signal observed throughout the chloroplast in addition to the
251 envelope). Tic20, a known component of the chloroplast translocon machinery, was enriched in
252 the chloroplast envelope on the nuclear side and at the cytosolic lobe junction in agreement with
253 previous immunofluorescence data (Schottkowski et al., 2012). By contrast, other chloroplast
254 envelope proteins such as Cre11.g467759 and CGLD28 were evenly distributed. LCIA (Low
255 CO_2 Inducible A), a putative HCO_3^- transporter (Yamano et al., 2015), and LCI20, a putative 2-
256 oxoglutarate/malate translocator (Johnson and Alric, 2013), both showed some homogeneous
257 chloroplast signal in addition to a clear envelope signal. Further biochemical analysis will be
258 required to confirm these dual localizations and to determine whether these proteins are
259 functional in both the chloroplast envelope and thylakoid membranes.
260 For three proteins, the signal was observed as punctate dots throughout the chloroplast
261 (Figure S3A). A protein with predicted 50S ribosome-binding GTPase activity (Cre12.g524950)
262 was seen as multiple puncta with enrichment in the translational zones (T-zones) located on
263 either side of the pyrenoid where chloroplast translation is thought to occur (Uniacke and
264 Zerges, 2009), supporting a potential role in chloroplast translation. Also showing a punctate
265 pattern with enrichment in the T-zone region was histone-like protein 1 (HLP1; Cre06.g285400),
266 which was shown to localize to chloroplast nucleoids found adjacent to the pyrenoid (Karcher et
267 al., 2009; Ris and Plaut, 1962). Finally, the fatty acid biosynthesis enzyme acetyl-CoA biotin
268 carboxyl carrier (BCC2; Cre01.g037850) also appears as several punctate dots including some
269 in the T-zone region. The similarity of the localization patterns of these proteins suggests that
270 chloroplast translation, chloroplast DNA and fatty acid synthesis may be co-localized in the
271 chloroplast.

We also analyzed the properties of chloroplast localized proteins and found that proteins with specific patterns of localization were often enriched in specific physical properties. As expected, all eight chloroplast envelope proteins contained one or more transmembrane domains (predicted by TMHMM v. 2.0; Krogh et al., 2001). Interestingly, proteins showing homogeneous chloroplast localization were enriched in transmembrane domains (Figure 3B), found in 9/14 homogeneous proteins vs 4/39 for chloroplast non-homogenous proteins (P = 0.0002, Fisher's exact test). This observation suggests that proteins with homogeneous localization are most likely thylakoid membrane-associated.

The Pyrenoid Appears to Show Selectivity to Stromal Contents

Because the pyrenoid is a non-membrane-bound organelle, its protein composition cannot be regulated by a membrane translocation step. We therefore sought to understand whether pyrenoid proteins are enriched for any specific physicochemical properties. We classified chloroplast localized proteins into two groups: 1) pyrenoid depleted, where the signal from the pyrenoid was weaker than the surrounding chloroplast and 2) not pyrenoid depleted, where the signal from the pyrenoid was comparable or brighter than the surrounding chloroplast. Interestingly, the two groups showed different protein molecular weight distributions (P = 0.001, Mann-Whitney U test). The 39 proteins that are not pyrenoid depleted are almost all smaller than ~50 kDa (Figure 3C; the value of ~50 kDa excludes the Venus YFP region, therefore the effective molecular weight is ~78kDa), suggesting that the pyrenoid may exclude larger proteins. One protein (Cre01.g030900) was larger than this cut-off. Analysis of its structure by Phyre2 predicts that it contains two transmembrane domains. It is possible that a substantial portion of this protein is in the thylakoid lumen, where it may not contribute to the size exclusion effect. Understanding the mechanisms behind protein selectivity of the pyrenoid could provide valuable insight into how other non-membrane-bound organelles control their protein composition.

298

299 **We Identified Multiple New Pyrenoid Components**

300 Electron microscopy-based techniques have shown that the *Chlamydomonas* pyrenoid contains
301 a dense matrix of Rubisco surrounded by a starch sheath and is traversed by membrane
302 tubules formed from merged thylakoids (Figure 4A; Engel et al., 2015). Currently, seven
303 proteins have been unambiguously localized to three different regions of the pyrenoid: the
304 pyrenoid matrix, periphery, and tubules. The pyrenoid matrix contains the Rubisco holoenzyme
305 (RBCS/RbcL; Lacoste-Royal and Gibbs, 1987), its chaperone Rubisco activase (RCA1; McKay
306 et al., 1991), essential pyrenoid component 1 (EPYC1; Cre10.g436550), a Rubisco linker
307 protein important for Rubisco packaging in the pyrenoid (Mackinder et al., 2016), and a protein
308 of unknown function (Cre06.g259100; Kobayashi et al., 2016). Under very low CO_2 conditions,
309 the LCIB/LCIC complex, whose role is still uncertain (Jin et al., 2016), is known to form puncta
310 around the pyrenoid periphery (Yamano et al., 2010). Recently, a calcium-binding protein, CAS,
311 has been shown to specifically localize to the pyrenoid tubules at low CO_2 (Wang et al., 2016).
312 Here, we identify seven additional pyrenoid-localized components and three previously un-
313 described sub-pyrenoid localization patterns (Figure 4B-D). These data offer insights into the
314 functional role of pyrenoid-localized components and provide necessary molecular details for
315 better characterization of sub-pyrenoid structures.

316

317 **The Pyrenoid Has at Least Four Distinct Outer Layers**

318 Our data suggest that the pyrenoid is surrounded by at least four distinct outer layers: 1) LCIB
319 and LCIC localize to puncta around the periphery; 2) PSBP4 (photosystem II subunit P4)
320 localizes to a different set of puncta; 3) STA2 (starch synthase 2) and SBE3 (starch branching
321 enzyme 3) localize to plate-like structures; and 4) LCI9 localizes to a mesh-like structure (Figure
322 4C-E).

Confocal microscopy indicated that LCIB, LCIC and PSBP4 showed punctate outer pyrenoid patterns, whereas SBE3, STA2 and LCI9 showed a more homogeneous distribution around the pyrenoid periphery (Figure 4B). Co-expression of LCIB-Venus and LCIC-mCherry showed that LCIB and LCIC were co-localized (Figure 4D), supporting the previous finding that they are part of the same complex in the stroma (Yamano et al., 2010).

PSBP4-Venus did not co-localize with LCIC-mCherry (Figure 4D), indicating that PSBP4 is in a different structure or complex. PPD1, the *Arabidopsis* homolog of PSBP4, has been shown to be in the thylakoid lumen (Liu et al., 2012). Therefore, the PSBP4 puncta likely represent proteins located in the thylakoid lumen. Consistent with this possibility, we also see a small amount of PSBP4-Venus signal within the pyrenoid, and this signal forms a network-like pattern reminiscent of pyrenoid tubules.

We studied the localization of STA2, a granule bound starch synthase, and SBE3, an enzyme involved in starch branching. Co-localization data indicated that STA2 was localized within the perimeter described by LCIC (Figure 4D). STA2 formed a clearly defined plate-like pattern around the pyrenoid core, which appeared to coincide with the location and shape of the starch sheath (Figure 4C). SBE3 also displayed this plate pattern, but was generally more diffuse than STA2 (Figure 4B). Our data suggest that both STA2 and SBE3 localize to the starch sheath.

Dual expression indicates that LCI9 is tightly apposed to the pyrenoid matrix and, like STA2, also localizes within the perimeter described by LCIC (Figure 4D). However, analysis of Z-sections showed that the pattern of LCI9 could be further distinguished from that of STA2 and SBE3 because LCI9 formed a mesh structure around the pyrenoid (Figure 4C). Intriguingly, the complementary localizations of STA2 and LCI9 suggest that LCI9 may be part of a protein layer that fills the gaps between the starch plates.

A Putative Methyltransferase Localizes to the Pyrenoid Matrix

We discovered that SMM7 (Cre03.g151650), a putative methyltransferase, localized to the pyrenoid matrix. This is intriguing because another putative methyltransferase, CIA6 (Cre10.g437829), was found to be required for pyrenoid assembly (Ma et al., 2011), although its localization was not determined. Unlike *CIA6*, *SMM7* is strongly transcriptionally upregulated under low CO_2 conditions (Brueggeman et al., 2012; Fang et al., 2012). Identification of the protein targets of CIA6 and SMM7 will likely provide critical insights into pyrenoid biogenesis and regulation.

Pyrenoid Tubules are Enriched in PSAH, a Component of Photosystem I

Traversing the pyrenoid are pyrenoid tubules, which are thought to deliver CO_2 at a high concentration to the matrix (Meyer et al., 2016). Previous work using immunogold labeling and photosystem (PS) I and PSII activity assays suggested that the pyrenoid tubules from several different algal lineages contain active PSI components and are depleted in PSII components (McKay and Gibbs, 1990; McKay and Gibbs, 1991; Mustardy et al., 1990). A hypothesis offered to explain these findings was that minimizing PSII activity in the pyrenoid would decrease O_2 release in the pyrenoid to minimize photorespiration (McKay and Gibbs, 1990). In contrast to these findings, we found that PSII components (PSBP3, PSBQ, PSBR) showed similar pyrenoid localization patterns to that of PSI (PSAG, PSAK and FDX1), cytochrome b_6f (CYC6) and ATP synthase (ATPC) components (Figure S3B). Even if assembled PSII is present in the tubules, its activity could be reduced, as has been previously suggested (McKay and Gibbs, 1991).

Strikingly, we found that unlike other PSI components, the PSI protein PSAH was enriched within the pyrenoid-tubules (Figure 4B). PSAH is a 130 amino-acid protein with a single transmembrane helix that in land plants binds to the core PSI at the site where light harvesting complex II (LHCII) docks in state transitions (Ben-Shem et al., 2003; Lunde et al., 2000). The enrichment of PSAH could indicate an additional, pyrenoid related, role for this

protein in algae. Together, our localization data for pyrenoid components allow us to propose a model for the spatial organization of the pyrenoid (Figure 4E).

We Generated a Spatially Defined Protein-Protein Interaction Network of the CCM

In an effort to understand the interconnectivity of the protein components of the CCM, we developed a large-scale affinity purification mass spectrometry (AP-MS) approach. The Venus-tagged fusion proteins contain a 3xFLAG tag, allowing for affinity purification of localized proteins. We chose 38 candidates for AP-MS, focusing on proteins previously implicated in the CCM and on those we found in the pyrenoid. We ran all 38 bait lines in duplicate as either biological, affinity purification or MS replicas (Table S3).

To aid in filtering out nonspecific bait-prey interactions from true interactions, we used ^{15}N labeling. We affinity purified baits and associated proteins from lines grown in ^{14}N media, and, before mass spectrometry, we mixed each sample with affinity-purified Venus-3xFLAG and associated proteins from lines grown in ^{15}N media. However, we found that $^{14}N/^{15}N$ ratios alone were not sufficient to identify true interactors from false positives. This was primarily due to ubiquitous contaminants showing variation between and across samples. This resulted in cases where contaminants had larger $^{14}N/^{15}N$ ratios than low abundant, but specific interactors.

To overcome this challenge, we adapted a method developed by Sowa et al. (2009) to determine a WD-score for each protein-prey interaction. The WD-score incorporates the reproducibility, specificity and abundance (in our approach we used ratio data) of each interaction (Figure 5A; see Experimental Procedures for further details). The WD-score is empirical by nature, so a cut-off value has to be generated. To determine a suitable cut-off value, we assumed that interactions between baits and preys in different organelles are nonspecific, and thus the distribution of their WD-scores approximates the distribution of WD-scores for false positive interactions. We took the highest WD-score value of 47.5 in this subset and used it as the WD-score cut-off. Approximately 3.8% of the interactions had WD-scores

above this value, giving 513 true positive interactions involving 398 proteins (Figure 5B and C). These proteins were considered high-confidence interacting proteins (HCIPs). This cut-off value is more stringent than in other studies in which a simulated dataset was used to determine a cut-off resulting in approximately 5% of data being determined as HCIPs (Behrends et al., 2010; Christianson et al., 2012; Sowa et al., 2009). One inherent limitation of AP-MS is that it cannot distinguish between direct and indirect interactions, for example this can result in large protein complexes being affinity purified even though a bait protein only directly interacts with one member of the complex.

We Used Several Approaches to Validate the Network

Given that PredAlgo was the best performing localization prediction software for *Chlamydomonas* (Figure 2E), we expected that the PredAlgo localization prediction for HCIPs would match the observed localization of the baits (Figure 5D and E). As expected, HCIPs of chloroplast-localized baits were enriched for proteins with predicted chloroplast localizations, and the same was true for mitochondria and ER/extracellular localized baits, which were enriched for predicted mitochondrial and secretory pathway localizations, respectively. HCIPs of nuclear, flagella and plasma membrane localized baits were enriched for the PredAlgo prediction “Other.”

Our network recapitulated previously known physical interactions (Figure 5F). Rubisco large subunit showed strong interactions with RBCS1 and RBCS2. EPYC1, a proposed Rubisco linker protein, is known to be in a complex with Rubisco (Mackinder et al., 2016), and here it showed strong physical interactions with both the Rubisco large subunit and RBCS2. Because RBCS1 and RBCS2 differ by only four amino acids, not all peptide fragments could be unambiguously assigned one of these specific isoforms. Finally, LCIB and LCIC have previously been shown to form a complex (Yamano et al., 2010), and in our dataset they are strongly associated with each other, with high reciprocal WD-scores >132 (Figure 5F).

To further validate this network, we performed a Gene Ontology (GO) enrichment analysis of HCIPs of baits localized to specific compartments (Figure 5G). HCIPs of baits from a specific compartment (i.e. chloroplast) are significantly enriched in GO function and localization terms related to that compartment, providing further support of the dataset quality.

HCIPs are Transcriptionally Co-Upregulated at Low CO_2

Cellular adaption to changing CO_2 requires transcriptional regulation of a large number of genes (Brueggeman et al., 2012; Fang et al., 2012). Most protein complexes display tight control of subunit stoichiometry, with subunit transcripts generally showing similar transcriptional patterns (Jansen et al., 2002). Analysis of transcriptional changes of baits and preys shows that most HCIPs show the same type of transcriptional regulation as their baits (Figure S4).

We Identified Many Novel Rubisco Interacting Proteins

Rubisco is thought to be the most abundant enzyme in the biosphere (Ellis, 1979), with its assembly and function extensively studied for several decades (Portis and Parry, 2007). *Chlamydomonas* has two nuclear encoded Rubisco small subunit proteins, RBCS1 and RBCS2, which are differentially regulated (Goldschmidt-Clermont and Rahire, 1986). It was formally possible that the four amino acid difference between the two proteins would lead to a difference in localization; however, we see that both localize to the pyrenoid (Figure 4B and D).

To identify novel protein complexes and new members of known complexes, we performed hierarchical clustering on HCIPs (Figure 6; see Figure S5 for all bait-prey interactions with a WD-score ≥1). The baits RBCS1 and RBCS2 clustered together and shared 15 HCIPs, four of which were also HCIPs of EPYC1. RBCS1- and RBCS2-associated proteins were enriched in uncharacterized proteins. Several of these interactors have homologs in other green algae but lack any conserved domains (Cre01.g054700, Cre01.g054850, Cre02.g088950, Cre16.g655050). Using the structural prediction software Phyre2, we found that Cre16.g655050

contains an N-terminal RbcX fold, which is found in a class of Rubisco chaperones, and the rest of the protein is predicted to be disordered (Figure S6). A BLAST analysis using Cre16.g655050 as the query showed that its full sequence is conserved in the closely related species *Volvox carteri* and *Gonium pectorale.* We also found that that the N-terminal RbcX-like region is conserved in several more evolutionarily distant Chlorophytes such as *Micromonas pusilla* (Table S4). In addition to Cre16.g655050, *Chlamydomonas* contains two copies of RbcX: RbcX-IIa (Cre01.g030350) and RbcX-IIb (Cre07.g339000). Functional characterisation of RbcX-IIa shows that it is a bona fide RbcX, binding to RbcL and aiding in Rubisco holoenzyme assembly (Bracher et al., 2015). Whether Cre16.g655050 is a chaperone for Rubisco or performs an alternative function is unknown.

Carbohydrate binding domains were found in three Rubisco interactors, including the two starch branching enzymes, SBE1 and SBE4, the latter of which also interacts with EPYC1. Given the concave shape of the pyrenoid-surrounding starch sheaths, there may be variation in starch synthesis and/or breakdown occurring between the two faces. One way to target a subset of starch metabolic enzymes to the inner concave face would be through an interaction with pyrenoid core proteins. The functional roles of the different SBE isoforms in *Chlamydomonas* have yet to be determined.

Interestingly, RBCS1 and RBCS2 interact with an ATP binding cassette (ABC) family transporter (Cre06.g271850). The specific role of this protein may help us elucidate transmembrane transport processes occurring across pyrenoid tubules.

EPYC1 Interacts with a Kinase and Two 14-3-3 Proteins

The putative Rubisco linker protein EPYC1 is phosphorylated at low CO_2 (Turkina et al., 2006). Interestingly, we see that EPYC1 associates with a predicted serine/threonine protein kinase (KIN4-2; Cre03.g202000). Understanding the role of this kinase may shed light on post-translational modifications associated with pyrenoid biogenesis and/or function.

EPYC1 interacts with two 14-3-3 proteins FTT1 and FTT2. 14-3-3 proteins are known to bind phosphorylated proteins; hence the interaction of 14-3-3 proteins with EPYC1 could potentially be regulated by the phosphorylation state of EPYC1. 14-3-3 proteins can influence the stability, function, interactions and localization of their targets (Chevalier et al., 2009). It is therefore possible that these 14-3-3 proteins are regulating an interaction between EPYC1 and Rubisco, possibly by changing the availability of protein-binding domains.

CAH3 Interacts with PSBP3, STT7 and Redox-Related Proteins

The carbonic anhydrase CAH3 is essential for the CCM (Karlsson et al., 1998) and is thought to convert HCO_3^- to CO_2 in the thylakoid membranes that traverse the pyrenoid, supplying the pyrenoid with a high concentration of CO_2. In our study, CAH3 associated with the TAT2 and TAT3 proteins of the twin-arginine translocation (Tat) pathway (Figure 6 and 7; Table S5), which delivers substrate proteins to the thylakoid lumen. This observation is consistent with work showing that CAH3 contains a predicted Tat signal peptide (Benlloch et al., 2015) and with previous biochemical studies suggesting that CAH3 localizes to the thylakoid lumen (Karlsson et al., 1998).

At low CO_2, CAH3 is phosphorylated, and this phosphorylation correlates with increased CA activity and localization to the pyrenoid (Blanco-Rivero et al., 2012). Here, we find that CAH3 has a strong interaction (WD-score = 209) with the kinase STT7 (Figure 6). The role of STT7 in LHCII phosphorylation and state transitions is well documented (Depège et al., 2003). However, it is unlikely that STT7 is directly phosphorylating CAH3, because the kinase domain of STT7 has been shown to be on the stromal side (Lemeille et al., 2009) and CAH3 is thought to be localized in the lumen (Karlsson et al., 1998). A direct interaction between STT7 and CAH3 may be occurring via the N-terminus of STT7, which is thought to be luminal via a single membrane traversing domain (Lemeille et al., 2009).

CAH3 also interacts with PSBP3 (Figures 6 and 7), a homolog of *Arabidopsis* PSBP-like 1 (PPL1). *Arabidopsis* PPL1 is involved in repair of photodamaged PSII (Ishihara et al., 2007). This interaction is intriguing in light of previous work indicating that CAH3 co-fractionates with PSII (Blanco-Rivero et al., 2012; Villarejo et al., 2002). However, because we did not detect other components of PSII, CAH3 may interact with PSBP3 without the rest of PSII.

PSBP4 is in a Complex with PSI Assembly Factors

PSBP4 is a PsbP domain (PPD)-containing protein whose *Arabidopsis* homolog is essential for photosystem I assembly and function (Liu et al., 2012). In our data, PSBP4 interacted with four proteins associated with PSI assembly: ycf3, ycf4, CGL71 and TAB2 (Heinnickel et al., 2016; Rochaix et al., 2004), suggesting that PSBP4 and these factors form a PSI assembly complex. PSBP4 also interacts with three uncharacterized conserved green lineage proteins (CGL30, CGL59 and CPLD12) and nine other proteins of unknown function (Figure 7). These interactions suggest that these uncharacterized proteins have roles in PSI assembly and function. Notably, PSBP4's localization suggests that PSI assembly occurs at the pyrenoid periphery.

The LCIB/LCIC Complex Interacts with Two Bestrophin-Like Proteins

Our data confirm that LCIB and LCIC, known stromal soluble proteins, are in a tight complex (Yamano et al., 2010). The *lcib* mutant has an "air-dier" phenotype: it exhibits WT growth in either very low CO_2 (0.01% CO_2 v/v) or high CO_2 (3% v/v), but dies in air levels of CO_2 (0.04%) (Wang and Spalding, 2006). The functional role of the LCIB/C complex is still unknown. This complex is hypothesized to act as either a CO_2 leakage barrier at the pyrenoid periphery or as a CO_2 recapture system, acting as a vectorial CO_2 to HCO_3^- conversion module to recapture CO_2 released from HCO_3^- by CAH3 in the thylakoid lumen (Duanmu et al., 2009). A role in the conversion of CO_2 to HCO_3^- is likely, as several homologs of LCIB were recently shown to be functional β-carbonic anhydrases. However, LCIB/C has no carbonic anhydrase function (Jin et

al., 2016), indicating that the complex may be tightly regulated or may require additional factors for proper function.

Both LCIB and LCIC interact with LCI11 (Cre16.g663450), and LCIC also interacts with Cre16.g662600 (Figure 6 and 7). Both LCI11 and Cre16.g662600 are putative bestrophins, which typically transport chloride but have been shown to be permeable to HCO_3^- (Qu and Hartzell, 2008). Furthermore, both proteins are upregulated at low CO_2 levels (Table S1 and Figure S4). LCI11 and Cre16.g662600 directly interact, and both also interact with another bestrophin-like protein, Cre16.g663400.

LCI9 Interacts with PFK1, PFK2 and SBE3 to Form a Carbohydrate Metabolism Module

As described above, LCI9 forms a mesh structure, likely in the gaps between starch plates. LCI9 contains two CBM20 (carbohydrate binding module 20) domains and is predicted to function as a glucan 1,4-α-glucosidase. Glucan 1,4-α-glucosidases hydrolyze glucosidic bonds, releasing glucose monomers from glucan chains. Therefore, LCI9 most likely plays a role in starch breakdown at the pyrenoidal starch plate junctions. AP-MS analysis shows that the strongest HCIPs of LCI9 are PFK1 and PFK2 (phosphofructokinases 1 and 2). PFK is a key regulator of glycolysis and is important for maintaining cellular ATP levels (Johnson and Alric, 2013). The exact metabolic role of an LCI9, PFK1 and PFK2 assemblage is still unclear. LCI9 also associates with SBE3, which in turn associates with STA3 and DPE2 (disproportionating enzyme 2), a putative α-1,4-glucanotransferase. Because SBE3 and its HCIPs are involved in starch synthesis and modification, components for control of starch breakdown and starch synthesis are potentially in close proximity, allowing tight regulation of starch structure and energy release. It should be noted that a potential caveat of performing AP-MS on proteins containing CBMs is that proteins could co-precipitate due to binding a common carbohydrate substrate, not due to direct protein-protein interactions.

Bicarbonate Transporters LCI1 and HLA3 Form a Complex with a P-type ATPase

HLA3 (high light activated 3) and LCI1 have both been implicated in HCO_3^- uptake at the plasma membrane (Ohnishi et al., 2010; Yamano et al., 2015). HLA3 is an ABC transporter, and its absence under low CO_2 conditions results in a reduced uptake of inorganic carbon by *Chlamydomonas* cells (Yamano et al., 2015). HLA3 expressed in *Xenopus* oocytes showed moderate uptake of HCO_3^- (Atkinson et al., 2016). LCI1 lacks any conserved functional or structural domains and contains four predicted transmembrane regions. Knock-down of LCI1 protein resulted in a small reduction in inorganic carbon uptake (Ohnishi et al., 2010); however, the function of LCI1 has not been demonstrated in a heterologous system.

Unexpectedly, we found that HLA3 and LCI1 are found together in a complex. The two proteins showed a reciprocal, strong interaction, each having WD scores >125. In addition, they appear to be in a complex with ACA4 (Autoinhibited Ca^{2+}-ATPase 4; Cre10.g459200), a P-type ATPase/cation transporter. Alignment of ACA4 with functionally characterized P-type ATPases shows that it is a member of the group IIIA family of P-type ATPases (Figure S7). Group IIIA members are known H^+-exporting ATPases (see Thever and Saier 2009). ACA4 may be aiding HCO_3^- uptake either by maintaining a H^+ gradient that HLA3 and/or LCI1 is using to drive HCO_3^- uptake, or by generating localized cytosolic alkaline regions similar to those that form near anion exchanger I during HCO_3^- uptake (Johnson and Casey, 2011). A localized alkaline region could decrease HCO_3^- to CO_2 conversion and hence diffusion out of the cell. HLA3 and LCI1 also share three other HCIPs: MRP2 (multidrug resistance protein 2), another ABC transporter; GFY5, which is transcriptionally upregulated upon acetate addition (Goodenough et al., 2014) and whose bacterial and fungal homologs are associated with acetate transport (Robellet et al., 2008; Sá-Pessoa et al., 2013); and Cre15.g635067, which contains a periplasmic phosphate-binding domain found in phosphate ABC transporters (Panther: PTHR30570).

The regulation of inorganic carbon transport is critical for the efficiency of the CCM. Recent work has shown that Ca^{2+} signalling is key for proper regulation of the CCM, with the

581 Ca^{2+}-binding protein CAS1 transcriptionally regulating HLA3 and other components (Wang et
582 al., 2016). One HCIP of HLA3 is an EF-hand-containing Ca^{2+}/calmodulin-dependent protein
583 kinase (Cre13.g571700), which could potentially regulate HLA3 post-translationally. Additionally,
584 HLA3 physically interacts with an adenylate/guanylate cyclase (CYG63: Cre05.g236650).
585 Adenylate and guanylate cyclases are known to play a role in sensing inorganic carbon across a
586 broad range of taxa including diatoms (Harada et al., 2006; Tresguerres et al., 2010). Thus,
587 Cre13.g571700 and Cre05.g236650 may represent another mode of CCM regulation, possibly
588 by sensing inorganic carbon availability at the plasma membrane.
589 HLA3 is a primary candidate for enhancing HCO_3^- uptake in higher plants (Atkinson et
590 al., 2016). Modeling shows that addition of HCO_3^- transporters to the chloroplast envelope
591 should enhance photosynthesis (McGrath and Long, 2014). The discovery that HLA3 and LCI1
592 are part of the same complex, and the identification of additional HLA3 and LCI1 HCIPs, may
593 aid in the assembly of a functional HCO_3^- transport module in higher plants.
594
595 **Perspective**
596 By developing an efficient fluorescent protein-tagging and AP-MS pipeline in *Chlamydomonas*,
597 we have generated a spatially defined network of the *Chlamydomonas* CCM. This large-scale
598 approach gives a comprehensive view of the CCM by revealing missing components, by
599 redefining the localization of others, and by identifying specific protein-protein interactions. Our
600 work also provides insight into the function and regulation of these known and newly discovered
601 CCM proteins, and represents a valuable resource for their further characterization.
602 Indeed, work building on this resource has already led to fundamental advances in our
603 understanding of the CO_2 concentrating mechanism. Early stages of this project identified the
604 EPYC1 protein as an abundant pyrenoid matrix-localized protein. Through in-depth
605 characterization, we found that EPYC1 acts as a molecular glue that links Rubisco holoenzymes
606 to form the pyrenoid matrix (Mackinder et al., 2016), solving the decades-old mystery of how

607 Rubisco is held together in the matrix. More recently, working with tagged lines produced in this
608 project, we discovered that the pyrenoid matrix is not solid but rather behaves as a liquid droplet
609 that mixes internally, and dissolves and condenses with the cell cycle (Freeman Rosenzweig et
610 al., in revision).
611 Our observation that the pyrenoid matrix appears to exclude proteins larger than ~78
612 kDa may be related to the liquid-like nature of the matrix. Interestingly, another liquid-like non-
613 membrane organelle, the *C. elegans* P granule, shows size exclusion of fluorescently labelled
614 dextrans 70 kDa and larger (Updike et al., 2011). This behavior may result from surface tension
615 generated by the proteins that produce the liquid phase (Bergeron-Sandoval et al., 2016)
616 Our approach for determining the localizations and physical interactors of candidate
617 proteins is a general strategy that can be used to quickly elucidate the protein composition and
618 organization of poorly understood organelles, cellular compartments and cellular processes.
619 When we began this project, our knowledge of the pyrenoid protein composition and structural
620 organization was extremely limited. We naively thought that the pyrenoid was primarily
621 composed of Rubisco and Rubisco activase, and therefore did not expect that any of the
622 transcriptionally identified candidate proteins would localize to the pyrenoid. We were surprised
623 when we not only identified new pyrenoid-localized proteins, but also observed proteins that
624 displayed previously undescribed patterns of localization. This included proteins localizing to the
625 starch sheath, a mesh surrounding the pyrenoid, and a new class of puncta at the pyrenoid
626 periphery. Beyond advancing our knowledge of the protein composition and structure of the
627 pyrenoid, our data reveal a possible size selectivity of the pyrenoid matrix.
628 Our results suggest changes to the existing model of inorganic carbon flux to the
629 pyrenoid (Figure 7). The apparent absence of carbonic anhydrase in the chloroplast stroma
630 aligns the *Chlamydomonas* CCM model more with the cyanobacterial model, in which the
631 absence of carbonic anhydrase in the cytosol is critical for inorganic carbon accumulation in the

form of HCO_3^- (Price and Badger, 1989; Price et al., 2008). The localization of the carbonic anhydrase CAH6 in flagella suggests potential roles in inorganic carbon sensing.

We anticipate that the tagged strains we have generated will be a valuable resource for the research community. These tagged strains not only provide markers for nearly every known organelle and sub-organelle compartment, but also provide molecular handles for characterizing organelle structure and function. The lines allow visual analysis of the dynamics of proteins and organelles on a wide range of time scales. Time-lapse imaging of fluorescently tagged strains will enable the exploration of protein and organelle relocalization in response to environmental changes, and fluorescence recovery after photobleaching (FRAP) experiments will allow the study of protein and organelle diffusion kinetics. More broadly, our tagging pipeline opens the door for a proteome-wide localization study in a photosynthetic organism.

Due to a rapidly rising global population and a finite agricultural land area, novel approaches are essential to maintain food security. One potential approach for improving yields is the transfer of a CCM into higher plants to increase CO_2 fixation rates (Long et al., 2015). Recent work has found that nearly all algal CCM proteins localize correctly in higher plants with no changes to their protein sequence, suggesting that the transfer of algal components could be relatively straightforward (Atkinson et al., 2016). However, engineering efforts were constrained by our limited knowledge of the components of the algal CCM. The work we present here provides a detailed blueprint of the algal CCM, revealing dozens of new targets for transfer into crop plants to improve carbon fixation, and enhancing our basic molecular understanding of a fundamental cellular process that drives global biogeochemical cycles.

SUPPLEMENTAL INFORMATION

Supplemental Information includes 7 figures and 5 tables.

AUTHOR CONTRIBUTIONS

L.C.M.M. and M.C.J. designed and supervised the study. L.C.M.M., C.C. and M.R. performed the cloning, L.C.M.M. did the microscopy and L.C.M.M. and C.C. carried out the AP-MS. S.R. and L.C.M.M. developed the affinity purification protocol. W.P. and S.R.B. provided bioinformatics support. R.L. and C.M.A. oversaw the mass spectrometry and peptide mapping. L.C.M.M., C.C. and M.C.J. analysed and interpreted the data. L.C.M.M created the figures. C.C. created the online viewing platform. L.C.M.M. and M.C.J. wrote the manuscript with input from all authors.

ACKNOWLEDGEMENTS

We thank the members of the Jonikas laboratory for helpful discussions, Z. Friedberg and R. Vasquez for help with gene cloning, A. Okumu for mass spectrometry sample preparation, H. Cartwright for confocal microscopy support, and U. Goodenough, R. Milo, A. Smith, N. Wingreen, T. Silhavy, M. Meyer and A. McCormick for helpful comments on the manuscript. Stanford University Mass Spectrometry is thankful to the NIH, Award Number S10RR027425 from the National Center For Research Resources for assistance in purchasing the mass spectrometer. The project was funded by National Science Foundation Grants EF-1105617 and IOS-1359682 (M.C.J.) and the Carnegie Institution for Science (L.C.M.M. and M.C.J.). The authors declare no conflicts of interest.

FIGURES

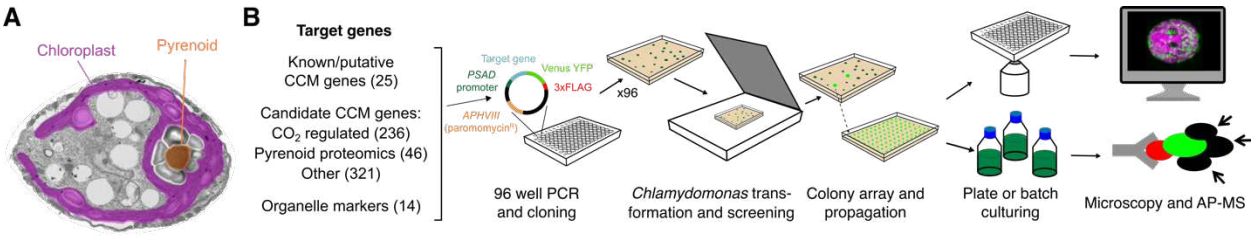
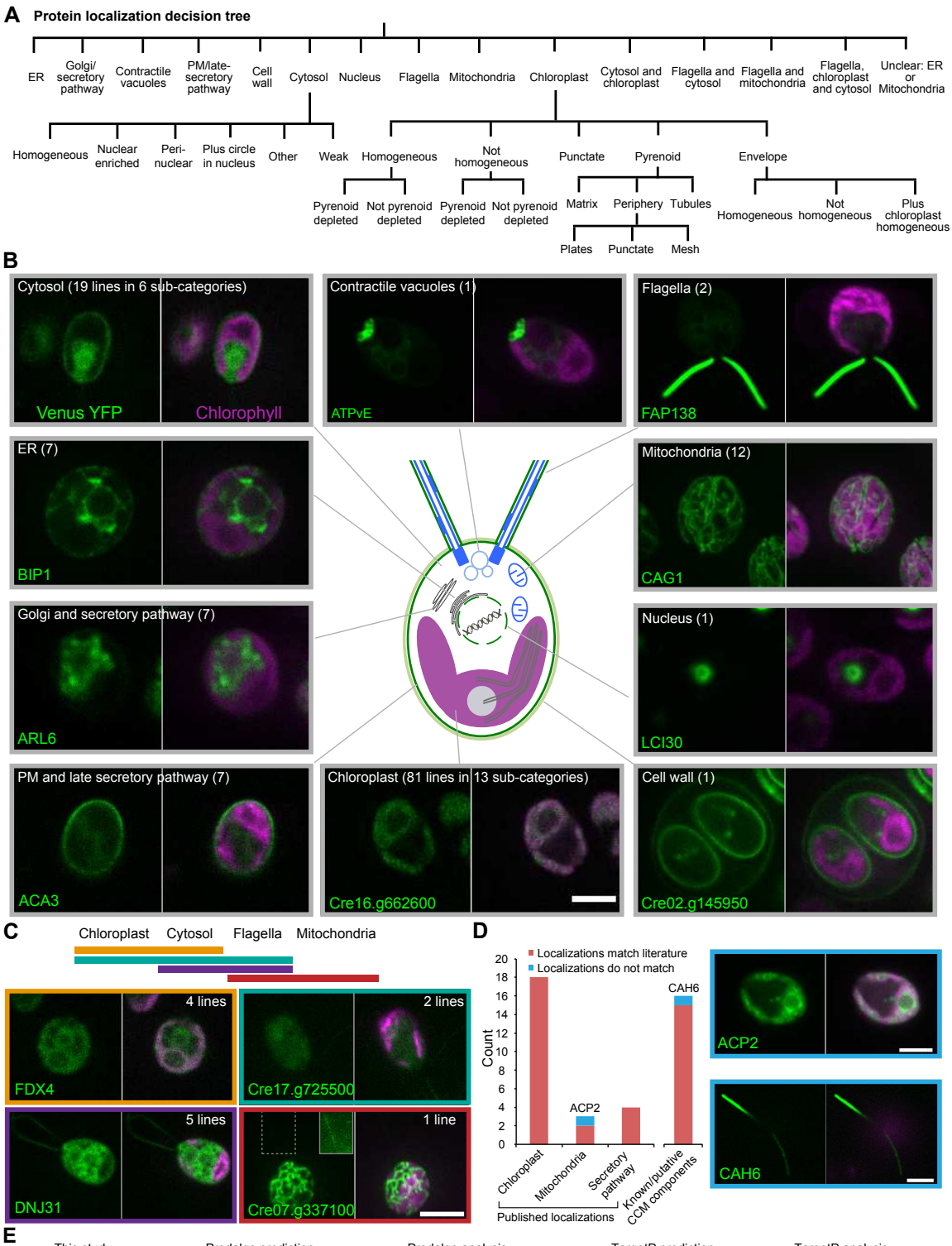


Figure 1. We Developed a High-Throughput Pipeline to Determine the Localization and Physical Interactions of Algal Proteins.

(A) A false-color transmission electron micrograph of a *Chlamydomonas reinhardtii* cell. The chloroplast is highlighted in magenta and the pyrenoid matrix in blue.

(B) Tagging and mass spectrometry pipeline. Target genes were amplified by PCR and Gibson assembled in frame with Venus-3xFLAG, under the constitutive *PSAD* promoter. Transformants were screened for fluorescence using a scanner, and arrayed to allow robotic propagation. Lines were either imaged using confocal microscopy to determine their spatial distribution or batch cultured for affinity purification-mass spectrometry (AP-MS).



E

This study		Predalgo prediction				Predalgo analysis			TargetP prediction				TargetP analysis		
		C	M	SP	O	Match	False negative rate	False positive rate	C	M	SP	O	Match	False negative rate	False positive rate
Chloroplast	67	60	3	2	2	90%	10%	19%	21	41	0	5	31%	69%	16%
Mitochondria	13	3	4	0	6	31%	69%	3%	4	2	0	7	15%	85%	39%
Secretory pathway	21	2	0	8	11	38%	62%	4%	0	3	5	13	24%	76%	1%
Other	41	15	0	2	24	59%	41%	15%	7	6	1	27	66%	34%	22%

Figure 2. Tagged Proteins Localized to a Diverse Range of Cellular Locations, and Revealed That CAH6 Localizes to Flagella.

(A) The decision tree used to assign proteins to specific subcellular locations.

(B) Representative images of proteins localized to different cellular locations. The number of different lines showing each localization pattern is in parentheses.

(C) Representative images of proteins that localized to more than one compartment. The solid outer line inset in the bottom right panel is an overexposed image, highlighting flagellar fluorescence, of the dashed line box.

(D) Comparison of our observations with published localizations. Images show the two proteins that did not match their published locations. All scale bars: 5 μm

(E) Comparison of our observations with localization prediction software programs PredAlgo and TargetP.

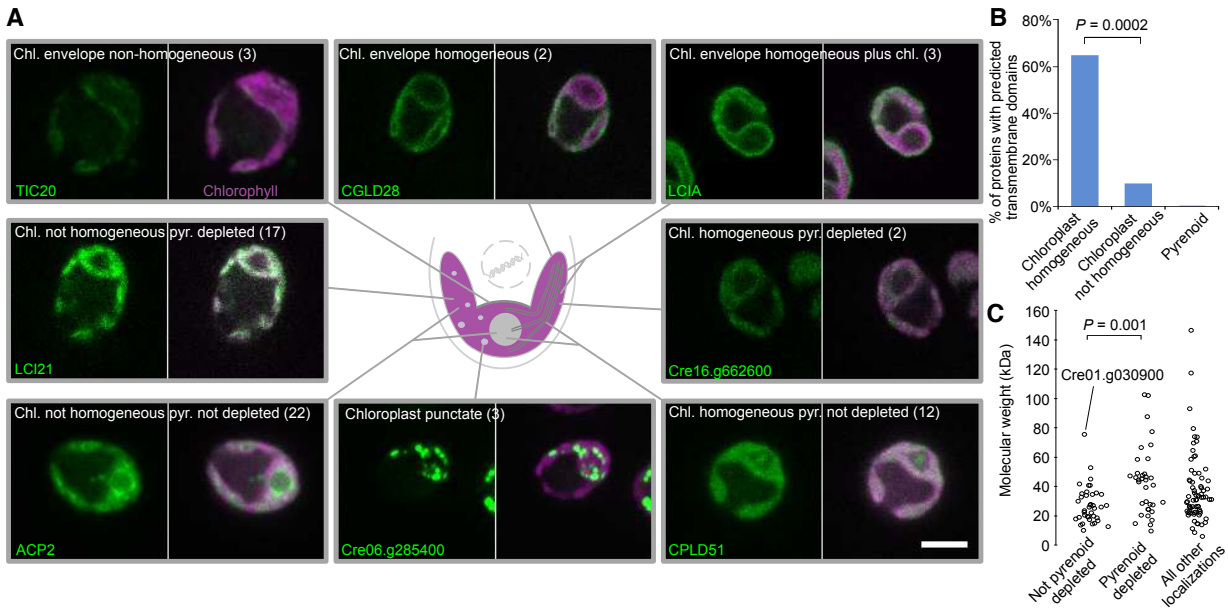
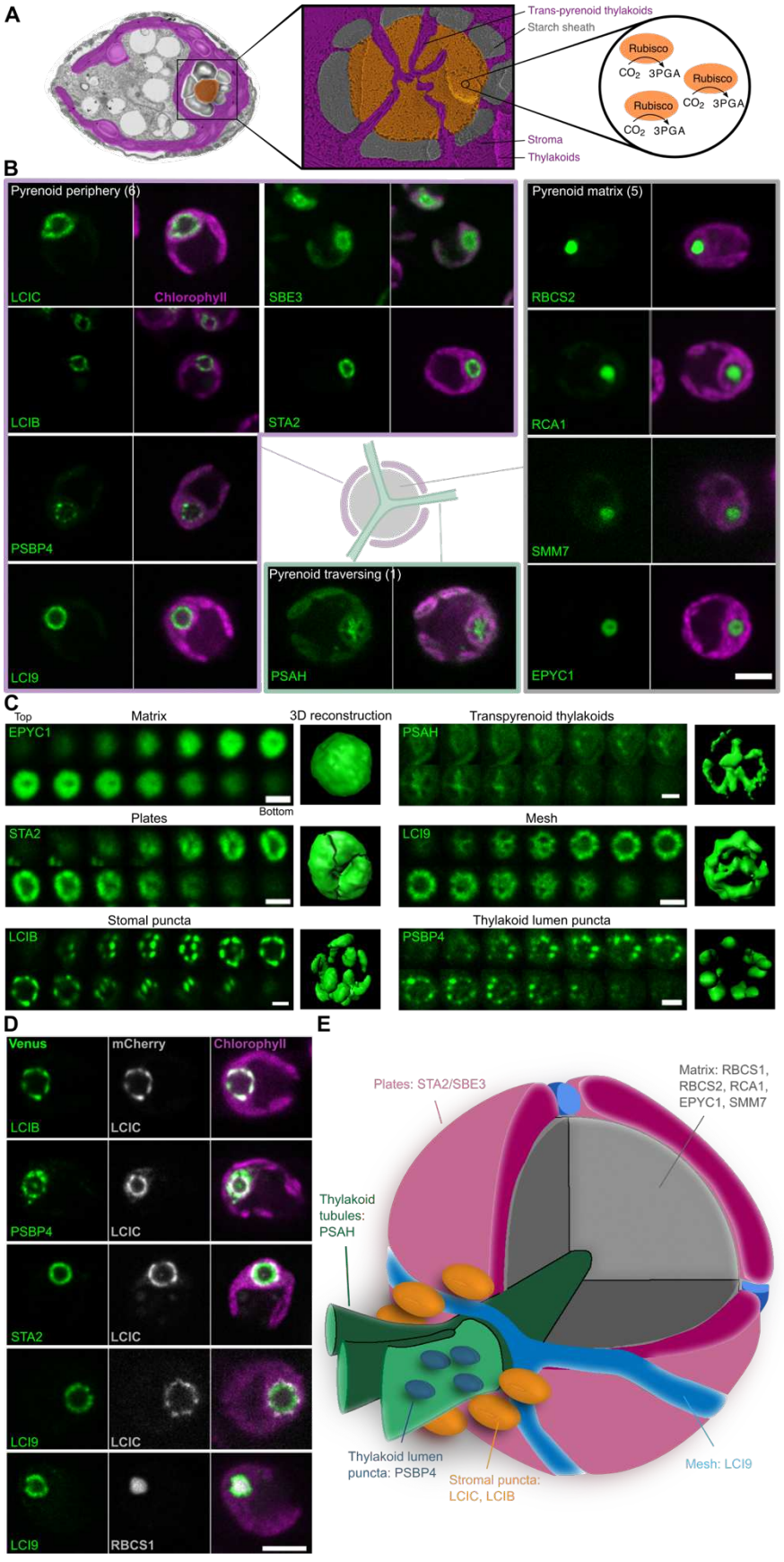


Figure 3. Chloroplast Proteins Show 13 different Localization Patterns.

(A) Representative images of proteins localized to different chloroplast regions. The number of proteins showing each pattern is in parentheses. Scale bar: 5 μm.

(B) The percentage of proteins with predicted transmembrane domains is shown for different localization patterns. Bracket shows a significant difference using Fisher's exact test.

(C) Predicted molecular weight of proteins is shown as a function of pyrenoid signal intensity. Cre01.g030900 that has a pyrenoid signal and is above the 50 kDa cut-off is labeled. Bracket shows significant difference using a Mann-Whitney U test.


A
Trans-pyrenoid thylakoids
Starch sheath
Rubisco
CO2 3PGA
Rubisco
CO2 3PGA
Rubisco
CO2 3PGA
Stroma
Thylakoids
B
Pyrenoid periphery (6)
LCIC
Chlorophyll
SBE3
LCIB
STA2
PSBP4
LCI9
Pyrenoid traversing (1)
PSAH
Pyrenoid matrix (5)
RBCS2
RCA1
SMM7
EPYC1
C
Top
Matrix
3D reconstruction
Transpyrenoid thylakoids
EPYC1
Bottom
PSAH
Plates
Mesh
STA2
LCI9
Stomal puncta
Thylakoid lumen puncta
LCIB
PSBP4
D
Venus
mCherry
Chlorophyll
LCIB
LCIC
PSBP4
LCIC
STA2
LCIC
LCI9
LCIC
LCI9
RBCS1
E
Plates: STA2/SBE3
Matrix: RBCS1, RBCS2, RCA1, EPYC1, SMM7
Thylakoid tubules: PSAH
Thylakoid lumen puncta: PSBP4
Stromal puncta: LCIC, LCIB
Mesh: LCI9

713

Figure 4. Pyrenoid Proteins Show at Least Six Distinct Localization Patterns and Reveal Three New Protein Layers.

(A) A false-color transmission electron micrograph and deep-etched freeze-fractured image of the pyrenoid highlight the pyrenoid tubules, starch sheath and pyrenoid matrix where the principal carbon fixing enzyme, Rubisco, is located.

(B) Proteins showing various localization patterns within the pyrenoid are illustrated. Scale bar: 5 μm.

(C) Confocal sections distinguish different localization patterns within the pyrenoid. Each end panel is a space-filling reconstruction. Scale bars: 2 μm.

(D) Dual tagging refined the spatial distribution of proteins in the pyrenoid. Scale bar: 5 μm.

(E) A proposed pyrenoid model highlighting the distinct spatial protein-containing regions.

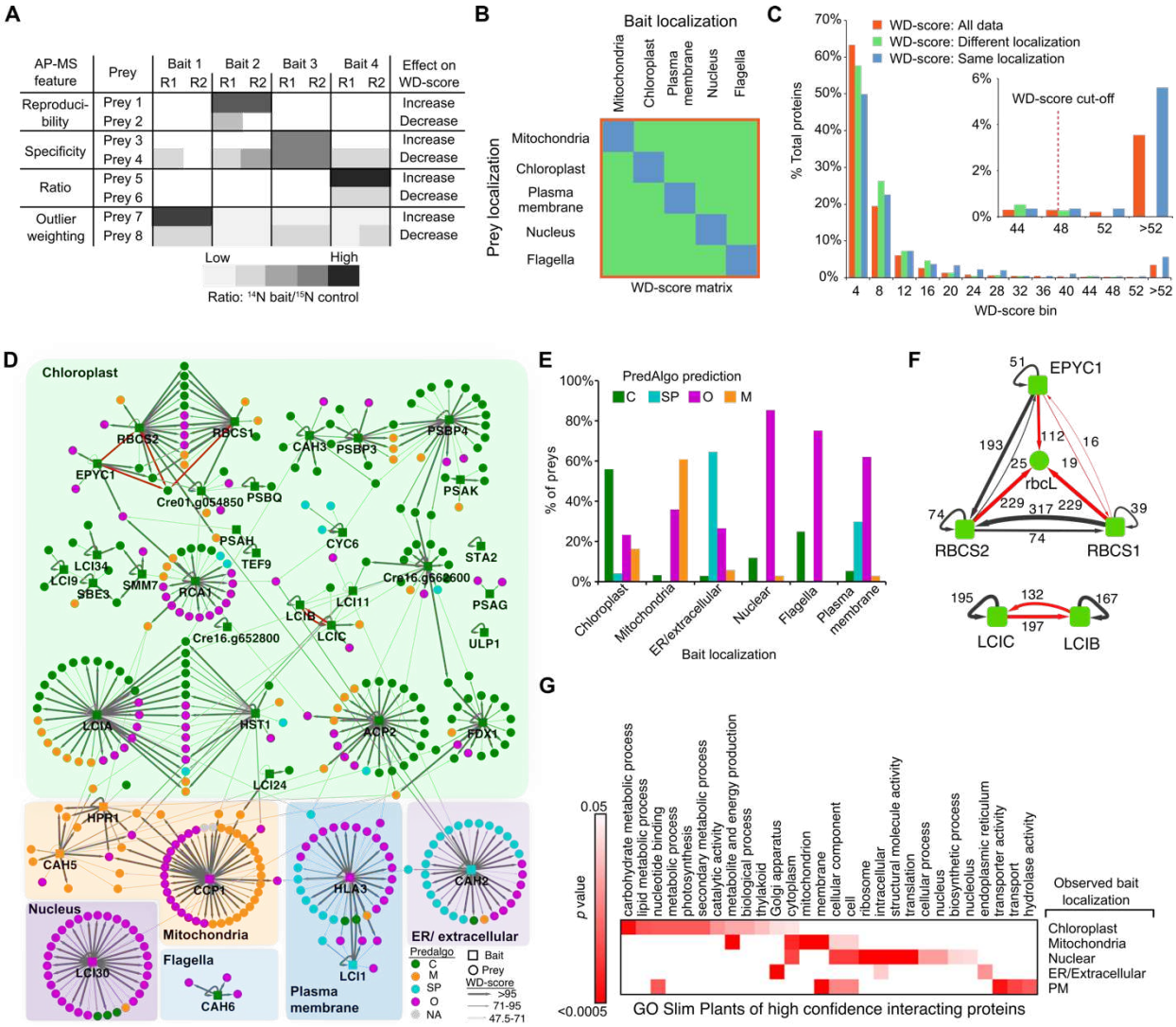


Figure 5. The AP-MS Data are of High Quality.

(A) Illustration of the influence of different AP-MS features (reproducibility, specificity, ratio and outlier weighting) on the WD-score. R1 and R2 represent replica 1 and 2.

(B) To determine a WD-score cut-off value a bait-prey matrix of WD-scores was formed containing only baits and preys whose localizations were determined in this study. The WD-scores from this matrix were then used to generate (C).

(C) A histogram of WD-scores for “All data,” “Different localization,” “Same localization.” A conservative WD-score cut-off was chosen as the point where all data fell above the highest

735 “Different localization” WD-score. Proteins with a WD-score greater than the cut-off are
736 classified as high confidence interacting proteins (HCIPs).
737 (D) Protein-protein interaction network of baits and HCIPs. Bait proteins are grouped according
738 to their localization pattern as determined by confocal microscopy. Baits and preys are colored
739 based on their predicted localization by PredAlgo. Previously known interactions are indicated
740 by red arrows.
741 (E) Comparison of prey PredAlgo predictions with bait localization. C, chloroplast; SP, secretory
742 pathway; O, Other; M, mitochondria.
743 (F) Confirmation of known interactions from the literature (red arrows). Values are WD-scores.
744 (G) Significantly enriched gene ontology (GO) terms for interactors of baits localized to different
745 cellular structures.

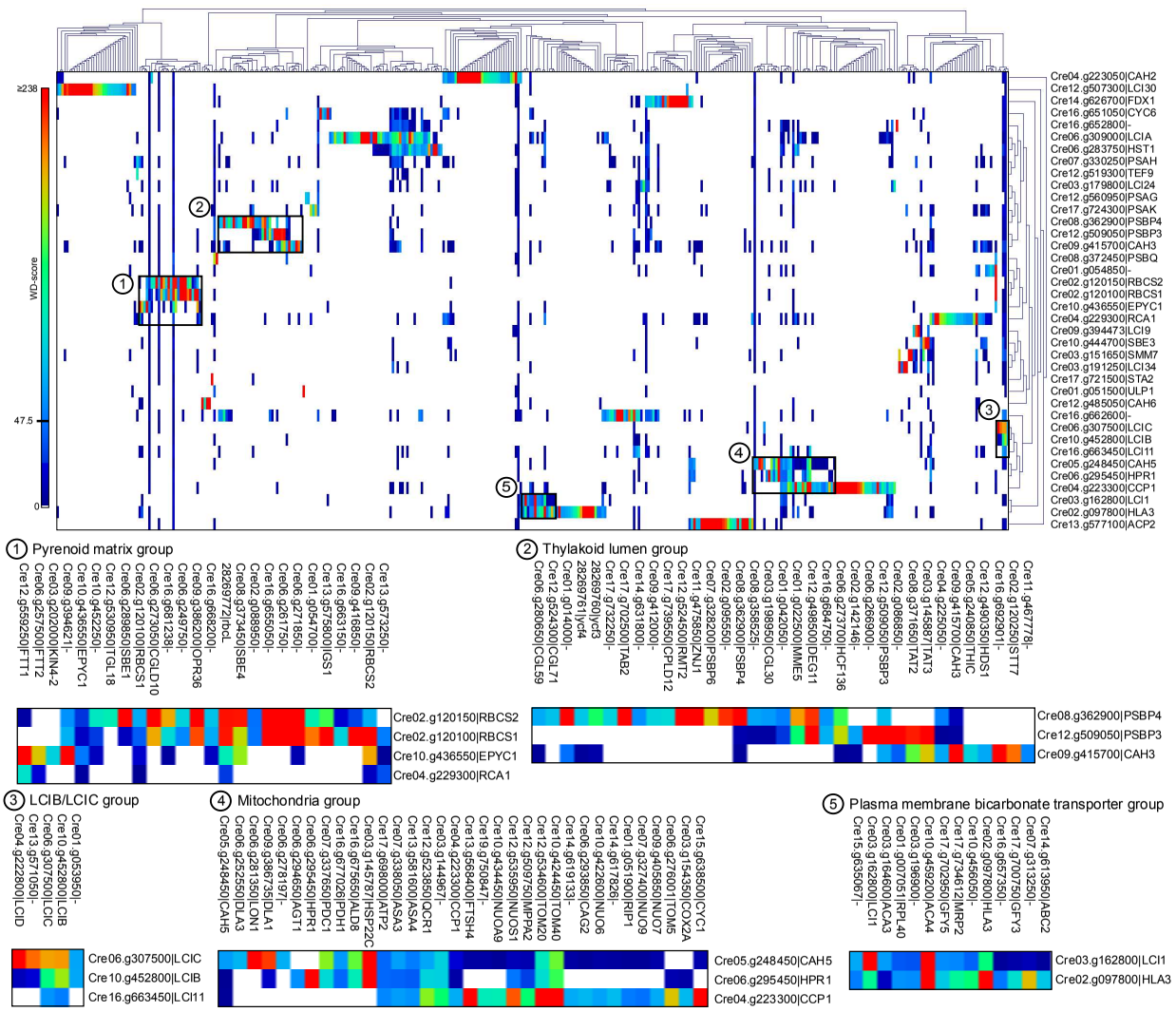


Figure 6. The AP-MS Data Reveals Previously Undescribed Physical Interactions, Including That Inorganic Carbon Transporters LCI1 and HLA3 Form a Physical Complex. Hierarchical clustering of all 38 baits with 398 HCIP preys. Specific groups of interest are boxed and highlighted below. Clustering of all baits and preys with interaction WD-scores ≥ 1 is provided in Figure S5.

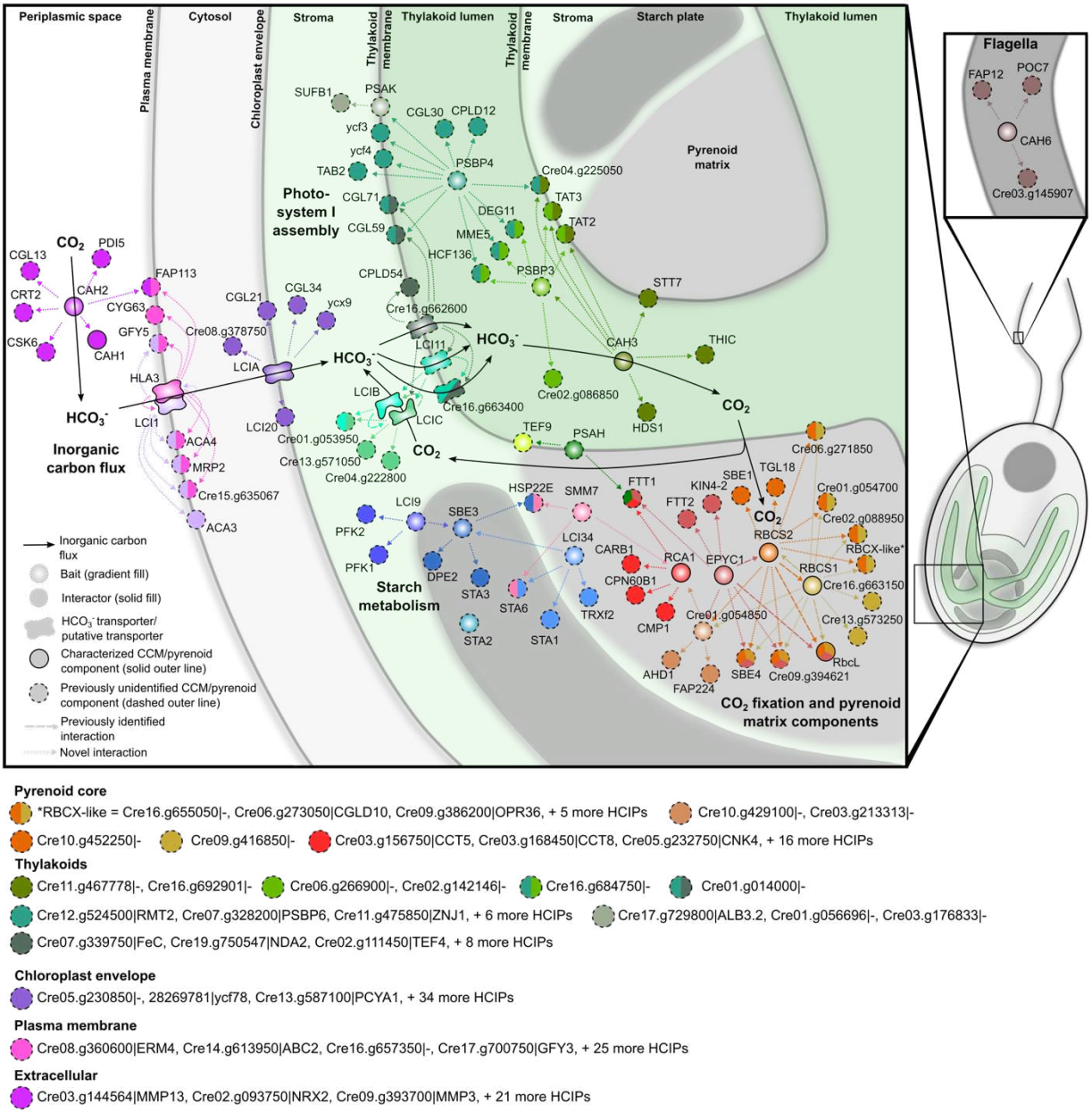
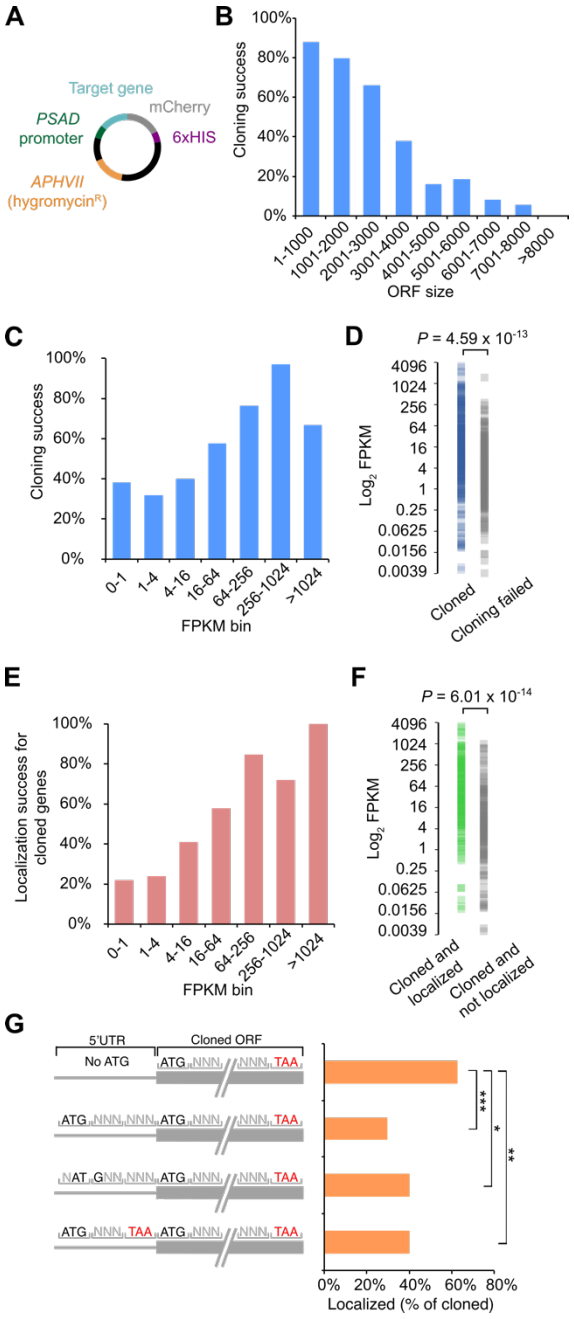


Figure 7. Combining Localization, Protein-Protein Interaction and Protein Function Data Reveals a Spatially Defined Interactome of the *Chlamydomonas* CCM.

A spatially defined protein-protein interaction model of the CCM. Baits have an outer black line, prey have no outer line. Each bait has a unique color. Prey are colored according to their bait, with proteins that interact with multiple baits depicted as pies with each slice colored according

760 to one of their interacting baits. Interactors are connected to their bait by a dashed line
761 representing the direction of interaction. Baits are arranged based on their localization observed
762 in this study. Interactors with predicted transmembrane domains are placed on membranes.
763 Prey of membrane localized baits lacking transmembrane domains are arranged according to
764 their PredAlgo localization prediction. Solid black arrows indicate inorganic flux through the cell.
765 For clarity, a selection of interactors are not included in the map but are highlighted below. All
766 interaction data with corresponding WD-scores can be found in Table S5.
767



768

769 **Figure S1. Cloning Success Correlated With Short ORF Size and High Gene Expression;**

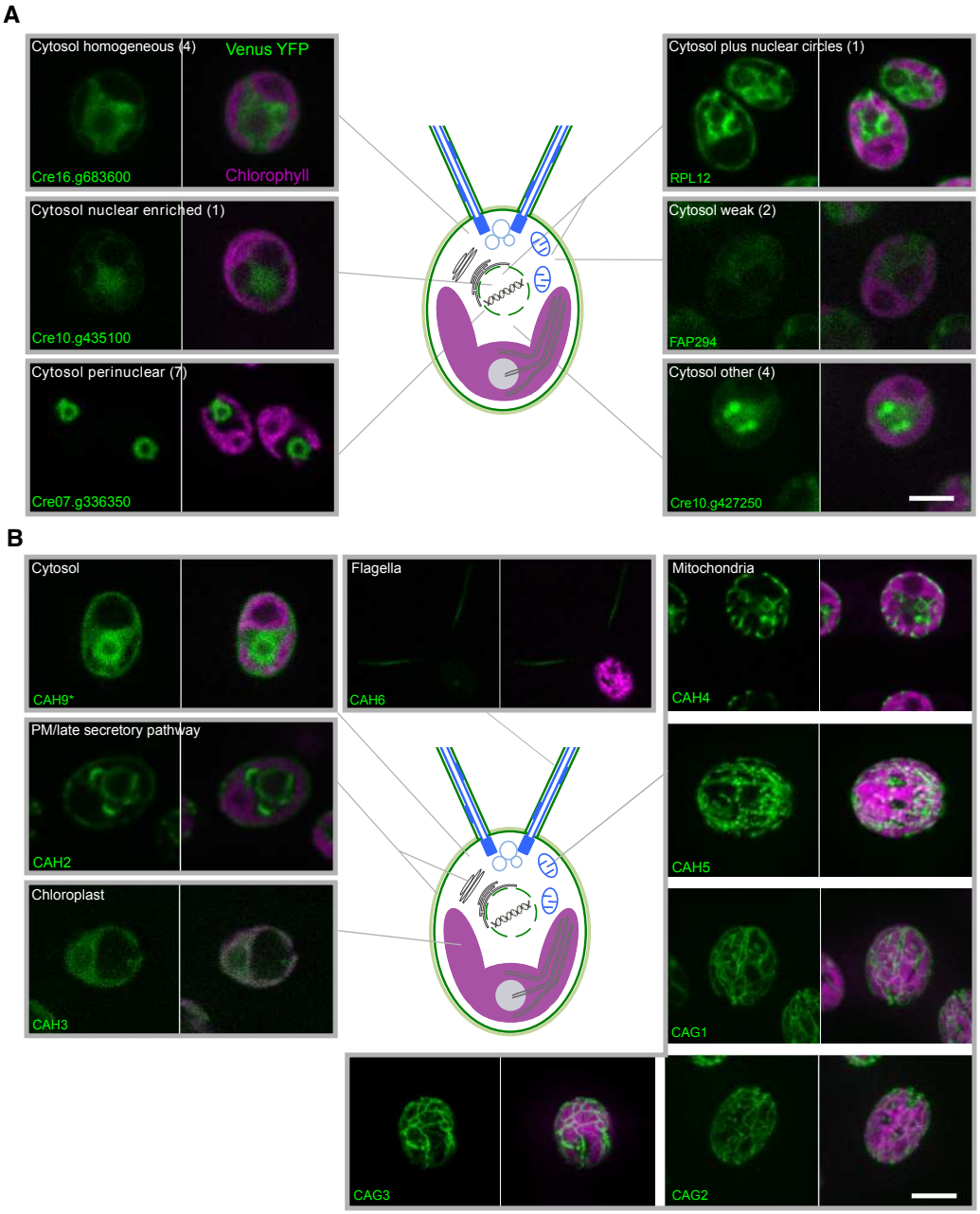
770 **and Protein Localization Success Correlated With Expression and Absence of Upstream**

771 **ATGs, Related to Figure 1**

772 (A) The pLM006 vector used for dual tagging of proteins with mCherry.

773 (B) Dependence of cloning success on open reading frame (ORF) size.

774 (C) Relationship of cloning success to the number of fragments per kilobase of transcript per
775 million mapped reads (FPKM) from phototrophic air-grown cells.
776 (D) Distribution of FPKM values of cloned genes and genes where cloning failed.
777 (E) Relationship of localization success to the FPKM from phototrophic air-grown cells.
778 (F) Distribution of FPKM values of cloned and localized genes vs. cloned and not localized
779 genes. (D) and (F) Brackets show significant difference using a Mann-Whitney U test.
780 (G) The relationship of localization success to presence of uATGs in transcripts. Asterisks
781 denote significant differences using Fisher's exact test: *** P <0.0001, ** P = 0.0025, * P =
782 0.025



783

784

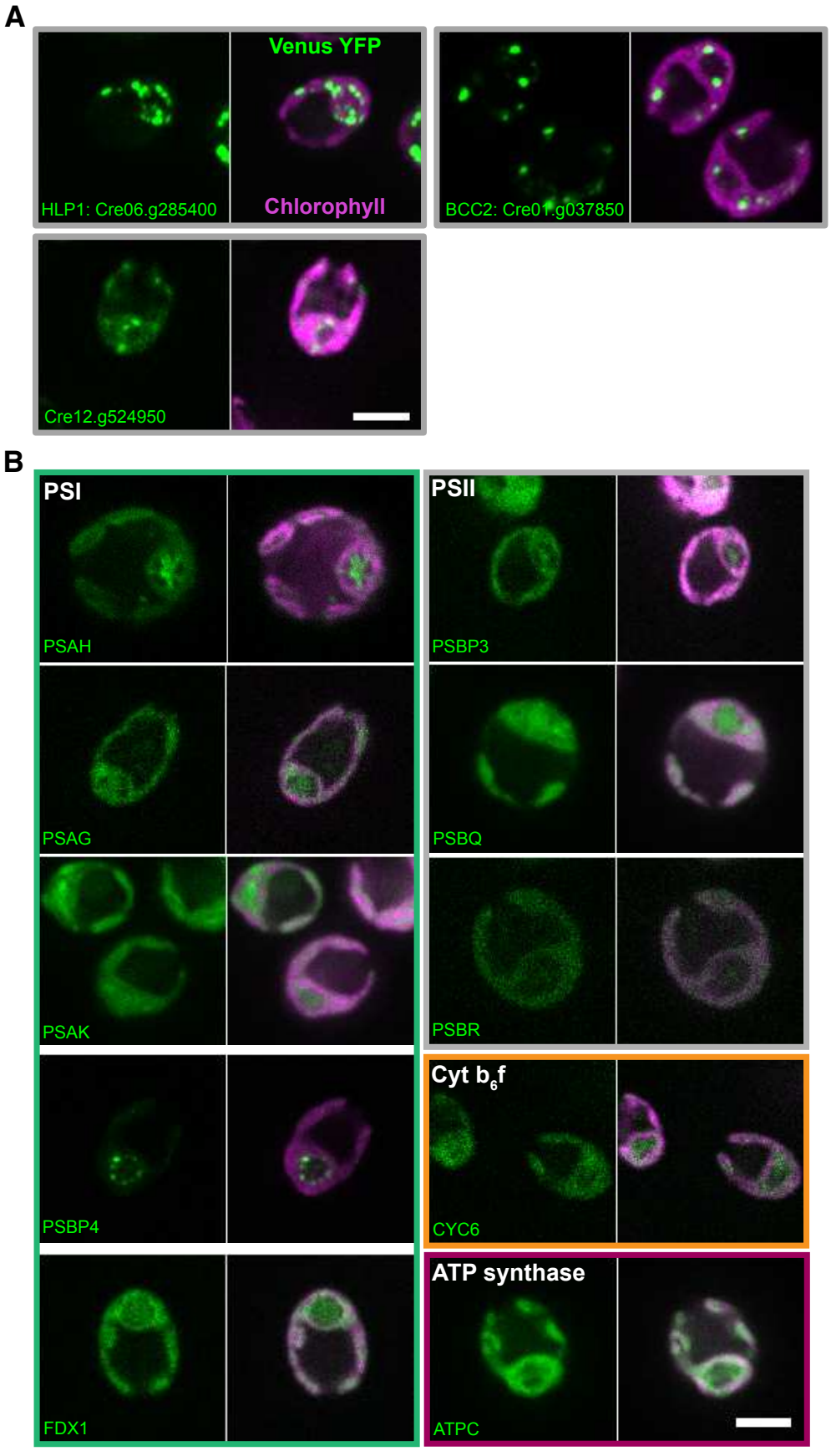
785 **Figure S2. Diverse Cytosolic Patterns and Carbonic Anhydrase Localizations Were**

786 **Observed, Related to Figure 2**

787 (A) Representative confocal images demonstrating a diverse range of cytosolic localization

788 patterns.

(B) Confocal images of successfully tagged and localized carbonic anhydrases. *The cloned construct was based on the CAH9 Augustus v5.0 gene model. Images for CAH5 and CAG1-3 are projected Z-stacks. (A) and (B) Scale bars: 5 μm.

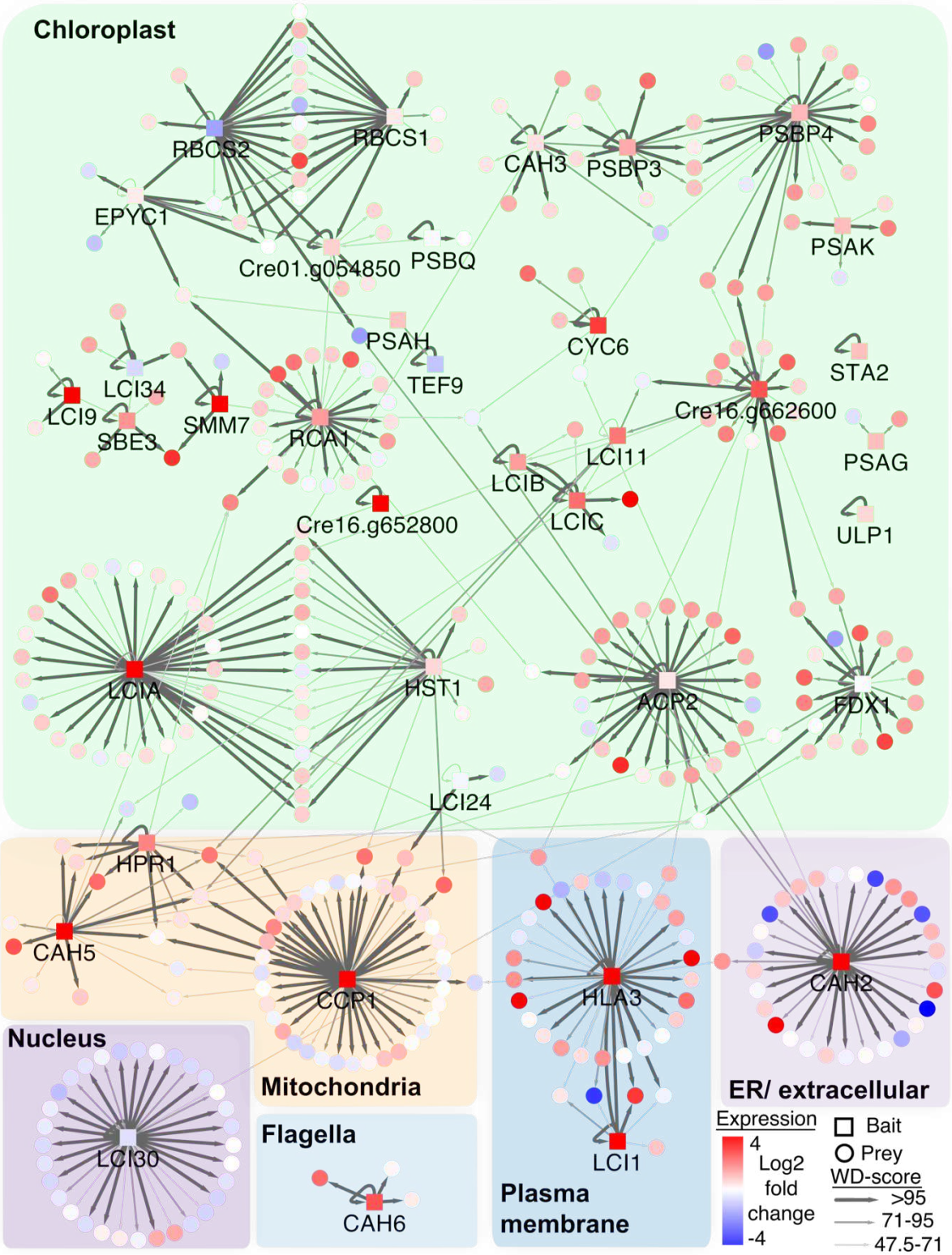

A
Venus YFP
HLP1: Cre06.g285400
Chlorophyll
BCC2: Cre01.g037850
Cre12.g524950
B
PSI
PSAH
PSAG
PSAK
PSBP4
FDX1
PSII
PSBP3
PSBQ
PSBR
Cyt b6f
CYC6
ATP synthase
ATPC

792

Figure S3. Proteins Forming Puncta Within the Chloroplast and Localization of Proteins Associated with Photosynthetic Electron Transport, Related to Figure 3

(A) Confocal images of proteins with signals in defined puncta within the chloroplast.

(B) Localization of Proteins Associated with Photosynthetic Electron Transport**.** The images for PSBP4 and PSAH are the same as used in Figure 4B. (A) and (B) Scale bars: 5 μm.

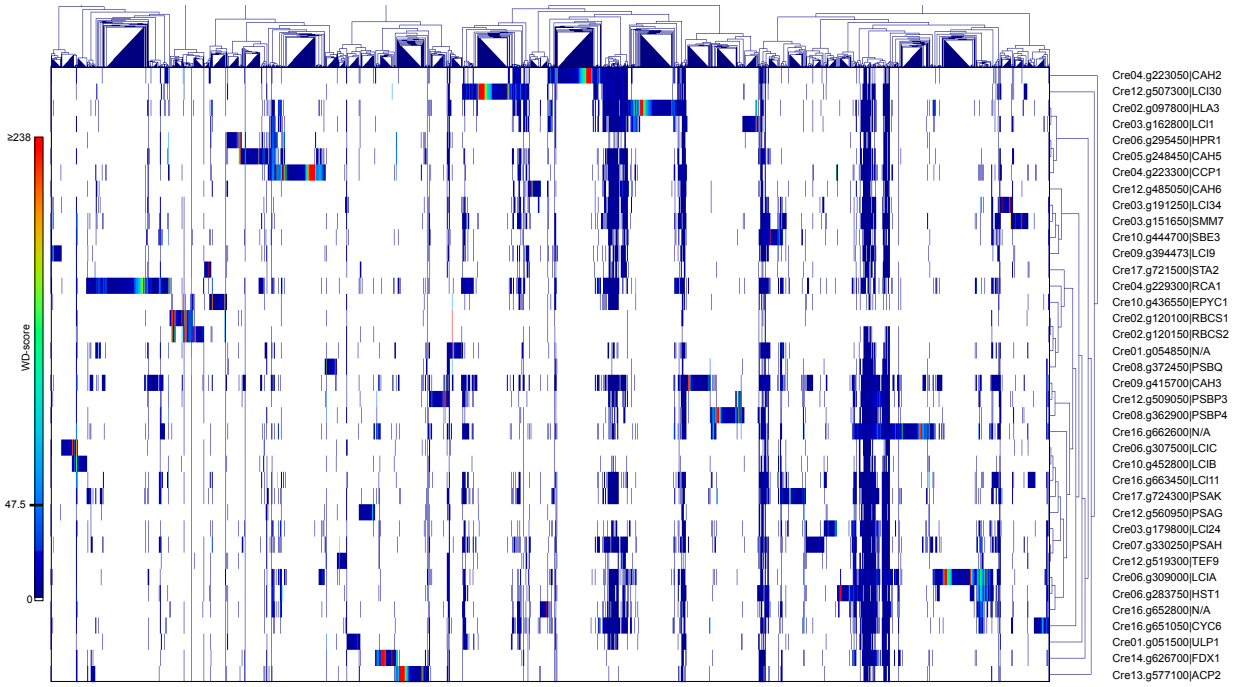

Chloroplast
RBCS2
RBCS1
EPYC1
Cre01.g054850
PSBQ
CAH3
PSBP3
PSBP4
PSAK
PSAH
TEF9
CYC6
STA2
LCI34
LCI9
SBE3
SMM7
RCA1
Cre16.g662600
PSAG
LCIB
LCI11
LCIC
Cre16.g652800
ULP1
LCIA
HST1
ACP2
FDX1
LCI24
HPR1
CAH5
CCP1
Nucleus
LCI30
Mitochondria
Flagella
CAH6
HLA3
LCI1
Plasma membrane
CAH2
ER/ extracellular
Expression
4
Log2 fold change
-4
Bait
Prey
WD-score
>95
71-95
47.5-71

799

800

Figure S4. Transcriptional Regulation of the CCM Protein-Protein Interaction Network, Related to Figure 5

Log_2 fold changes of proteins upregulated (red) or downregulated (blue) in response to low CO_2 are overlaid onto the HCIP protein-protein interaction network.



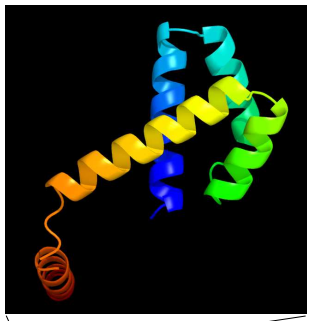
805

806

807 **Figure S5. Hierarchical Clustering to Identify Protein Complexes, Related to Figure 6**

808 Hierarchical clustering of all 38 baits and preys having an interaction WD-score ≥1. Large

809 regions of blue across most/all baits correspond to clusters of non-specific interactors.



Phyre2 template	Confidence	% ID	PDB Title/Family
c4gr6B	97.5	25	Crystal structure of AtRbcX2 from arabidopsis thaliana
d2py8a1	97.1	30	RbcX-like
d2peqa1	96.7	31	RbcX-like
c2py8B	96.6	33	RbcX
d2peoa1	96.6	28	RbcX-like
c2peoA	96.6	28	Crystal structure of RbcX from *Anabaena* CA
c2penE	96.6	31	Crystal structure of RbcX, crystal form i
c3ka1A	96.5	25	Crystal structure of RbcX from Thermosynechococcus elongatus
c4gr2A	95.7	19	Structure of AtRbcX1 from *Arabidopsis thaliana*
c4whjA	95.3	9	Myxovirus resistance protein 2 (mxb)

Figure S6. Further Analysis of Cre16.g655050, Related to Figure 6

Cre16.g655050 has a RbcX N-Terminal Domain and a Disordered C Terminus. Top: A predicted Phyre2 structural model of Cre16.g655050. The table shows the ten best template matches for Cre16.g655050 by Phyre2. The confidence score is the probability that the match between Cre16.g655050 and the template is a true homology. The % ID shows the percentage identity between Cre16.g655050 and the template.

819

820

821 **Figure S7. ACA4 Belongs to P-type ATPase Family IIIA, Related to Figure 6**

822 Phylogenetic tree analysis of 259 eukaryotic P-type ATPases, including functionally
823 characterized members representing the different P-type ATPase families. *Chlamydomonas*
824 ACA4 groups with family IIIA P-type ATPases, which are involved in H^+ pumping.

Table S1. Summary of Target Gene Features, Cloning and Localization.

(Attached excel spreadsheet)

Table S2. Proteins That Showed Multiple Localizations.

Phytozome v5.5 (Augustus u111.6) ID	Name	Localization	Predalgo predicted localization	Putative function	Predicted Mw
Cre09.g416800	-	Cytosol and chloroplast	O	-	22.78
Cre12.g552450	-	Cytosol and chloroplast	O	-	11.85
Cre16.g685000	-	Cytosol and chloroplast	C	-	25.67
Cre07.g334800	FDX4	Cytosol and chloroplast	C	Ferredoxin	14.05
Cre09.g396400	UBQ2	Flagella and cytosol	O	Bi-ubiquitin	17.2
Cre03.g204577	DNJ31	Flagella and cytosol	C	DnaJ-like protein	61.75
Cre07.g321800	-	Flagella and cytosol	O	-	24.88
Cre11.g467617	LCI19	Flagella and cytosol	O	Gamma hydroxybutyrate dehydrogenase	30.24
Cre16.g685050	LCI15	Flagella and cytosol	C	PRLI-interacting factor L	34.34
Cre03.g158000	GSA1	Flagella, chloroplast and cytosol	C	Glutamate-1-semialdehyde aminotransferase	49.23
Cre17.g725500	-	Flagella, chloroplast and cytosol	C	-	13.82
Cre07.g337100	-	Flagella and mitochondria	C	-	10.75

832 **Table S3. Proteins Used as Baits for the AP-MS Study.**

Phytozome JGI v5.5 (Augustus u111.6) ID	Name	Protein description	Localization	Replica type (Biological, BR; Affinity Purification AR; Mass Spectrometry, MR)	Number of Preys with WD-score >1	Number of HCIPs (WD-score >47.52)
Cre01.g051500	ULP1	Uncharacterized thylakoid lumenal polypeptide	Chloroplast not homogeneous with pyrenoid signal	MR	179	1
Cre01.g054850	-	-	Chloroplast not homogeneous with pyrenoid signal	AR	211	6
Cre02.g097800	HLA3	ABC transporter	Plasma membrane and late-secretory pathway	AR	372	36
Cre02.g120100	RBCS1	Rubisco small subunit 1	Pyrenoid matrix	BR	81	19
Cre02.g120150	RBCS2	Rubisco small subunit 2	Pyrenoid matrix	AR	104	20
Cre03.g151650	SMM7	-	Pyrenoid matrix	AR	303	5
Cre03.g162800	LCI1	Low-CO2-inducible membrane protein	Plasma membrane and late-secretory pathway	AR	266	7
Cre03.g179800	LCI24	Low-CO2-inducible membrane protein	Chloroplast homogeneous with pyrenoid signal	AR	249	3
Cre03.g191250	LCI34	Low-CO2-inducible protein	Chloroplast not homogeneous with pyrenoid signal	AR	239	5
Cre04.g223050	CAH2	Carbonic anhydrase, alpha type, periplasmic	ER	MR	323	34
Cre04.g223300	CCP1	Low-CO2-inducible chloroplast envelope protein	Mitochondria	AR	328	48
Cre04.g229300	RCA1	Rubisco activase	Pyrenoid matrix	AR	463	26
Cre05.g248450	CAH5	Mitochondrial carbonic anhydrase	Mitochondria	BR	290	19
Cre06.g283750	HST1	Homogentisate solanesyltransferase	Chloroplast not homogeneous with pyrenoid signal	AR	309	22
Cre06.g295450	HPR1	Hydroxypyruvate reductase	Mitochondria	MR	162	12
Cre06.g307500	LCIC	Low-CO2 inducible protein	Pyrenoid periphery punctate	MR	113	6
Cre06.g309000	LCIA	Anion transporter	Chloroplast envelope plus chloroplast homogeneous	AR	393	43
Cre07.g330250	PSAH	Subunit H of photosystem I	Pyrenoid tubules	AR	351	2
Cre08.g362900	PSBP4	Lumenal PsbP-like protein	Pyrenoid periphery punctate	AR	234	24
Cre08.g372450	PSBQ	Oxygen-evolving enhancer protein 3	Chloroplast not homogeneous with pyrenoid signal	AR	134	3
Cre09.g394473	LCI9	Low-CO2-inducible protein	Pyrenoid periphery mesh	MR	200	4
Cre09.g415700	CAH3	Carbonic anhydrase 3	Chloroplast homogeneous with pyrenoid signal	AR	500	11
Cre10.g436550	EPYC1/LCI5	Low-CO2-inducible protein	Pyrenoid matrix	BR	146	9
Cre10.g444700	SBE3	Starch branching enzyme	Pyrenoid periphery spherical	AR	212	4
Cre10.g452800	LCIB	Low-CO2-inducible protein	Pyrenoid periphery punctate	AR	136	3
Cre12.g485050	CAH6	Carbonic anhydrase 6	Flagella	MR	190	4
Cre12.g507300	LCI30	Low-CO2-inducible protein	Nucleus	MR	320	34
Cre12.g509050	PSBP3	OEE2-like protein of thylakoid lumen	Chloroplast not homogeneous with pyrenoid signal	AR	245	10
Cre12.g519300	TEF9	Predicted protein	Chloroplast homogeneous with pyrenoid signal	MR	180	1
Cre12.g560950	PSAG	Photosystem I reaction center subunit V	Chloroplast not homogeneous with pyrenoid signal	AR	145	2
Cre13.g577100	ACP2	Acyl-carrier protein	Chloroplast not homogeneous with pyrenoid signal	MR	189	26
Cre14.g626700	Fd/FDX1	Ferredoxin	Chloroplast not homogeneous with pyrenoid signal	AR	199	20
Cre16.g651050	CYC6	Cytochrome c_6	Chloroplast not homogeneous with pyrenoid signal	AR	288	7
Cre16.g652800	-	-	Chloroplast homogeneous with pyrenoid signal	AR	281	2
Cre16.g662600	-	-	Chloroplast homogeneous pyrenoid disenriched	AR	371	23
Cre16.g663450	LCI11	Low-CO2-inducible membrane protein	Chloroplast homogeneous with pyrenoid signal	AR	284	6
Cre17.g721500	STA2	Granule-bound starch synthase I	Pyrenoid periphery spherical	AR	142	1
Cre17.g724300	PSAK	Photosystem I reaction center subunit psaK	Chloroplast not homogeneous with pyrenoid signal	AR	319	5
				Totals	9451	513

833

Table S4. Cre16.g655050 BLAST Results.

Query cover is the percentage of the query sequence that matches the hit sequence. E-value is the expected value, the lower the E-value the more significant the hit.

Accession	Species	Query cover	E-value	Identity
XP_002950714.1	*Volvox carteri*	77%	4.0E-137	56%
KXZ52617.1	*Gonium pectorale*	71%	9.0E-127	56%
XP_005849673.1	*Chlorella variabilis*	31%	2.0E-53	56%
XP_005645512.1	*Coccomyxa subellipsoidea*	28%	2.0E-50	55%
XP_005847655.1	*Chlorella variabilis*	27%	5.0E-48	53%
XP_001698126.1	*Chlamydomonas reinhardtii*	10%	1.0E-29	100%
XP_013896920.1	*Monoraphidium neglectum*	16%	2.0E-29	60%
XP_002501227.1	*Micromonas commoda*	24%	1.0E-18	41%
XP_003062310.1	*Micromonas pusilla*	24%	2.0E-17	39%

Table S5. Protein-Protein Interaction Data.

All interactions with a WD-score ≥1 are shown. Rows highlighted in blue were classified as HCIPs.

(Attached excel spreadsheet)

EXPERIMENTAL PROCEDURES

Strains and Culturing

The background *Chlamydomonas reinhardtii* strain for all experiments was wild-type (WT) cMJ030 (CC-4533). WT cells were maintained on 1.5% Tris-acetate-phosphate (TAP) agar with revised Hutner's trace elements (Kropat et al., 2011) at 22°C in low light (~10 µmol photons m^{-2} s^{-1}). Lines harboring Venus-3xFLAG-tagged genes in the pLM005 plasmid were maintained in the same conditions with solid media supplemented with 20 µg mL^{-1} paromomycin. For lines also harbouring the pLM006 plasmid, the media was further supplemented with 25 µg mL^{-1} hygromycin. During liquid growth for imaging and affinity purification mass spectrometry, antibiotic concentrations were used at $1/10^{th}$ these concentrations.

Plasmid Construct and Cloning

For the tagging and AP-MS pipeline, we used the pLM005 plasmid, and for dual-tagging experiments, we used the pLM006 plasmid (Mackinder et al., 2016). Open reading frames were PCR amplified from genomic DNA and cloned in-frame with either a C-terminal Venus-3xFLAG (pLM005) or an mCherry-6xHIS (pLM006) tag by Gibson assembly as previously described (Mackinder et al., 2016). Primers were designed to amplify target genes from their predicted start codon up to, but not including, the stop codon. To allow efficient assembly into *Hpa*I-cut pLM005 or pLM006, primers contained the following adapters: Forward primers (5'-3'), GCTACTCACAACAAGCCCAGTT and reverse primers (5'-3'), GAGCCACCCAGATCTCCGTT. To increase our success with larger genes, we split some of these into multiple fragments that were reassembled following PCR amplification. However, due to a multiplicative effect, the cloning efficiency dropped off rapidly: only a 20% efficiency for two fragments (14/69) and 8% for three fragments (6/74). All junctions were sequence verified by Sanger sequencing and constructs were linearized by either *Eco*RV or *Dra*I prior to electroporation into WT *Chlamydomonas reinhardtii*. Transformations were performed as described in Zhang et al.,

2014. Lines expressing fluorescent tagged proteins were identified and selected as previously described (Mackinder et al., 2016). For each construct, three fluorescent colonies were isolated and maintained in 96 arrays using a Singer Rotor propagation robot.

Microscopy

For microscopy of Venus-tagged lines, colonies were transferred from agar to Tris-phosphate (TP) liquid medium (Kropat et al., 2011) in a 96-well microtiter plate and grown with gentle agitation in air at 150 µmol photons m^{-2} s^{-1} light intensity. After ~2 days of growth, 15 µL of cells were pipetted onto a 96-well optical bottom plate (Brooks Automation Inc.) and a 120 µL of 1% TP low-melting-point agarose at ~34°C was overlaid to minimize cell movement. Lines grown for detailed Z-stack analysis and dual-tagged lines containing proteins with both Venus and mCherry tags were grown in 80 mL of TP, bubbled with 0.01% CO_2 (with 21% O_2, balanced with N_2) for ~12 hours at 150 µmol photons m^{-2} s^{-1} light intensity. Cells were the prepared for imaging as described in Mackinder et al. (2016). All imaging was performed using a spinning-disk confocal microscope (custom modified Leica DMI6000) according to Mackinder et al. (2016).

Affinity Purification

Cell lines expressing Venus-3xFLAG-tagged proteins were grown in 50 mL of TAP media at 100 µmol photons m^{-2} s^{-1} light intensity until they reached a cell density of ~2-4 x 10^6 cells mL^{-1}. Cells were then pelleted at 1000 *g* for 4 minutes, resuspended in TP medium and transferred to 800 mL of TP medium. They were then bubbled with air with constant stirring and 150 µmol photons m^{-2} s^{-1} light intensity to a density of ~2-4 x 10^6 cells mL^{-1}. All liquid media contained 2 µg mL^{-1} paromomycin. In parallel, control strains expressing only the Venus-3xFLAG tag were grown under identical conditions except that, during liquid growth, $^{14}NH_4Cl$, the sole nitrogen source, was replaced with $^{15}NH_4Cl$. This ensured ^{15}N growth for at least eight generations.

Cells from Venus-3xFLAG-tagged protein lines and control lines were separately harvested and affinity purified as described in Mackinder et al., 2016 except that the affinity purification incubation step was reduced to 1.5 hours. After competitive elution by incubation with the 3xFLAG peptide, samples were diluted 1:1 with 2X SDS-PAGE buffer (BioRad) containing 50 mM β-mercaptoethanol and heat denatured for 10 minutes at 70°C. Tagged protein and control denatured elutions were then mixed 1:1 (16µL:16µL), and 28 µL of sample was partially purified by electrophoresing on a Tris-glycine gel (Criterion TGX gel ; BioRad) until the protein moved 1.8 to 2 cm (~40 minutes at 50V). Gel slices were then fixed in 1 mL of 10% acetic acid, 50% methanol, 40% deionised water for 1 hour, with a change of the fixing solution after 15 minutes, 30 minutes and 1 hour. Gel slices were soaked twice in 1mL of deionized water for 2 minutes, then stored in 1% acetic acid at 4°C until processing for mass spectrometry.

Mass Spectrometry

Limited gel slices representing 3xFLAG AP eluates were diced into 1x1mM squares and then incubated in 50 mM ammonium bicarbonate for ~15 minutes. After pH neutralization, the diced gel slices were reduced with 5 mM DTT for 30 minutes at 55°C. The reducing buffer was removed and samples were alkylated with 10 mM propionamide at 10 mM for 30 minutes at room temperature. Gel samples were washed with multiple rounds of 1:1 acetonitrile:50mM ammonium bicarbonate until the gels were free of all dye. 10 uL of 125 nanogram trypsin/lysC (Promega) was added to each gel band and gels were allowed to swell for 10 minutes, followed by the addition of 25 to 35uL 50 mM ammonium bicarbonate. The gels were digested overnight at 37°C. Peptide extraction was performed in duplicate, and the peptide pools dried in a speed vac until readied for LCMS/MS. Each peptide pool was reconstituted in 12.5 uL 0.1% formic acid, 2% acetonitrile, 97.9% water and loaded onto a NanoAcquity UPLC (Waters). The mobile phases were A: 0.585% acetic acid, 99.415% water and B: 0.585% acetic acid, 10% water,

89.415% acetonitrile. The analytical column was a picochip (New Objective) packed with 3 µM C18 reversed phase material approximately 10.5cm in length. The flow rate was 600 nL/min during the injection phase and 450 nL/min during the analytical phase. The mass spectrometer was a orbitrap Elite, operated in a data-dependant acquisition (DDA) schema in which the fifteen most intense multiply charged precursor ions were selected for fragmentation in the ion trap. The precursor mass settings were a resolution of 120,000 and an ion target value of 750,000, max fill time 120 usec. The MS/MS settings were 50,000 ions and a maximum fill time of 25 µsec.

Mass Spectrometry Data Analysis

Peptide identification

MS/MS data were analyzed using an initial screening by Preview for validation of data quality, followed by Byonic v2.6.49 (Bern et al., 2012) for peptide identification and protein inference against version 5.5 of the *Chlamydomonas reinhardtii* translated genome. In a typical analysis, each data file was searched in two parallel Byonic analyses: one for the unlabeled peptides, and one treating the incorporation of ^{15}N isotopic labels as a fixed modification. In both cases, these data were restricted to 12 ppm mass tolerances for precursors, with 0.4 Da fragment mass tolerances assuming up to two missed cleavages and allowing for only fully tryptic peptides. These data were validated at a 1% false discovery rate using typical reverse-decoy techniques as described previously (Elias and Gygi, 2007). The combined identified peptide spectral matches and assigned proteins were then exported for further analysis using custom tools developed in MatLab (MathWorks) to provide visualization and statistical characterization.

Background to CompPASS analysis

To identify *bona fide* interactions, we used an $^{14}N/^{15}N$ labeling strategy. Bait-Venus-3xFLAG fusion proteins were grown in ^{14}N media in parallel to ^{15}N grown controls expressing only

Venus-3xFLAG. 3xFLAG affinity purification was performed for target and control lines in parallel, proteins were eluted by 3xFLAG competition, and then target and control elutions were mixed prior to SDS-PAGE purification and MS. In theory, this approach should control for non-specific proteins interacting with the resin, 3xFLAG peptide, Venus and tubes and it should also control for MS variation between runs, resulting in only large ratios for specific interactors. However, analysis of the complete data set showed that using only $^{14}N/^{15}N$ ratios was insufficient to identify real interactors from false positives. This is generally due to the spurious nature of some preys, and in several cases the ratios diverged from 1 across all baits for some preys. Therefore, to analyze our $^{14}N/^{15}N$ labeled dataset, we decided to adapt the CompPASS method (Sowa et al., 2009), an approach previously developed to analyze AP-MS studies of this size using unlabeled proteins.

Identification of carry-over proteins from previous MS runs

Carry-over of proteins from previous MS runs is a common source of contamination, and increases with protein abundance and hydrophobicity (Morris et al., 2014). To reduce carry-over contamination, column wash steps and MS blanks were frequently included, and placed between samples that were previously identified to be prone to carry-over. In addition, an *in silico* filtering step was included to remove carry-over contamination prior to CompPASS analysis. Data was sorted by MS run order and half-life-like patterns of decreasing raw values were scanned for. To confirm contamination was due to carry-over and not true interactions, half-life-like patterns between MS replicas ran in a different order were compared. Raw values for carry-over contamination that showed the same patterns between replicas were set to zero.

Generating WD-scores

The CompPASS method uses spectral counts and devises a score (WD-score) based on the specificity of the prey, spectral count number and reproducibility. Instead of using spectral

counts we used $^{14}N/^{15}N$ ratios. Using $^{14}N/^{15}N$ ratios helps clean out abundant common contaminants. Based on the CompPASS method, we generated WD-scores for each bait-prey interaction. First, we determined the $^{14}N/^{15}N$ ratios for the bait-prey interaction for each replica. If a protein had no spectral counts in one of the ^{14}N or ^{15}N, the spectral count was set to 1 to generate a ratio. If it was not detected in both the ^{14}N and ^{15}N, its $^{14}N/^{15}N$ ratio value was therefore 1. The ratios for each replica were then averaged to populate a stats table of 38 baits and 3251 preys.

Stats table

	Bait 1	Bait 2	Bait 3	Bait *k*	
Prey 1	$X_{1,1}$	$X_{2,1}$	$X_{3,1}$	$X_{k,1}$	$\overline{X}_1$
Prey 2	$X_{1,2}$	$X_{2,2}$	$X_{3,2}$	$X_{k,2}$	$\overline{X}_2$
Prey 3	$X_{1,3}$	$X_{2,3}$	$X_{3,3}$	$X_{k,3}$	$\overline{X}_3$
Prey *m*	$X_{1,m}$	$X_{2,m}$	$X_{3,m}$	$X_{k,m}$	$\overline{X}_m$

$X_{i,j}$ is the average $^{14}N/^{15}N$ ratio from two replicas for prey *j* from bait *i.*

m is the total number of unique prey proteins identified (3251).

k is the total number of unique baits (38).

We plugged the above values into the WD-score equation, which is defined as follows (Behrends et al., 2010):

$$WD_{i,j} = \sqrt{\left(\frac{k}{\sum_{j=1}^{i=k} f_{i,j}}\,\omega_j\right)^p}\,X_{i,j}$$

$$\omega_j = \left(\frac{\sigma_j}{\overline{X}_j}\right), \overline{X}_j = \frac{\sum_{i=1,j=n}^{i=k} X_{i,j}}{k}, n = 1,2,\ldots m, \quad \begin{matrix} if\ \omega_j \leq 1 \rightarrow \omega_j = 1 \\ if\ \omega_j > 1 \rightarrow \omega_j = 0 \end{matrix}$$

$$f_{i,j} = \{^{1;X_{i,j} > 0}_{X_{i,j}}$$

The WD-score has 3 main components taking into account the uniqueness, the reproducibility and the $^{14}N/^{15}N$ ratio. $\frac{k}{\sum_{j=1}^{i=k} f_{i,j}}$ is a "uniqueness" measure that up-weights unique interactors and down-weights promiscuous interactors. It counts the number of baits that a given prey was detected in. Therefore, the less often the prey is seen across the baits, the larger the value. *k* is constant for all preys, in our case it is 38. Therefore, if a prey is unique to one bait, this term will equal 38 (38/1), whereas if is a prey is seen interacting with all baits this value would be 1 (38/38). In addition to the uniqueness measurement is a weighting term, ω_j. This term is only applied if the standard deviation is greater than the mean for a prey across all baits. It was introduced in Behrends et al. (2010) to offset the low uniqueness value for true interactors that are seen in many baits.

p is a reproducibility measure that upweights preys that are seen in both replicas if the ratio averages were ≤10.2 fold of each other. We decided to add a "closeness" value of replica ratios because, for spurious and general contaminant preys, they frequently would be detected in both replicas but would have a large $^{14}N/^{15}N$ ratio difference between replicas, whereas in true interactors $^{14}N/^{15}N$ ratios between replicas are generally very similar. To determine a cut-off, we looked at all preys that were only detected in one bait and which were also replicated in both MS runs (this gave 173 high-confidence true interactions). We then took the largest fold change between the replica $^{14}N/^{15}N$ ratios where more than 1 spectral count was used to determine the ratio.

$X_{i,j}$ is the $^{14}N/^{15}N$ ratio. In Sowa et al, 2009, this is the average of total spectral counts for the replicas. In our case the $X_{i,j}$ is the average of the $^{14}N/^{15}N$ of both replicas. By using the $^{14}N/^{15}N$ ratio we in effect have performed an initial clean up of the data, with background contaminants (seen in both the ^{14}N bait and ^{15}N control) down-weighted.

If the protein was not detected in either replica it was assigned a WD-score of 0.

Determining the WD-score threshold

Due to the empirical nature of the WD-score, a cut-off must be determined. Sowa et al. (2009) generated a random dataset and used a cut-off value above which 5% of the random dataset fell. Interestingly, this also corresponded to ~5% of the real dataset, which they recommend as a suitable approximation for the threshold. Due to potential pitfalls in the generation of a random dataset, we decided to use an alternate approach to determine the WD-score cut-off. We made a new stats table that included all baits (38) and just preys (83) that we had obtained localization data for. We then made the assumption that interactions between baits and preys in spatially different regions (at the organelle level) were non-specific. We took the highest WD-score value in this new stats table and used it as the WD-score cut-off, which, in our case was 47.516. Approximately 3.78% of the data lies above this value, giving 513 interactions involving 398 proteins. A WD-score >47.516 was thus considered a high confidence interacting protein (HCIP).

Comparison of Localization Data with PredAlgo and TargetP

To allow the direct comparison of PredAlgo and TargetP predictions to our localization data, we classified our data as follows: Chloroplast (C) includes “Chloroplast,” "Cytosol and chloroplast," and "Flagella, chloroplast and cytosol." Mitochondria (M) includes “Mitochondria,” "Flagella and mitochondria,” and "Unclear ER or mitochondria." Secretory pathway (SP) includes “Plasma

membrane and late-secretory pathway," "ER," "Unclear ER or mitochondria," "Golgi and secretory pathway," "Cell wall," and "Contractile vacuoles." Other (O) includes "Cytosol," "Flagella," "Flagella and cytosol," "Flagella and mitochondria," "Flagella, chloroplast and cytosol," and "Nucleus." The data used for analysis excluded proteins used in the PredAlgo training set (Tardif et al., 2012).

Gene Expression Values and Presence of Upstream ATGs

Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values were downloaded from Phytozome (https://phytozome.jgi.doe.gov/phytomine/begin.do). For analysis of cloning and localization success relative to transcript abundance, FPKM values for "photo.HighLight MidLog" from the GeneAtlas experiment group were used. These experiments were performed at ambient CO_2 levels (~400 ppm), a CO_2 concentration reflective of our experimental conditions. For an approximation of CCM induction, $\log_2$ FPKM changes were calculated by dividing FPKM values from photo.HighLight MidLog and hetero.Ammonia MidLog experiments of the GeneAtlas experiment group.

An analysis of genes for upstream ATGs (uATGs) was recently performed on version 5.5 of the *Chlamydomonas* genome (Cross, 2016). Comparison of our localization data to the presence of uATGs showed that localization success was 63% (89/141) in the absence of upstream ATGs (uATGs), relative to only 30% (17/57; Figure S1G) when uATGs were found in-frame to the annotated start site in the mRNA (Cross, 2016).

Interestingly, localization success only rose to 40% for both cloned genes that contained an out-of-frame uATG (12/30) and cloned genes that contained an uATG followed by an in-frame stop codon (26/65). This suggests that in some cases out-of-frame uATGs may be the correct translation initiation sites due to unannotated splicing events. Our data is in general

agreement with the analysis by Cross (2016), which proposed that ~10% of current transcript models would result in incorrect translation initiation and incorrect encoded peptides.

P-Type ATPase Tree Assembly

Protein sequences of diverse P-type ATPases (Thever and Saier, 2009) were downloaded from the National Centre for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/). NCBI sequences were combined with six P-type ATPases found in *Chlamydomonas* for a total of 259 sequences. Sequence alignment was performed using ClustalW and a phylogenetic tree created using FastTree2 (Price et al., 2010).

GO Term Analysis

HCIPs of baits that localized to either the chloroplast, mitochondria, nucleus, ER/extracellular or PM were analyzed for GO-term enrichment using the Cytoscape plugin, BINGO (Maere et al., 2005). Preys also included some baits that were detected as HCIPs of other baits. The GO-term, “Generation of precursor metabolites and energy” was shortened to “metabolite and energy production” in Figure 5.

Transmembrane and Protein Disorder Prediction

Protein transmembrane regions were predicted using TMHMM 2.0 (Krogh et al., 2001). The percentage of protein disorder was predicted using ESpritz v1.3 (Walsh et al., 2012) with the prediction type set to Disprot and decision threshold set to Best Sw.

Pyrenoid Enrichment Analysis

To determine whether the pyrenoid showed selectivity regarding protein size we categorized chloroplast localized proteins into pyrenoid depleted or not pyrenoid depleted. The “all other localizations” included all non-chloroplast proteins.

Data Analysis and Visualization

All confocal microscopy images were analyzed using Fiji (Schindelin et al., 2012). WD-score analysis, bait-prey matrix assembly and statistical tests were performed in Microsoft Excel. Hierarchical clustering was done using Multi Experiment Viewer (Saeed et al., 2003). Network visualization and GO analysis was done in Cytoscape (Shannon et al., 2003).

REFERENCES

Atkinson, N., Feike, D., Mackinder, L., Meyer, M.T., Griffiths, H., Jonikas, M.C., Smith, A.M., and McCormick, A.J. (2016). Introducing an algal carbon-concentrating mechanism into higher plants: location and incorporation of key components. Plant Biotechnol J 14, 1302–1315.

Atteia, A., Adrait, A., Brugière, S., Tardif, M., Van Lis, R., Deusch, O., Dagan, T., Kuhn, L., Gontero, B., and Martin, W. (2009). A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the α-proteobacterial mitochondrial ancestor. Molecular Biology and Evolution 26, 1533-1548.

Badger, M. (2003). The roles of carbonic anhydrases in photosynthetic CO_2 concentrating mechanisms. Photosynthesis Res 77, 83-94.

Bauwe, H., Hagemann, M., and Fernie, A.R. (2010). Photorespiration: players, partners and origin. Trends Plant Sci 15, 330-336.

Behrends, C., Sowa, M.E., Gygi, S.P., and Harper, J.W. (2010). Network organization of the human autophagy system. Nature 466, 68-76.

Behrenfeld, M.J., Randerson, J.T., McClain, C.R., Feldman, G.C., Los, S.O., Tucker, C.J., Falkowski, P.G., Field, C.B., Frouin, R., Esaias, W.E*., et al.* (2001). Biospheric Primary Production During an ENSO Transition. Science 291, 2594-2597.

Ben-Shem, A., Frolow, F., and Nelson, N. (2003). Crystal structure of plant photosystem I. Nature 426, 630-635.

Benlloch, R., Shevela, D., Hainzl, T., Grundstrom, C., Shutova, T., Messinger, J., Samuelsson, G., and Sauer-Eriksson, A.E. (2015). Crystal structure and functional characterization of photosystem II-associated carbonic anhydrase CAH3 in *Chlamydomonas reinhardtii*. Plant Physiol 167, 950-962.

Bergeron-Sandoval, L.-P., Safaee, N., and Michnick, S.W. (2016). Mechanisms and consequences of macromolecular phase separation. Cell 165, 1067-1079.

Bern, M., Kil, Y.J., and Becker, C. (2012). Byonic: Advanced Peptide and Protein Identification Software. In Current Protocols in Bioinformatics (John Wiley & Sons, Inc.).

Blanco-Rivero, A., Shutova, T., Roman, M.J., Villarejo, A., and Martinez, F. (2012). Phosphorylation controls the localization and activation of the lumenal carbonic anhydrase in *Chlamydomonas reinhardtii*. PloS one 7, e49063.

Bracher, A., Hauser, T., Liu, C., Hartl, F.U., and Hayer-Hartl, M. (2015). Structural Analysis of the Rubisco-Assembly Chaperone RbcX-II from *Chlamydomonas reinhardtii*. PloS one 10, e0135448.

1132 Brueggeman, A.J., Gangadharaiah, D.S., Cserhati, M.F., Casero, D., Weeks, D.P., and
1133 Ladunga, I. (2012). Activation of the carbon concentrating mechanism by CO_2 deprivation
1134 coincides with massive transcriptional restructuring in *Chlamydomonas reinhardtii*. The Plant
1135 cell 24, 1860-1875.

1136 Cameron, J.C., Wilson, S.C., Bernstein, S.L., and Kerfeld, C.A. (2013). Biogenesis of a bacterial
1137 organelle: the carboxysome assembly pathway. Cell 155, 1131-1140.

1138 Chandrashekar, J., Yarmolinsky, D., von Buchholtz, L., Oka, Y., Sly, W., Ryba, N.J., and Zuker,
1139 C.S. (2009). The taste of carbonation. Science 326, 443-445.

1140 Chevalier, D., Morris, E.R., and Walker, J.C. (2009). 14-3-3 and FHA domains mediate
1141 phosphoprotein interactions. Annu Rev Plant Biol 60, 67-91.

1142 Choi, H.I., Kim, J.Y.H., Kwak, H.S., Sung, Y.J., and Sim, S.J. (2016). Quantitative analysis of
1143 the chemotaxis of a green alga, *Chlamydomonas reinhardtii*, to bicarbonate using diffusion-
1144 based microfluidic device. Biomicrofluidics 10, 014121.

1145 Christianson, J.C., Olzmann, J.A., Shaler, T.A., Sowa, M.E., Bennett, E.J., Richter, C.M., Tyler,
1146 R.E., Greenblatt, E.J., Harper, J.W., and Kopito, R.R. (2012). Defining human ERAD networks
1147 through an integrative mapping strategy. Nat Cell Biol 14, 93-105.

1148 Cross, F.R. (2016). Tying down loose ends in the Chlamydomonas genome: functional
1149 significance of abundant upstream open reading frames. G3: Genes|Genomes|Genetics 6, 435-
1150 446.

1151 Depège, N., Bellafiore, S., and Rochaix, J.-D. (2003). Role of chloroplast protein kinase Stt7 in
1152 LHCII phosphorylation and state transition in *Chlamydomonas*. Science 299, 1572-1575.

1153 Dismukes, G., Klimov, V., Baranov, S., Kozlov, Y.N., DasGupta, J., and Tyryshkin, A. (2001).
1154 The origin of atmospheric oxygen on Earth: the innovation of oxygenic photosynthesis. PNAS
1155 98, 2170-2175.

1156 Duanmu, D., Wang, Y., and Spalding, M.H. (2009). Thylakoid lumen carbonic anhydrase
1157 (CAH3) mutation suppresses air-Dier phenotype of LCIB mutant in *Chlamydomonas reinhardtii*.
1158 Plant Physiol 149, 929-937.

1159 Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in
1160 large-scale protein identifications by mass spectrometry. Nat Methods 4, 207-214.

1161 Ellis, R.J. (1979). The most abundant protein in the world. Trends Biochem Sci 4, 241-244.

1162 Emanuelsson, O., Nielsen, H., Brunak, S., and Von Heijne, G. (2000). Predicting subcellular
1163 localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300, 1005-
1164 1016.

Engel, B.D., Ishikawa, H., Wemmer, K.A., Geimer, S., Wakabayashi, K.-i., Hirono, M., Craige, B., Pazour, G.J., Witman, G.B., and Kamiya, R. (2012). The role of retrograde intraflagellar transport in flagellar assembly, maintenance, and function. The Journal of cell biology 199, 151-167.

Engel, B.D., Schaffer, M., Kuhn Cuellar, L., Villa, E., Plitzko, J.M., and Baumeister, W. (2015). Native architecture of the *Chlamydomonas* chloroplast revealed by in situ cryo-electron tomography. eLife 4.

Fang, W., Si, Y., Douglass, S., Casero, D., Merchant, S.S., Pellegrini, M., Ladunga, I., Liu, P., and Spalding, M.H. (2012). Transcriptome-wide changes in *Chlamydomonas reinhardtii* gene expression regulated by carbon dioxide and the CO_2-concentrating mechanism regulator CIA5/CCM1. The Plant cell 24, 1876-1893.

Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. Science 281, 237-240.

Fuhrmann, M., Oertel, W., and Hegemann, P. (1999). A synthetic gene coding for the green fluorescent protein (GFP) is a versatile reporter in *Chlamydomonas reinhardtii*. The Plant Journal 19, 353-361.

Fujiu, K., Nakayama, Y., Iida, H., Sokabe, M., and Yoshimura, K. (2011). Mechanoreception in motile flagella of *Chlamydomonas*. Nat Cell Biol 13, 630-632.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6, 343-345.

Goldschmidt-Clermont, M., and Rahire, M. (1986). Sequence, evolution and differential expression of the two genes encoding variant small subunits of ribulose bisphosphate carboxylase/oxygenase in *Chlamydomonas reinhardtii*. J Mol Biol 191, 421-432.

Goodenough, U., Blaby, I., Casero, D., Gallaher, S.D., Goodson, C., Johnson, S., Lee, J.-H., Merchant, S.S., Pellegrini, M., Roth, R.*, et al.* (2014). The Path to Triacylglyceride Obesity in the sta6 Strain of *Chlamydomonas reinhardtii*. Eukaryotic Cell DOI: 10.1128/ec.00013-14.

Guruharsha, K.G., Rual, J.F., Zhai, B., Mintseris, J., Vaidya, P., Vaidya, N., Beekman, C., Wong, C., Rhee, D.Y., Cenaj, O.*, et al.* (2011). A protein complex network of *Drosophila melanogaster*. Cell 147, 690-703.

Harada, H., Nakajima, K., Sakaue, K., and Matsuda, Y. (2006). CO_2 sensing at ocean surface mediated by cAMP in a marine diatom. Plant Physiol 142, 1318-1328.

Heinnickel, M., Kim, R.G., Wittkopp, T.M., Yang, W., Walters, K.A., Herbert, S.K., and Grossman, A.R. (2016). Tetratricopeptide repeat protein protects photosystem I from oxidative disruption during assembly. PNAS 113, 2774-2779.

Hu, H., Boisson-Dernier, A., Israelsson-Nordström, M., Böhmer, M., Xue, S., Ries, A., Godoski, J., Kuhn, J.M., and Schroeder, J.I. (2010). Carbonic anhydrases are upstream regulators of CO_2-controlled stomatal movements in guard cells. Nat Cell Biol 12, 87-93.

Huh, W.-K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature 425, 686-691.

Ishihara, S., Takabayashi, A., Ido, K., Endo, T., Ifuku, K., and Sato, F. (2007). Distinct functions for the two PsbP-like proteins PPL1 and PPL2 in the chloroplast thylakoid lumen of Arabidopsis. Plant Physiol 145, 668-679.

Jansen, R., Greenbaum, D., and Gerstein, M. (2002). Relating whole-genome expression data with protein-protein interactions. Genome Research 12, 37-46.

Jin, S., Sun, J., Wunder, T., Tang, D., Cousins, A.B., Sze, S.K., Mueller-Cajar, O., and Gao, Y.-G. (2016). Structural insights into the LCIB protein family reveals a new group of β-carbonic anhydrases. PNAS 113, 14716-14721.

Johnson, D.E., and Casey, J.R. (2011). Cytosolic H^+ microdomain developed around AE1 during AE1-mediated Cl^-/HCO_3^- exchange. The Journal of physiology 589, 1551-1569.

Johnson, X., and Alric, J. (2013). Central carbon metabolism and electron transport in *Chlamydomonas reinhardtii*: metabolic constraints for carbon partitioning between oil and starch. Eukaryotic Cell 12, 776-793.

Karcher, D., Koster, D., Schadach, A., Klevesath, A., and Bock, R. (2009). The *Chlamydomonas* chloroplast HLP protein is required for nucleoid organization and genome maintenance. Molecular plant 2, 1223-1232.

Karlsson, J., Clarke, A.K., Chen, Z.Y., Hugghins, S.Y., Park, Y.I., Husic, H.D., Moroney, J.V., and Samuelsson, G. (1998). A novel α-type carbonic anhydrase associated with the thylakoid membrane in *Chlamydomonas reinhardtii* is required for growth at ambient CO_2. The EMBO Journal 17, 1208-1216.

Kobayashi, Y., Takusagawa, M., Harada, N., Fukao, Y., Yamaoka, S., Kohchi, T., Hori, K., Ohta, H., Shikanai, T., and Nishimura, Y. (2016). Eukaryotic components remodeled chloroplast nucleoid organization during the green plant evolution. Genome Biology and Evolution 8, 1-16.

Koroleva, O.A., Tomlinson, M.L., Leader, D., Shaw, P., and Doonan, J.H. (2005). High-throughput protein localization in Arabidopsis using Agrobacterium-mediated transient expression of GFP-ORF fusions. The Plant Journal 41, 162-174.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P.*, et al.* (2006). Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. Nature 440, 637-643.

1234 Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E.L. (2001). Predicting
1235 transmembrane protein topology with a hidden Markov model: application to complete
1236 genomes. J Mol Biol 305, 567-580.

1237 Kropat, J., Hong-Hermesdorf, A., Casero, D., Ent, P., Castruita, M., Pellegrini, M., Merchant,
1238 S.S., and Malasarn, D. (2011). A revised mineral nutrient supplement increases biomass and
1239 growth rate in *Chlamydomonas reinhardtii*. The Plant Journal 66, 770-780.

1240 Lacoste-Royal, G., and Gibbs, S.P. (1987). Immunocytochemical localization of ribulose-1, 5-
1241 bisphosphate carboxylase in the pyrenoid and thylakoid region of the chloroplast of
1242 *Chlamydomonas reinhardtii*. Plant Physiol 83, 602-606.

1243 Lemeille, S., Willig, A., Depege-Fargeix, N., Delessert, C., Bassi, R., and Rochaix, J.D. (2009).
1244 Analysis of the chloroplast protein kinase Stt7 during state transitions. PLoS Biol 7, e45.

1245 Liu, J., Yang, H., Lu, Q., Wen, X., Chen, F., Peng, L., Zhang, L., and Lu, C. (2012). PsbP-
1246 domain protein1, a nuclear-encoded thylakoid lumenal protein, is essential for photosystem I
1247 assembly in Arabidopsis. The Plant cell 24, 4992-5006.

1248 Long, S.P., Marshall-Colon, A., and Zhu, X.G. (2015). Meeting the global food demand of the
1249 future by engineering crop photosynthesis and yield potential. Cell 161, 56-66.

1250 Lunde, C., Jensen, P.E., Haldrup, A., Knoetzel, J., and Scheller, H.V. (2000). The PSI-H subunit
1251 of photosystem I is essential for state transitions in plant photosynthesis. Nature 408, 613-615.

1252 Ma, Y., Pollock, S.V., Xiao, Y., Cunnusamy, K., and Moroney, J.V. (2011). Identification of a
1253 novel gene, CIA6, required for normal pyrenoid formation in *Chlamydomonas reinhardtii*. Plant
1254 Physiol 156, 884-896.

1255 Mackinder, L.C.M., Meyer, M.T., Mettler-Altmann, T., Chen, V.K., Mitchell, M.C., Caspari, O.,
1256 Freeman Rosenzweig, E.S., Pallesen, L., Reeves, G., Itakura, A.*, et al.* (2016). A repeat protein
1257 links Rubisco to form the eukaryotic carbon-concentrating organelle. PNAS 113, 5958-5963.

1258 Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess
1259 overrepresentation of gene ontology categories in biological networks. Bioinformatics 21, 3448-
1260 3449.

1261 McGrath, J.M., and Long, S.P. (2014). Can the cyanobacterial carbon-concentrating mechanism
1262 increase photosynthesis in crop species? A theoretical analysis. Plant Physiol 164, 2247-2261.

1263 McKay, R., Gibbs, S.P., and Vaughn, K. (1991). Rubisco activase is present in the pyrenoid of
1264 green algae. Protoplasma 162, 38-45.

1265 McKay, R.M.L., and Gibbs, S.P. (1990). Phycoerythrin is absent from the pyrenoid of
1266 *Porphyridium cruentum*: photosynthetic implications. Planta 180, 249-256.

McKay, R.M.L., and Gibbs, S.P. (1991). Composition and function of pyrenoids: cytochemical and immunocytochemical approaches. Canadian Journal of Botany 69, 1040-1052.

Meyer, M.T., McCormick, A.J., and Griffiths, H. (2016). Will an algal CO_2-concentrating mechanism work in higher plants? Curr Opin Plant Biol 31, 181-188.

Mitra, M., Lato, S.M., Ynalvez, R.A., Xiao, Y., and Moroney, J.V. (2004). Identification of a new chloroplast carbonic anhydrase in *Chlamydomonas reinhardtii*. Plant Physiol 135, 173-182.

Moroney, J.V., Ma, Y., Frey, W.D., Fusilier, K.A., Pham, T.T., Simms, T.A., DiMario, R.J., Yang, J., and Mukherjee, B. (2011). The carbonic anhydrase isoforms of *Chlamydomonas reinhardtii*: intracellular location, expression, and physiological roles. Photosynthesis Res 109, 133-149.

Morris, J.H., Knudsen, G.M., Verschueren, E., Johnson, J.R., Cimermancic, P., Greninger, A.L., and Pico, A.R. (2014). Affinity purification–mass spectrometry and network analysis to understand protein-protein interactions. Nature protocols 9, 2539-2554.

Mustardy, L., Cunningham, F.X., and Gantt, E. (1990). Localization and quantitation of chloroplast enzymes and light-harvesting components using immunocytochemical methods. Plant Physiol 94, 334-340.

Neupert, J., Karcher, D., and Bock, R. (2009). Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. The Plant Journal 57, 1140-1150.

Ohnishi, N., Mukherjee, B., Tsujikawa, T., Yanase, M., Nakano, H., Moroney, J.V., and Fukuzawa, H. (2010). Expression of a low CO_2-inducible protein, LCI1, increases inorganic carbon uptake in the green alga *Chlamydomonas reinhardtii*. The Plant cell 22, 3105-3117.

Pazour, G.J., Agrin, N., Leszyk, J., and Witman, G.B. (2005). Proteomic analysis of a eukaryotic cilium. The Journal of cell biology 170, 103-113.

Portis, A.R., and Parry, M.A. (2007). Discoveries in Rubisco (Ribulose 1, 5-bisphosphate carboxylase/oxygenase): a historical perspective. Photosynthesis Res 94, 121-143.

Price, G., and Badger, M. (1989). Expression of Human Carbonic Anhydrase in the Cyanobacterium *Synechococcus* PCC7942 Creates a High CO_2-Requiring Phenotype Evidence for a Central Role for Carboxysomes in the CO_2 Concentrating Mechanism. Plant Physiol 91, 505-513.

Price, G.D., Badger, M.R., Woodger, F.J., and Long, B.M. (2008). Advances in understanding the cyanobacterial CO_2-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. J Exp Bot 59, 1441-1461.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. PloS one 5, e9490.

1301 Qu, Z., and Hartzell, H.C. (2008). Bestrophin Cl^- channels are highly permeable to HCO_3^-.
1302 American Journal of Physiology-Cell Physiology 294, C1371-C1377.

1303 Rae, B.D., Long, B.M., Badger, M.R., and Price, G.D. (2013). Functions, compositions, and
1304 evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO_2
1305 fixation in cyanobacteria and some proteobacteria. Microbiol Mol Biol Rev 77, 357-379.

1306 Rasala, B.A., Lee, P.A., Shen, Z., Briggs, S.P., Mendez, M., and Mayfield, S.P. (2012). Robust
1307 expression and secretion of Xylanase1 in *Chlamydomonas reinhardtii* by fusion to a selection
1308 gene and processing with the FMDV 2A peptide. PloS one 7, e43349.

1309 Reinfelder, J.R. (2011). Carbon concentrating mechanisms in eukaryotic marine phytoplankton.
1310 Annual review of marine science 3, 291-315.

1311 Ris, H., and Plaut, W. (1962). Ultrastructure of DNA-containing areas in the chloroplast of
1312 *Chlamydomonas*. The Journal of cell biology 13, 383.

1313 Robellet, X., Flipphi, M., Pégot, S., MacCabe, A.P., and Vélot, C. (2008). AcpA, a member of
1314 the GPR1/FUN34/YaaH membrane protein family, is essential for acetate permease activity in
1315 the hyphal fungus *Aspergillus nidulans*. Biochemical Journal 412, 485-493.

1316 Rochaix, J.-D., Perron, K., Dauvillée, D., Laroche, F., Takahashi, Y., and Goldschmidt-
1317 Clermont, M. (2004). Post-transcriptional steps involved in the assembly of photosystem I in
1318 *Chlamydomonas*. Biochem Soc Trans 32, 567-570.

1319 Rosenbaum, J.L., and Witman, G.B. (2002). Intraflagellar transport. Nature Reviews Molecular
1320 Cell Biology 3, 813-825.

1321 Sá-Pessoa, J., Paiva, S., Ribas, D., Silva, I.J., Viegas, S.C., Arraiano, C.M., and Casal, M.
1322 (2013). SATP (YaaH), a succinate–acetate transporter protein in *Escherichia coli*. Biochemical
1323 Journal 454, 585-595.

1324 Saeed, A., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier,
1325 T., Thiagarajan, M.*, et al.* (2003). TM4: a free, open-source system for microarray data
1326 management and analysis. Biotechniques 34, 374.

1327 Sarov, M., Murray, J.I., Schanze, K., Pozniakovski, A., Niu, W., Angermann, K., Hasse, S.,
1328 Rupprecht, M., Vinis, E., Tinney, M.*, et al.* (2012). A genome-scale resource for in vivo tag-
1329 based protein function exploration in *C. elegans*. Cell 150, 855-866.

1330 Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch,
1331 S., Rueden, C., Saalfeld, S., Schmid, B.*, et al.* (2012). Fiji: an open-source platform for
1332 biological-image analysis. Nat Methods 9, 676-682.

1333 Schottkowski, M., Peters, M., Zhan, Y., Rifai, O., Zhang, Y., and Zerges, W. (2012). Biogenic
1334 membranes of the chloroplast in *Chlamydomonas reinhardtii*. PNAS 109, 19286-19291.

1335 Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N.,
1336 Schwikowski, B., and Ideker, T. (2003). Cytoscape: A Software Environment for Integrated
1337 Models of Biomolecular Interaction Networks. Genome Research 13, 2498-2504.

1338 Shively, J., Ball, F., Brown, D., and Saunders, R. (1973). Functional organelles in prokaryotes:
1339 polyhedral inclusions (carboxysomes) of *Thiobacillus neapolitanus*. Science 182, 584-586.

1340 Sinetova, M.A., Kupriyanova, E.V., Markelova, A.G., Allakhverdiev, S.I., and Pronina, N.A.
1341 (2012). Identification and functional role of the carbonic anhydrase Cah3 in thylakoid
1342 membranes of pyrenoid of *Chlamydomonas reinhardtii*. Biochim Biophys Acta Biochim Biophys
1343 Acta 1817, 1248-1255.

1344 Sowa, M.E., Bennett, E.J., Gygi, S.P., and Harper, J.W. (2009). Defining the human
1345 deubiquitinating enzyme interaction landscape. Cell 138, 389-403.

1346 Tardif, M., Atteia, A., Specht, M., Cogne, G., Rolland, N., Brugière, S., Hippler, M., Ferro, M.,
1347 Bruley, C., and Peltier, G. (2012). PredAlgo: a new subcellular localization prediction tool
1348 dedicated to green algae. Molecular Biology and Evolution, mss178.

1349 Terashima, M., Specht, M., Naumann, B., and Hippler, M. (2010). Characterizing the anaerobic
1350 response of *Chlamydomonas reinhardtii* by quantitative proteomics. Molecular & Cellular
1351 Proteomics 9, 1514-1532.

1352 Thever, M.D., and Saier, M.H. (2009). Bioinformatic characterization of p-type ATPases
1353 encoded within the fully sequenced genomes of 26 eukaryotes. Journal of Membrane Biology
1354 229, 115-130.

1355 Tian, G.-W., Mohanty, A., Chary, S.N., Li, S., Paap, B., Drakakaki, G., Kopec, C.D., Li, J.,
1356 Ehrhardt, D., and Jackson, D. (2004). High-throughput fluorescent tagging of full-length
1357 Arabidopsis gene products in planta. Plant Physiol 135, 25-38.

1358 Tresguerres, M., Buck, J., and Levin, L.R. (2010). Physiological carbon dioxide, bicarbonate,
1359 and pH sensing. Pflügers Archiv-European Journal of Physiology 460, 953-964.

1360 Turkina, M.V., Blanco-Rivero, A., Vainonen, J.P., Vener, A.V., and Villarejo, A. (2006). CO_2
1361 limitation induces specific redox-dependent protein phosphorylation in *Chlamydomonas*
1362 *reinhardtii*. Proteomics 6, 2693-2704.

1363 Uniacke, J., and Zerges, W. (2009). Chloroplast protein targeting involves localized translation
1364 in *Chlamydomonas*. PNAS 106, 1439-1444.

1365 Updike, D.L., Hachey, S.J., Kreher, J., and Strome, S. (2011). P granules extend the nuclear
1366 pore complex environment in the *C. elegans* germ line. The Journal of cell biology 192, 939-
1367 948.

Villarejo, A., Shutova, T., Moskvin, O., Forssén, M., Klimov, V.V., and Samuelsson, G. (2002). A photosystem II-associated carbonic anhydrase regulates the efficiency of photosynthetic oxygen evolution. The EMBO Journal 21, 1930-1938.

Walsh, I., Martin, A.J.M., Di Domenico, T., and Tosatto, S.C.E. (2012). ESpritz: accurate and fast prediction of protein disorder. Bioinformatics 28, 503-509.

Wang, L., Yamano, T., Takane, S., Niikawa, Y., Toyokawa, C., Ozawa, S.-i., Tokutsu, R., Takahashi, Y., Minagawa, J., Kanesaki, Y.*, et al.* (2016). Chloroplast-mediated regulation of CO_2-concentrating mechanism by Ca^{2+}-binding protein CAS in the green alga *Chlamydomonas reinhardtii*. PNAS 113, 12586-12591.

Wang, Y., and Spalding, M.H. (2006). An inorganic carbon transport system responsible for acclimation specific to air levels of CO_2 in *Chlamydomonas reinhardtii*. PNAS 103, 10110-10115.

Wang, Y., Stessman, D.J., and Spalding, M.H. (2015). The CO_2 concentrating mechanism and photosynthetic carbon assimilation in limiting CO_2: how *Chlamydomonas* works against the gradient. The Plant Journal 82, 429-448.

Yamano, T., Miura, K., and Fukuzawa, H. (2008). Expression analysis of genes associated with the induction of the carbon-concentrating mechanism in *Chlamydomonas reinhardtii*. Plant Physiol 147, 340-354.

Yamano, T., Sato, E., Iguchi, H., Fukuda, Y., and Fukuzawa, H. (2015). Characterization of cooperative bicarbonate uptake into chloroplast stroma in the green alga *Chlamydomonas reinhardtii*. PNAS 112, 7315-7320.

Yamano, T., Tsujikawa, T., Hatano, K., Ozawa, S., Takahashi, Y., and Fukuzawa, H. (2010). Light and low-CO_2-dependent LCIB-LCIC complex localization in the chloroplast supports the carbon-concentrating mechanism in *Chlamydomonas reinhardtii*. Plant Cell Physiol 51, 1453-1468.

Zimmerman, K., and Yoder, B.K. (2015). SnapShot: sensing and signaling by cilia. Cell 161, 692.