# Exploring the organelle assembly mechanism of the *Chlamydomonas reinhardtii* pyrenoid through Proximity Labelling

Chun Sing Lau

Doctor of Philosophy

University of York
Biology

July 2024

## Abstract

A significant fraction of global carbon fixation occurs in a phase-separated organelle called the pyrenoid. To achieve this, the pyrenoid participates in a process called the $CO_2$ concentrating mechanism (CCM) that raises the cell's internal inorganic carbon concentration by up to a 100-fold above the external concentration, thereby suppressing photorespiration and increasing the photosynthetic efficiency. The CCM of *Chlamydomonas reinhardtii* represents one of the best characterised pyrenoid-based CCM to date, however the assembly mechanism of the pyrenoid structural components: the starch sheath and thylakoid tubules is relatively unclear. To accelerate this process and identify proteins which might contribute to their biogenesis, I developed a TurboID-based proximity labelling pipeline which allows efficient pulldown of proteins based on their proximity to the tagged Protein-of-interest. By targeting Rubisco, and other pyrenoid proteins thought to be critical for the assembly of the starch sheath and thylakoid tubules, I obtained a dataset containing previously identified pyrenoid proteins and a list of high-confidence novel pyrenoid proteins. Within this list, PINS2 was taken forward for further characterisation due to its mesh-like localisation which parallels another starch-related protein LCI9. This homology and proximity in their chromosomal position led to the identification of LCI9/PINS1, PINS2, and PINS3 as a three-gene family sharing a similar domain arrangement. Via TEM imaging, I observed that the pyrenoid starch sheath in the *pins2* mutant exhibits aberrant overlap that are uncommon in wildtype strains. Further, the *pins2* mutant displays a high-$CO_2$ requiring phenotype suggesting a defect in the CCM. The carbonic anhydrase LCIB alters its localisation from a diffuse stromal pattern to the pyrenoid periphery in sub-air level $CO_2$. This re-localisation is proposed to be critical for maintaining CCM efficiency by scavenging the $CO_2$ leaked from the pyrenoid. I observed that the LCIB re-localisation towards the pyrenoid is perturbed in the *pins2* mutant. This highlights the importance of a correctly formed starch sheath to the LCIB-mediated $CO_2$-scavenging system. Together, our results indicate that the PINS proteins collectively alter the pyrenoid starch sheath through their different localisation and interaction with starch synthesis enzymes on the starch sheath gap, this in turn builds a canonical starch sheath necessary for LCIB recruitment.

## Author's Declaration

I declare that this thesis is a presentation of original work, and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

The work in this thesis has either been published or are in preparation for submission.

**Chapter 2 - A phase-separated $CO_2$-fixing pyrenoid proteome determined by TurboID in *Chlamydomonas reinhardtii***

*The Plant Cell*
Published: 17 May 2023
DOI: https://doi.org/10.1093/plcell/koad131
PMID: 37195994

**Chapter 3 - Proteomic mapping of *Chlamydomonas* pyrenoid interfaces**

*In preparation for submission to Plant Physiology*

**Chapter 4 - The starch sheath peripheral coiled-coil protein PINS2 is important for pyrenoid starch sheath formation**

*In preparation for submission to Plant Physiology*

## Acknowledgements

I could not have possibly finished this degree without the care and support from all my family and friends. First, I would like to extend my most sincere gratitude to my parents, who has been nothing but patient and supportive of my decision to chase an academic dream in a country far away from home. From my undergraduate to this PhD, I've only been home ever so shortly, and always hopping back on a plane after a short stay back home. Our video calls have always given me the strength continue my work. I'm forever in debt to the both of you and I hope I have made you proud.

This PhD would not have been remotely as enjoyable as it was if not for my lovely lab mates past and present. I'd like to thank Christine for our little gossip sessions. It always cheers me up to chat with you in the lab, and to confuse everybody by throwing in a few English words in between our chats. I'd also like to thank Sabina and Matt who are a joy to be around, and for the countless drunken nights we share. I'd like to thank James who continues to push me to be a more creative scientist. Collectively I'd like to give a big shoutout to my senior colleagues who I was very fortunate to have given me immense support and was never annoyed by my oftentimes idiotic questions about basic science. Specifically, Tommy who was very patient in training me when I started my masters and Gary who is really my first lab work mentor. Both of whom have taught me many valuable lessons on not only practical lab work but also the importance of planning. I'd like to thank Charlotte for her gentle reminder that the precious work-life balance is what makes us love what we do, and perhaps more importantly, provided emotional support when I really needed it. I wish to also extend my gratitude to the rest of the Mackinder lab, Philipp, Onyou, Katy, Gaurav, Mihris, and Abi. All of you have made the lab an incredibly friendly place and have made some of my late nights in the lab infinitely more fun.

My PhD would not have been possible without my two supervisors. I wish to thank Prof. Luke Mackinder, who has been the most supportive and kind supervisor I've ever had. Without his direction, support and ideas, I have no doubt this PhD would have been much more difficult. On top of that, I'd like to thank him for taking a chance on a student who has little experience in the lab, allowing me to study for both my master's and PhD in his group. It is during this time that I felt that I truly matured as an independent researcher, I'm eternally grateful for his gentle encouragement when I doubt my own capability, and for fostering such a friendly and compassionate environment which people can thrive. I'd also like to thank my co-supervisor Prof Gavin Thomas, who has been highly supportive throughout my PhD and welcomed the addition of a little green research into their primarily bacterial research group. It was always an enjoyable and perhaps nerve-wrecking challenge to be probed on the fundamental details of my work, which I sometimes would not have thought about.

I’d also like to acknowledge the incredible technology facility team who has been very helpful throughout this PhD for their mass spectrometry analysis, confocal microscopy and TEM training. Particularly I’d like to thank Dr. Adam Dowle, Dr. Grant Calder, Dr. Meg Stark, and Dr. Clare Steele-King who have either taught me how to operate different machines (and put up with my incessant call for help); or have worked their magic to make sure I obtained the data that best answers my questions.

I’m incredibly privileged to have worked on such a collaborative project and I’d like to thank the Jonikas group and McCormick group who have given me a lot of constructive feedback during my PhD. I’d also like to thank the Gates foundation for providing the funding for this PhD project.

Outside of the lab, I’d like to thank my partner Tom who has lovingly supported me during this long writing process. He was never annoyed even when I’m inevitably late to dinner because of yet another western blot; and has always cheered me on whenever I’m feeling down. I truly cherish our time together and here is to many more exciting adventures and holidays. To my friends Jack, Simon, Oliver, Tony, Adam, Danny and Preston. While we have all been separated geographically since university, our gatherings are always so fun and take me back in time. Thank you for keeping an eye on me and making me feel so loved.

This PhD was an incredible experience that I wouldn’t trade anything for. While it was not without its low moments, I felt that I have come out more resilient, confident and more curious than when I started. I hope that I can go on many more years in science to keep exploring the unknown.

# Table of Contents

# Table of Figures

# Table of Supplemental Figure

# Table of Supplemental Data Set

# Chapter 1 - Introduction

## The need for $CO_2$ concentrating mechanisms in aquatic photoautotrophs

Much of the world's photosynthesis occurs not in the luscious greenery on land but in the expansive aquatic environment encompassing the vast oceans and freshwater ecosystems. There, photosynthesis is driven by the diverse species of aquatic photosynthetic microbes including the prokaryotic cyanobacteria and the eukaryotic algae. Estimates of their contribution to the global carbon cycling suggests that aquatic photosynthetic organisms are responsible for nearly 50% of all the photosynthesis that takes place on earth, which marks their striking importance to the global carbon cycle (Falkowski and Raven 2013; Prasad et al. 2021). A common limitation shared in aquatic environments is the shortage of the primary substrate of photosynthesis, $CO_2$. The low $CO_2$ concentration in water stems from two major root causes, the limited $CO_2$ gas dissolution into a liquid, which is compounded by the slow diffusion of dissolved $CO_2$ across the liquid phase; and the natural interconversion of dissolved $CO_2$ into membrane impermeable $HCO_3^-$ species. The resulting accessible dissolved $CO_2$ concentration in water is thus almost constantly sub-optimal to support photosynthesis. Especially as the principal carbon-fixation enzyme – Rubisco (Ribulose-1,5-bisphosphate Carboxylase/Oxygenase) - exhibits affinity to oxygen, which drives the production of the wasteful compound 2-phosphoglycolate (2-PG). The salvaging of which by the photorespiration pathway expends both ATP and NADPH and results in a net-loss of carbon. While recent evidence of the photorespiration pathway suggests that the 2-PG metabolism is important for nitrogen assimilation and amino acid synthesis, renewing the perspective of photorespiration as possibly not an energetically wasteful process (Eisenhut et al. 2008; Hagemann et al. 2016; Shi and Bloom 2021), it is undeniable that this alternative oxygenation pathway further restricts photosynthesis efficiency in a $CO_2$-scarce environment in which many aquatic photosynthetic organisms are exposed to. Such difficulty likely drove the evolution of molecular mechanism to suppress photorespiration by either altering the $CO_2/O_2$ specificity of Rubisco or implementing a $CO_2$ concentrating mechanism (CCM) (Rickaby and Eason Hubbard 2019). As its name suggests, the latter functions by driving the active delivery of $CO_2$ towards the site of Rubisco, which saturates and thereby suppresses Rubisco's oxygenation activity. To achieve this, algae evolved both Inorganic carbon ($C_i$) delivery (such as $C_i$ transporters and carbonic anhydrases) and structural components (pyrenoid and carboxysome) that are able to concentrate $CO_2$ surrounding Rubisco. The form and organisation of CCM components across algae species differs greatly.

### The diversity of organisation and structure in CCMs

Algal CCMs are incredibly diverse in their configuration. On a simpler configuration spectrum, a study in *Nanochloropsis oceanica* reports its CCM is driven primarily through an

$C_i$ uptake system consisting of $HCO_3^-$ transporters and a carbonic anhydrase CAH1. It was proposed that putative $HCO_3^-$ transporters deliver $HCO_3^-$ from external environment towards the epiplastid Endoplasmic Reticulum (ER). There the ER-luminal carbonic anhydrase CAH1 converts the accumulated $HCO_3^-$ into $CO_2$, raising the local concentration of $CO_2$ surrounding the chloroplast. This builds a concentration gradient of $CO_2$ to drive their passive diffusion into the chloroplast stroma and thereby raising the $CO_2$ concentration surrounding Rubisco (Gee and Niyogi 2017). A similar mechanism has also been proposed for *Cyanidioschyzon merolae* (Steensma et al. 2023). These $HCO_3^-$ transporter/carbonic anhydrase systems represent a relatively simpler CCM organisation, and might be “leaky”, referring to the loss of $CO_2$ from passive diffusion outwards of the cytosol that reduces their efficiency. Land plants similarly evolved CCMs and counter this leakiness by sequestering $CO_2$ (in the form of $HCO_3^-$) into C4 sugars and partitioning the $CO_2$ release process into specialised bundle sheath cells, where a suberin deposition limits $CO_2$ diffusion (Ermakova et al. 2020; Danila et al. 2021). Lacking the multicellular structure necessary to divide $CO_2$ sequestration and release, many CCM-containing algae instead exhibit a biophysical CCM that takes the form of a Rubisco-dense microcompartment which allows a similar partitioning of $CO_2$ release at the site of Rubisco. The Rubisco microcompartments vary greatly in their structure and organisation, though most aquatic photosynthetic organisms present two major class of microcompartments. One being carboxysomes found in cyanobacteria species; and the other the pyrenoid that is common to eukaryotic algae.

Cyanobacteria species can be classified into two distinct groups based on the Rubisco form it possesses (Form 1A and 1B). Cyanobacteria that carry the form 1A Rubisco within their α-carboxysomes are called α-cyanobacteria while those carrying form 1B Rubisco within β-carboxysomes are called β-cyanobacteria. The two carboxysomes systems evolved convergently, with α-carboxysome possibly arising from proteobacteria via horizontal gene transfer event, while β-carboxysome evolution remains unclear (Kerfeld and Melnicki 2016). As both classes of carboxysomes share a similar architectural and functional principle, here I focus on the β-carboxysome as an example. In *Synechococcus elongatus* PCC 7942 (Synechococcus hereafter), the carboxysomes is represented as an icosahedral proteinaceous microcompartment. The β-carboxysome resides in the cytosol and consists of a putative $CO_2$ impermeable protein shell which contains the Rubisco holoenzyme in a paracrystalline-like packing, and a carbonic anhydrase (Kaneko et al. 2006; Rae et al. 2013). The unique icosahedral geometry of the protein shell stems from the hexameric CcmK2 – 4 and CcmO proteins which occupy the polyhedron’s face (Kerfeld et al. 2005), while a pentameric CcmL protein caps the vertices of the polyhedron (Tanaka et al. 2008). Rubisco within the β-carboxysome is organised by the two isoforms of CcmM protein: CcmM-58 and

CcmM-35. Both CcmM isoforms contain a Rubisco small subunit-like domain (SSUL) which binds to the Rubisco holoenzyme between two rbcL dimers at the equatorial region and drives the liquid-liquid phase separation of the Rubisco enzyme (Wang et al. 2019b; Zang et al. 2021). The longer isoform (CcmM-58) additionally contains a carbonic anhydrase like domain (γCAL) which mediates the intermolecular interaction between the bound Rubisco, carbonic anhydrase, and the M58 isoform. These interactions are hypothesised to be important during carboxysome assembly, which initiates with the nucleation of Rubisco holoenzyme, likely driven by their interaction with CcmM-35. The nascent Rubisco condensate then rapidly recruits shell proteins CcmN/O/K2 further encapsulating the condensate. Vertex formation driven by the CcmL protein ultimately caps the carboxysome shell leading to a fully matured carboxysome (Cameron et al. 2013). New carboxysomes can therefore arise from Rubisco assemblies initiated independently or from ruptured carboxysomes that act as a new assembly template (Chen et al. 2013).

The $C_i$ delivery mechanism which transports external $C_i$ into the carboxysome is equally essential to the CCM functioning. In Synechococcus, external $HCO_3^-$ is transported into the cytosol by three different transporters, namely SbtA, BicA and BCT1. The three transporters vary greatly in their configuration, with SbtA and BicA being single subunit $Na^+/HCO_3^-$ transporters while BCT1 is a 4 subunit ATP-binding cassette (ABC) transporter (Omata et al. 1999; Shibata et al. 2002; Wang et al. 2019a; Rottet et al. 2021). Aside from their classification, the transporter kinetics differ greatly between the three transporters. BCT1 and SbtA exhibit a high $C_i$ affinity but a lower flux, BicA, in contrast, shows a low $C_i$ affinity and high flux (Rottet et al. 2021). These differences in their kinetics likely enabled the cyanobacteria cell to efficiently transport $C_i$ across the plasma membrane even in a wide range of external $C_i$ concentrations.

Once within the cytosol, $HCO_3^-$ needs to cross the carboxysome shell to reach the Rubisco matrix. This transport step is proposed to occur via the small central pore formed on the oligomeric CCM shell proteins. The modelling results of these pores suggests that they exhibit an anion permeability and would therefore promote the transport of $HCO_3^-$ and necessary metabolite into the carboxysome for Rubisco function (Mahinthichaichan et al. 2018; Faulkner et al. 2020). Additionally, a recent report suggests that the homo- and hetero-hexameric complex of CCM shell proteins are able to form higher-order oligomerisation with a capped (doubled) arrangement (Samborska and Kimber 2012; Sommer et al. 2019). As the capped arrangement changes the central pore structure, it was proposed that these formations alter the $HCO_3^-$ or metabolite permeability, and suggests that the process of $HCO_3^-$ uptake into carboxysome might be tightly controlled. (Samborska and Kimber 2012; Sommer et al. 2019). Within the carboxysome, a carbonic anhydrase CcaA is

present to convert the accumulated $HCO_3^-$ into $CO_2$, therefore supplying the condensed Rubisco with a saturating concentration of $CO_2$ (Long et al. 2007). All together, these components form the basis of the carboxysome-based CCM that is found across many prokaryotic species.

## An overview of the Chlamydomonas CCM

### The Rubisco-rich pyrenoid condensate

Moving across to the eukaryotic tree of life, the biophysical CCM within *Chlamydomonas reinhardtii* is one of the best characterised systems in eukaryotic algae. In *Chlamydomonas reinhardtii* (Chlamydomonas hereafter), the Rubisco-dense microcompartment is represented by the pyrenoid organelle. The pyrenoid is an electron-dense organelle observable in transmission electron micrographs and was initially suggested to be a protein-rich cell body that might function as a site of protein storage (Griffiths 1980). Subsequent studies that probed for pyrenoid composition through immunogold labelling (Borkhsenious et al. 1998) identified that Rubisco is the main constituent of the pyrenoid, with ~90% of the cells Rubisco localised within the pyrenoid. This highlighted the pyrenoid's function in carbon fixation.

In the recent decade, an increasing body of research highlighted that cells possess dynamic biomolecular condensates that arise from a phenomenon termed phase separation. This phase-separation is characterised by multivalent interactions between intrinsically disordered proteins as well as nucleic acids. The resulting phase-separated bodies display liquid-like behaviour such as division by fission and fusion. In Chlamydomonas, an intrinsically disordered protein called EPYC1 (Essential Pyrenoid Component 1) co-immunoprecipitated with Rubisco holoenzyme. Abolishing the expression of *EPYC1* caused the pyrenoid matrix to essentially dissolve (Mackinder et al. 2016). Later, *in vivo* characterisation of the Chlamydomonas pyrenoid system using live-cell imaging of fluorescently tagged EPYC1 and RBCS1 provided unequivocal evidence on the pyrenoid's liquid-like behaviour (Freeman Rosenzweig et al. 2017). Combined with the observation that the addition of EPYC1 alone is able to demix Rubisco into phase separated condensates *in vitro* (Wunder et al. 2018), these data affirmed that the formation of the phase-separated pyrenoid matrix is driven by EPYC1-Rubisco interactions. Follow up work has further characterised the molecular mechanism of EPYC1 and Rubisco interaction. An earlier study identified the surface residues of the two α-helix of RBCS to be critical for pyrenoid formation (Meyer et al. 2012). This was later confirmed structurally through cryo-EM single particle analysis, which demonstrated the binding site of EPYC1 on the RBCS alpha helices through a combination of hydrophobic interactions and salt bridge formation (He et al. 2020). The repeated sequence motif on

EPYC1 that binds to Rubisco is termed the Rubisco binding motif (RBM), and was later shown to be common in many pyrenoid targeted proteins. This motif was further suggested to be important for directing the assembly of the pyrenoid organelle (Meyer et al. 2020).

**The $CO_2$ delivery system across multiple membranes**

The Chlamydomonas CCM system is not operated by the Rubisco condensate alone and instead requires the cooperative action of a $C_i$ uptake system that spans the plasma membrane, chloroplast envelope, and thylakoid membrane to direct the extracellular $C_i$ into the pyrenoid matrix and leading to the eventual release as $CO_2$. Starting from the plasma membrane, a periplasmic carbonic anhydrase, CARBONIC ANHYDRASE 1 (CAH1), was proposed to be important for $C_i$ uptake. There is a long-standing debate to CAH1's importance to the cells CCM, as the use of carbonic anhydrase inhibitor acetazolamide significantly reduces the wildtype's photosynthetic efficiency at high pH, which argues for the role of an external carbonic anhydrase in $C_i$ uptake (Moroney et al. 1985). This is however, contrasted by a study on a CAH1-deficient mutant which found a lack of high-$CO_2$ requiring phenotype, and the minimal reduction in oxygen evolution activity compared to wildtype (Van and Spalding 1999). A recent report further elucidated CAH1's molecular function by making use of a CRISPR-Cas9 generated *CAH1* mutant *cah1* (Shimamura et al. 2024). By measuring the $C_i$ affinity [$K_{1/2}$ ($C_i$)], a parameter that denotes the required external $C_i$ concentration necessary to attain the half-maximal rate of photosynthetic oxygen evolution, the authors found that *cah1* exhibits a high-pH dependent reduction in $C_i$ affinity (Shimamura et al. 2024). More importantly, the observed reduction in $C_i$ affinity is complemented by the addition of external carbonic anhydrase confirming that extracellular carbonic anhydrase activity is critical for maintaining $C_i$ availability.

Two plasma membrane proteins have been identified to be responsible for $C_i$ uptake into the cytosol. The two proteins being HIGH-LIGHT ACTIVABLE 3 (HLA3), and LOW $CO_2$ INDUCIBLE 1 (LCI1). HLA3 belongs to the ABC-transporter superfamily and was identified in a panel of genes which shows transcript dependence on both high-light and low $CO_2$ concentration (Im and Grossman 2002). It was additionally shown to be under the control of CIA5/CCM1 (Im and Grossman 2002), which is widely recognised as the master regulator of CCM genes in Chlamydomonas (Moroney et al. 1989; Wang et al. 2005; Duanmu et al. 2009a; Brueggeman et al. 2012), implicating the *HLA3* gene to CCM functioning. Two lines of evidence support HLA3 function as a $C_i$ transporter across the plasma membrane. The first one being that the *HLA3* RNAi knockdown strain exhibits a moderately reduced $C_i$ affinity when it is grown in alkaline condition (Duanmu et al. 2009a). Secondly, when HLA3 expression is induced at high $CO_2$ level, a condition where *HLA3* is not normally expressed, cells are found to exhibit a significantly higher $C_i$ affinity than its high $CO_2$ acclimated

background, indicating the partial activation of the CCM (Gao et al. 2015). Together, current data suggests HLA3 likely functions to preferentially transport $HCO_3^-$ across the plasma membrane. The second $C_i$ uptake component LCI1 was similarly induced by low $CO_2$ conditions. LCI1 response to low $CO_2$ is also under the regulation of CIA5 expression, particularly by its downstream transcriptional factor LOW $CO_2$ STRESS RESPONSE (LCR1) (Burow et al. 1996; Yoshioka et al. 2004; Ohnishi et al. 2010). The artificial induction of LCI1 in the *lcr1* background is found to increase the cells $C_i$ uptake as determined by [$^{14}$C] $C_i$ labelling (Ohnishi et al. 2010). Analysis of the genetic knock-out of the LCI1 further verified their contribution to $C_i$ uptake, where the single *lci1* mutant exhibits reduced $C_i$-dependent oxygen evolution activity that is more apparent in acidic or neutral pH condition, suggesting a preference to $CO_2$ uptake (Kono and Spalding 2020). Indeed, in the recently solved LCI1 crystal structure, LCI1 is found to form a trimeric protein channel. The protein channel is characterised by a strongly electronegative pore which infers the transport of either a positive or neutral metabolite, therefore precluding the transport of the anionic $HCO_3^-$ ions. A putative $CO_2$ ligand was also found on the cytoplasmic facing end of each LCI1 subunit, further supporting that $CO_2$ is the $C_i$ species transported by LCI1 (Kono et al. 2020). Collectively, these results demonstrate that the transport activity of HLA3 and LCI1 drives $C_i$ delivery into the cytosol with the help of CAH1.

Once reaching the cytosol, the accumulated $C_i$ still needs to be transported across one more membrane to the chloroplast stroma. Here a Chloroplast envelope protein called LOW $CO_2$ INDUICBLE A (LCIA) has been identified to drive an ATP-dependent $HCO_3^-$ transport across the chloroplast envelope. LCIA is a homologue of the nitrate transporter belonging to the formate/nitrite antiporter family. However, LCIA expression is independent of nitrogen source and is instead upregulated by low $CO_2$ concentration. Mutation in the *LCIA* gene was found to cause reduced $C_i$ affinity particularly at alkaline pH, suggesting their preference to transport $HCO_3^-$ (Wang and Spalding 2014). LCIA $C_i$ transporter activity is further validated by its exogenous expression in the *Xenopus* oocytes which found LCIA permeability to $HCO_3^-$ and $NO_3^-$ (Mariscal et al. 2006; Atkinson et al. 2016); and its expression in the CA-deficient *E. coli* strain (DCAKO) is able to complement DCAKO's high-$CO_2$ requiring phenotype, likely by repleting the $HCO_3^-$ that is diminished in the DCAKO strain (Förster et al. 2023).

At the chloroplast stroma, the protein LOW $CO_2$ INDUCIBLE B (LCIB) was identified to be critical for CCM functioning. Despite LCIB displaying little sequence homology to carbonic anhydrase, their solved crystal structures suggests LCIB belong to a class of β-Carbonic anhydrase (Jin et al. 2016). Their carbonic anhydrase activity was also subsequently demonstrated in yeast and Arabidopsis cells lacking native carbonic anhydrase (Kasili et al.

2023). The LCIB mutant *ad1* exhibits a unique growth phenotype where cells are found to die in air level $CO_2$ (0.03 – 0.05% $CO_2$) while surviving the more limiting very low $CO_2$ conditions (<0.02% $CO_2$) (Wang and Spalding 2006). This air-dier phenotype highlights that Chlamydomonas operates two separate but complementary $C_i$ uptake systems according to the surrounding $CO_2$ concentration. This differential contribution is further exemplified in the observation that the additive deletion of LCIA/LCIB causes cells to grow in neither limiting $C_i$ conditions (Wang and Spalding 2014). It is therefore posited that at air-level $CO_2$ concentration, LCIB drives the preferential conversion of the stromal $CO_2$ into $HCO_3^-$ due to the alkaline pH in the chloroplast stroma. Their activity generates a $CO_2$ sink which in turn drives the passive diffusion of external $CO_2$ into the chloroplast. Whereas in very low $CO_2$ conditions, the HLA3/LCIA transporters co-operatively brings $HCO_3^-$ into the chloroplast stroma through an ATP-dependent processes (Wang et al. 2015; Mackinder 2018).

The accumulated $HCO_3^-$ incurred by both processes is shuttled into the thylakoid lumen through the three putative $HCO_3^-$ channels called BESTROPHIN-LIKE 1 – 3 (BST1, BST2 and BST3 (Mukherjee et al. 2019). The BSTs proteins belong to the Bestrophin protein family that primarily function as chloride channels but have also been shown to exhibit $HCO_3^-$ permeability (Qu and Hartzell 2008; Herdean et al. 2016). The simultaneous RNAi knockdown line of *BST1/2/3* showed reduced $C_i$ affinity and caused growth impairment in $C_i$ limiting conditions. Rather than homogenously distributed on the thylakoid, the BST proteins are specifically enriched at the thylakoid membrane that immediately surrounds the pyrenoid. This proximity to the pyrenoid is hypothesised to accelerate the transport of $HCO_3^-$ directly to the $C_i$ sink – the pyrenoid (Mukherjee et al. 2019).

Since Rubisco at the pyrenoid is unable to access the membrane impermeable $HCO_3^-$ accumulated in the thylakoid lumen, a carbonic anhydrase is required to catalyse the dehydration of $HCO_3^-$ into $CO_2$ for their release. The CAH3 mutant *cah3* (or *ca-1*/*cia3*) cells exhibit reduced growth in $C_i$ limiting conditions despite its $C_i$ accumulation level above that of wildtype (Spalding et al. 1983; Moroney et al. 1986; Duanmu et al. 2009b). This high $C_i$ accumulation suggests that $C_i$-uptake is not impaired in *cah3,* instead the *cah3* mutant is unable to mobilise the accumulated $C_i$ (Duanmu et al. 2009b). Combined with its thylakoid luminal localisation (Karlsson 1998), CAH3 is proposed to be the final carbonic anhydrase that is responsible for releasing $CO_2$ into the pyrenoid. Contrary to LCIB, the conversion of $HCO_3^-$ to $CO_2$ is supported by the acidic pH environment stemming from photosynthesis. Together, the described CCM components and their functions highlight the two separate but complementary processes of $C_i$ uptake, which operate under different external $C_i$ concentrations.

**The transcriptional regulation of the CCM**

The Chlamydomonas CCM is inducible by low external $CO_2$ concentration. The induced expression of many identified CCM components including HLA3, LCI1, LCIA, LCIB, LCR1, CAH1, and CAH3 etc. are dependent on a Zinc-finger type transcription regulator called CIA5 (Fang et al. 2012). This vast association of CIA5 to CCM genes leads to its recognition as the master regulator of the Chlamydomonas CCM. The downstream signalling pathway of CIA5 is partially elucidated with the identification of LCR1. LCR1 is a myb-like transcription factor that is up-regulated by low $CO_2$ concentration in a CIA5-dependent manner. It is found that some CCM components such as CAH1 and LCI1 are no longer upregulated when grown in a low $CO_2$ concentration in the *lcr1* mutant strains, suggesting LCR1 is critical for their $CO_2$ response (Yoshioka et al. 2004). However, *lcr1* mutant does not abolish the low $CO_2$ response in all genes regulated by CIA5 suggesting that additional transcriptional regulators downstream of CIA5 are still to be identified. Moreover, multiple gene clusters were identified to dynamically respond to $CO_2$ limitation even in the *cia5* mutant which highlights the Chlamydomonas transcriptomic response to limiting $CO_2$ is not fully dependent on *CIA5* expression. These clusters includes genes enriched in metabolic pathways and extends to genes such as the *EPYC1* (*LCI5*) (Fang et al. 2012). Together these suggest that CIA5 is critical for the $CO_2$ response in many CCM genes but is not the sole regulator of the Chlamydomonas CCM.

Indeed, a calcium-binding protein in the chloroplast called CAS1 has been implicated in the parallel regulation of CCM components HLA3 and LCIA (Wang et al. 2016). The CAS1 protein is widely conserved in the green lineage. Its Arabidopsis homolog is known to be responsible for the transient elevation of cytosolic $Ca^{2+}$ signal that is important for stomatal closure (Han et al. 2003; Wang et al. 2016). The authors found that CAS1 displayed dynamic localisation in response to the $CO_2$ condition. They show that CAS1 is present on the thylakoid membrane external to the pyrenoid at high $CO_2$ concentration, and upon exposure to low $CO_2$ concentration, CAS1 transitioned into the thylakoid membrane within the pyrenoid. The dynamic localisation coincides with a concurrent increase of $Ca^{2+}$ concentration detected within the pyrenoid, further linking CAS1 and $Ca^{2+}$ signalling. Further, HLA3 and LCIA inducible expression in low $CO_2$ is also shown to be dependent on calcium concentration. This induction is lost in the *CAS1* insertional mutant, as both protein level of HLA3 and LCIA fails to accumulate in response to low $CO_2$ concentration. The exogenous addition of calcium alone is unable to rescue the *cas1* mutant's high-$CO_2$ requiring growth phenotype. Together, these results suggest that the CAS1-mediated calcium signal is responsible for sustaining a high HLA3/LCIA protein level in response to low $CO_2$.

**The energy budget of the CCM**

In the presence of photosynthetic light, photosystem 1 and 2 complexes on the thylakoid membrane drive the generation of a pH gradient across the thylakoid membrane through the combinatory processes called Linear electron flow (LEF), cyclic electron flow (CEF), and pseudocyclic electron flow (PCEF) (Burlacot 2023). The combined proton motive force (PMF) generated is then consumed by ATP synthases to produce ATP, while the final electron donor, ferredoxin, reduces NADP to generate NADPH. CEF and PCEF, additionally drives the increase in PMF without the reduction of NADP, and hence raises the ATP/NADPH ratio (Shikanai and Yamamoto 2017). In the perspective of the CCM, the multiple steps of the $C_i$ transport across cell membranes that delivers external $C_i$ into the thylakoid lumen involves active $C_i$ transporters such as HLA3 and LCIA, especially under very low $CO_2$ condition, with HLA3 also purported to be an energy-dependent (ATP) transporter (Burlacot and Peltier 2023). The release of accumulated $C_i$ in the form of $CO_2$ in the proximity of the pyrenoid matrix by CAH3 additionally consumes the pH gradient generated at the thylakoid lumen. Further downstream to the CCM machinery, the Calvin cycle and starch biosynthesis pathways necessitate the delicate balancing of the ATP/NADPH ratios, both substrates being the key outputs of photosynthesis (Burlacot and Peltier 2023). It's therefore evident that the energising mechanism – photosynthesis will be tightly coupled to the algal CCM machinery (Burlacot and Peltier 2023). Indeed, a high CEF/LEF ratios was measured from cells grown under $C_i$-limiting condition. Such high CEF/LEF ratios was shown to rapidly decrease in a state-transition independent manner upon the addition of external $C_i$ (Lucker and Kramer 2013). This suggests that Chlamydomonas cells exhibit rapid $C_i$-responsive fine-tuning of CEF/LEF ratios, likely as a response to compensate the increased ATP demand stemming from the CCM machinery (ATP-dependent $C_i$ transport processes) (Lucker and Kramer 2013). Further, a recent report proposed that the relationship between photosynthesis and the CCM might extend beyond the supplying of ATP (Burlacot et al. 2022). The authors found that the double mutants which lack both CEF and PCEF due to the mutations in PROTON GRADIENT REGULATOR (PGR1) and FLAVODIIRON PROTEIN B (FlvB) exhibit a CCM-dependent growth phenotype that is not found in the respective single mutant therefore suggesting that CEF/PCEF operates a compensatory process necessary for the CCM machinery. The authors further elucidated PGR1 and FlvB function by examining the mutant's non-photochemical quenching (NPQ) characteristics. NPQ is a mechanism where excess absorbed light energy is dissipated as heat (Roach 2020). The rapid and inducible component of NPQ, called energy-dependent quenching (qE), relies on the activity of the LHCSR3 protein that is responsive to thylakoid lumen acidification in Chlamydomonas (Bonente et al. 2011). Using NPQ as a proxy for trans-thylakoid pH gradient, the authors determined that Chlamydomonas cells lacking both CEF and PCEF fail to maintain a low thylakoid lumen pH.

It was then proposed that the CEF/PCEF ability to accumulate a low luminal pH possibly aids the BST-related function, by providing a favourable PMF for $HCO_3^-$ transport across the thylakoid membrane. However, the precise contributions of the proton gradient generated by CEF/PCEF to $HCO_3^-$ transport still awaits further examination. Especially as the *pgr1/flvb* mutant retains Ci uptake ability, in contrast to the BST1 – 3 knockdown strain (*bst-1*) which is unable to consume external Ci, suggesting that the BSTs transport activity is not abolished even when thylakoid luminal pH were affected (Burlacot et al. 2022).

## Structural components of the Chlamydomonas pyrenoid

### The starch sheath surrounds the pyrenoid matrix to limit $CO_2$ leakage

Unlike the cyanobacterial carboxysome, the Chlamydomonas pyrenoid does not exhibit a shell-like protein component which encases the Rubisco matrix to restrict the $CO_2$ leakage inherent to the rapid $CO_2$ release by the carbonic anhydrase. Instead, the Chlamydomonas pyrenoid is encircled by an arrangement of $CO_2$-impermeable starch granules, called the starch sheath. The $CO_2$-leakage reduction of starch sheath is partly associated with the recruitment of LCIB. During very low $CO_2$ conditions, LCIB is shown to re-localise and surround the pyrenoid starch sheath (Yamano et al. 2022). This placement enables LCIB to efficiently capture the leaked $CO_2$ and convert it back into $HCO_3^-$. The converted $HCO_3^-$ is then shuttled back into the thylakoid lumen creating a $CO_2$-recycling pathway. The dynamic re-localisation of LCIB has been shown to be dependent on correct starch sheath formation, as LCIB either aggregates at the chloroplast basal region in strains lacking a starch sheath (Toyokawa et al. 2020); or fail to show re-localisation in strains exhibiting multiple pyrenoid starch sheaths (Shimamura et al. 2023). Regardless, the association of LCIB with starch sheath suggests the pyrenoid starch sheath is a critical component of the CCM. Here I review the current understanding of the pyrenoid starch synthesis.

The process in which photosynthetically-derived metabolites are synthesised and stored in its osmotically inert form, starch, is hypothesised to have originated in Archaeplastida (Deschamps et al. 2008). Starch synthesis occurs in the chloroplast stroma. It initiates with an intermediate substrate of the Calvin cycle, fructose-6-phosphate (F6P). F6P is mutated into the Glucose-1-phosphate (G1P) with the phosphoglucoisomerase (PGI) (Preiser et al. 2020). Phosphoglucomutase (PGM) then catalyse the conversion between the G1P and G6P by changing the phosphate group from position α-1 to α-6. Subsequently, ADP-glucose pyrophosphorylase (AGPase) consumes an ATP and drives the addition of an adenosine monophosphate group at the same position, converting the G6P into ADP-glucose (Tetlow and Bertoft 2020). ADP-glucose production is considered a commitment step of starch synthesis as it represents a rate-limiting reaction in the starch biosynthesis pathway and is

strictly regulated allosterically by its substrates such as AMP, F6P, and redox modulation (Tiessen et al. 2002; Comino et al. 2017). After the generation of ADP-glucose, two paths of starch synthesis emerge which converts ADP-glucose into either amylopectin or amylose. Despite their identical building blocks (ADP-glucose), the two starch polymers exhibit drastically different properties which stem from the differential arrangement of the glucan chain; and the interplay of composition between the two polymers results in starch granules with a wide diversity of properties.

Amylopectin represents the major fraction within the starch granules (Tetlow and Bertoft 2020). It is formed via the addition of ADP-glucose on the α-1,4-glucosidic linkage through the activity of soluble starch synthases. Starch branching enzymes (SBE) then join the synthesised chain at the α-1,6-position resulting in linear branches which run at an angle along the same trajectory in a tree branch-like fashion. As neighbouring branches can form double helices, branches of different polymer are able to “hook” onto each other resulting in the semi-crystalline packing of the amylopectin that gives rise to starch granule’s insoluble properties. Additionally, starch debranching enzymes (SDE) aided in the trimming of α-1,6-bonds to modulate branch points and promote the semi-crystalline packing (Tetlow and Bertoft 2020). The combined activity of SBE and SDE brings the proportion of α-1,6-linkage to ~5% (Tetlow and Bertoft 2020). Amylose is similarly composed of ADP-glucose joined on the α-1,4-glucosidic linkage resulting in a linear polymer, however it exhibits little to no branching content (Seung 2020). Also unique to amylose, its synthesis is driven by the granule-bound starch synthase (GBSS), which suggests amylose production is unable to proceed in the absence of pre-existing starch granules (Smith et al. 2004; Grimaud et al. 2008; Seung et al. 2015). In starch granules, amylose is interspersed with amylopectin and is particularly enriched at the region of lower crystalline structure where amylopectin branching occurs. The organisation of the two polymers within the starch granules, results in the diverse properties of the starch granule. This is possible due to the array of starch synthesis enzymes where substrate affinity differs creating diverse branch length, crystalline packaging, and amylopectin/amylose ratios.

The starch granules that constitute the Chlamydomonas pyrenoid starch sheath are shaped as elongated cup-shaped granules which deviates from the lenticular morphology of “standard” starch granules that are present in the chloroplast stroma in *Arabidopsis* and Chlamydomonas. As with other components of the CCM, the starch sheath synthesis is similarly responsive to the external $CO_2$ conditions (Ramazanov et al. 1994). In elevated $CO_2$ conditions, most of the cell’s starch granules are present in the stroma having an oval shape, and a minimal starch sheath is seen. Cells exposed to ambient $CO_2$ conditions exhibit significant starch morphology change where a much thicker starch sheath and

reduction in stromal localised starch granules are observed. Despite their morphological differences, differentiating the starch granules between those originating from pyrenoid starch sheath; or the stromal localised starch granules after cell lysis is relatively difficult. Nitrogen starvation treatment is commonly applied to Chlamydomonas cells to induce lipid and starch production, this increase in starch production is seen to occur with the concomitant reduction in pyrenoid starch sheath thickness. Thus, the difference between starch composition and characteristics are often performed between the so called “transitory” and “storage starch” which refers to pyrenoid starch which are formed under low $CO_2$ conditions, and those formed under nitrogen deprivation respectively.

From these analyses, pyrenoid starch exhibits two main deviations from the stromal starch granules. First, pyrenoid starch exhibit much lower amylose content (~7%) compared to the high amylose content in stromal (storge) starch (~35%; (Courseaux et al. 2023)). Secondly, pyrenoid starch has an A-type crystal pattern and exhibits a higher crystallinity compared to stromal starch (Izumo et al. 2011). These two properties are likely interconnected, as increase in amylose content in maize starch granule has been seen with the concurrent reduction in its crystallinity, a consequence that has been attributed to the amorphous property of amylose (Cheetham and Tao 1998). In maize, a mutant that lacks SBEIIb produces starch granules where its amylopectin displays drastic reduction in branch points, and much longer chain lengths granting its name “amylose-extender” (Nishi et al. 2001). A high proportion of the amylose-extender starch granules are elongated and are generally smaller than wildtype starch granules due to the altered starch polymer structures (Lin et al. 2016). This contrasts with the pyrenoid starch which is similarly elongated despite displaying a low amylose content. It might be reasoned that the stromal starch produced from nitrogen starved cells differs from stromal starch granules that are common to high $CO_2$ culture conditions, and that the “low” amylose content detected of pyrenoid starch is in fact higher than that of the canonical stromal starch harvested from high $CO_2$ grown cells. In support of this, high $CO_2$ grown cells display stromal starch granules that are of comparable size to the pyrenoid starch sheath (Ramazanov et al. 1994), while nitrogen-starved cells stromal starch granules are often drastically larger and display irregular shape (Findinier et al. 2019b).

Alternatively, the direct relationship between starch granules and its amylose content might not hold true in Chlamydomonas. Mutation in the Chlamydomonas GBSSI, more commonly known as *STA2*, is expected to reduce amylose content within pyrenoid starch. However, it was determined that both pyrenoid and stromal starch purified from the *gbssI* mutant displayed lower crystallinity in comparison to wildtype background suggesting amylose/amylopectin ratio is likely regulated by additional interplay of starch synthases (Izumo et al. 2011). Regardless, despite the apparent difference in starch polymer

characteristics, the key morphological aspects of pyrenoid starch, Its cup-shaped granule, is still present. This is similarly observed in a separate STA2 mutant line where TEM micrographs clearly shows that the *sta2* pyrenoid starch sheath is thinned but does not compromise its overall coverage of the pyrenoid through their cup-shaped granule (Toyokawa et al. 2020). In a recent report, Chlamydomonas mutants of the starch branching enzymes SBE1, SBE2, and SBE3 have also been obtained which parallel the amylose-extender mutation in maize. The authors found that mutation in SBE3 yields the biggest impact on starch properties where the apparent amylose content in pyrenoid starch was increases by 2.4-fold when compared to the wildtype background and a reduction to total starch content by 20% (Courseaux et al. 2023). However, no imaging data has been provided which will allow us to determine whether pyrenoid starch sheath morphology is significantly changed due to its mutation. Therefore, the starch polymer characteristics likely contribute to certain aspects of the pyrenoid starch morphology such as thickness, however, the extent of starch composition's contribution to their morphology is still to be elucidated.

As the amylose content in both pyrenoid and stromal starch are increased in the SBE3 mutant line, it can be reasoned that an identical suite of starch synthesis enzymes is responsible for both pyrenoid and stromal starch synthesis. Indeed, while some starch synthesis protein has been shown to be induced by low $CO_2$ conditions such as STA2 and STA3 (Miura et al. 2004) suggesting they are likely important for the low $CO_2$ induced pyrenoid starch, most starch-related proteins showed little $CO_2$ response (Brueggeman et al. 2012; Fang et al. 2012). This raises an interesting question, as to what other proteins are responsible for the differences between stromal and pyrenoid starch.

To date, two proteins named STARCH GRANULE ABNORMAL 1 and 2 (SAGA1 and SAGA2) have been shown to be required for normal pyrenoid starch sheath morphology. Both proteins contain RBMs; a starch binding motif CBM20; and a significant proportion of the protein sequence are predicted as alpha helices that are homologous to coiled-coil domains (Itakura et al. 2019). Fluorescent protein tagging shows that SAGA1 localised to the pyrenoid with partial overlap to the starch sheath edge. The deletion of *SAGA1* caused the appearance of multiple pyrenoid, and the production of elongated and thinned starch granules with reduced contact with the pyrenoid matrix. The aberrant starch sheaths phenotype combined with their unique localisation leads to the hypothesis that they might directly participate in starch sheath synthesis, or alternatively regulate their formation by promoting the association of the Rubisco matrix with the pyrenoid starch sheath. Moreover, a recent study has expressed SAGA2, a homolog of SAGA1 that shares 30% sequence identity, in Arabidopsis that contains a proto-pyrenoid (Atkinson et al. 2024). The authors showed that the concomitant expression of the two SAGA proteins increased the proportion

of starch granules adjacent to the proto-pyrenoid matrix, further supporting SAGA's ability to promote matrix-starch interaction. However, much of the cells starch granules are still present in the stromal fraction, and the adjacent starch does not fully encircle the proto-pyrenoid matrix. Furthermore, the morphology of these adjacent starch granules still deviates from the cup-shaped canonical pyrenoid starch sheath suggesting that additional components of starch sheath synthesis are at play. One potential candidate is the protein named LOW $CO_2$ INDUCED 9 (LCI9; (Mackinder et al. 2017)). The fluorescent protein tagging of LCI9 shows that the protein localises on the edge in between the pyrenoid starch sheaths in a mesh-like manner. LCI9 also contains the starch-binding CBM20 domain and interacts with starch-related proteins such as SBE1 and PFK1 (Mackinder et al. 2017). Together these data highlights LCI9 as a promising candidate which might be important for shaping the pyrenoid starch sheath edge.

Additionally, in Arabidopsis, a class of coiled-coil proteins containing a carbohydrate-binding domain 48 (CBM48) called PROTEIN TARGETING TO STARCH (PTSTs) have been shown to be important for starch granule formation. They function by either directing the GBSS towards initiated starch granules for amylose synthesis (PTST1) or bringing short glucan chains towards starch synthase 4 that are critical for granule initiation (PTST2 and PTST3). This ability to promote the glucosyl substrates binding to the starch synthases drastically alters the amount and properties of starch granules, as seen in their respective mutants (Seung et al. 2015, 2017). Chlamydomonas also possesses one homolog of the PTST protein – Cre09.g387100. While no characterisation of their function has been reported to date, it is highly likely that Cre09.g387100 will also be crucial to produce starch granules, and in extension, the pyrenoid starch sheath.

In summary, the pyrenoid starch sheath is a critical component of the Chlamydomonas CCM, where its ability to limit $CO_2$ leakage is mediated by both the physical properties of starch sheath as well as its interaction with LCIB. The synthesis of this component is shown to be partly influenced by the differential expression of starch synthesis proteins, however, additional components such as SAGA1, SAGA2, LCI9 and PTST likely play an important role by either directly regulating starch synthesis or by indirectly regulating starch synthesis enzymes. Further work in characterising their function will be critical in elucidating the pyrenoid starch synthesis mechanism.

**The pyrenoid tubule network traversing the pyrenoid is integral to $C_i$ delivery**

The presence of thylakoid membranes traversing the pyrenoid is a common but not universal phenomenon observed in many pyrenoid-containing algal species (Barrett et al. 2021). The general morphology of the pyrenoid-traversing thylakoids also presents an impressive

diversity in their complexity (Meyer et al. 2017). This is especially apparent in *Porphyridium cruentum* where a complex disorganised network of thylakoid membranes is seen traversing the pyrenoid matrix (Gantt and Conti 1965). While on the opposite spectrum, in species such as *Chlorella vulgaris*, a single thylakoid bisects the pyrenoid, and at the resolution of TEM the pyrenoid traversing thylakoid membrane presents no visible difference to the surrounding (stromal) thylakoid (Zhang et al. 2020). In Chlamydomonas, (Engel et al. 2015) explored the chloroplast native architecture using cryo-electron tomography where the chloroplast structure is frozen in amorphous ice and thus preserving the native structure of the thylakoid membranes. They discovered that the thylakoid membrane in Chlamydomonas is organised as interconnected sheets which run parallel to the chloroplast envelope. Once these sheets are in proximity to the pyrenoid, they fuse and contort into tubular structures that are termed pyrenoid tubules. Nanoscale membrane structures called pyrenoid minitubules were also discovered within these pyrenoid tubules. These minitubules are formed via the infolding of thylakoid membranes, creating channels which bridges the external stromal space and pyrenoid matrix. It is hypothesised that the minitubule's <4 nm lumenal space likely permits metabolite transport. This is particularly important as Calvin cycle enzymes such as RPE1, PPDK, PRK1 are found external to the pyrenoid suggesting that a direct and efficient metabolite transport between the pyrenoid matrix and stroma is crucial for maintaining Calvin cycle efficiency (Küken et al. 2018). Apart from the morphology, it is likely that the membrane composition of the pyrenoid tubules is also different. Indeed, the pyrenoid tubule network is seen to become disorganised and swollen when Chlamydomonas cells are subjected to heat shock treatment, while the surrounding thylakoid membrane integrity is maintained (Zhang et al. 2022). This different response to heat treatment supports that the membrane composition of pyrenoid tubules is drastically different from its surrounding thylakoid.

The pyrenoid tubule network is a critical part of the Chlamydomonas CCM partly due to their association with two proteins – CAH3 and CAS1. Their molecular function was already described in detail in a previous section. Besides their association for CO2 delivery to the pyrenoid matrix, pyrenoid tubules might additionally play a role in pyrenoid fission. Goodenough (1970) showed clear TEM evidence where thylakoid membranes were seen appressed to the pyrenoid matrix in cells undergoing division. Their observation was echoed by the detailed confocal imaging during the pyrenoid division process (Freeman Rosenzweig et al. 2017). Together, this suggests that a positional signal external to the pyrenoid matrix recruits the plastid division machinery and determines the canonical division site. The Chlamydomonas division machinery is hypothesised to be inherited from bacterial system where the MIN protein system are responsible for positioning the FtsZ ring that tethers to

chloroplast envelope and driving thylakoid membrane constriction (Adams et al. 2008; TerBush et al. 2013). Recently, the MinD protein has been localised to the pyrenoid periphery (Wang et al. 2023). While no experimental evidence has been shown to establish their relationship, the proximity of MinD and pyrenoid tubules could suggests that the pyrenoid tubules serves as a positional benchmark for the chloroplast and pyrenoid division.

Despite being an integral part of the Chlamydomonas CCM, the mechanism in which the pyrenoid tubule is formed is largely unclear. Two proteins, RUBISCO BINDING MEMBRNAE PROTEIN 1 (RBMP1) and RUBISCO BINDING MEMBRNAE PROTEIN 2 (RBMP2), are localised to the pyrenoid tubules and possess multiple RBMs. This led to the hypothesis that these proteins might function as a pyrenoid tether which recruits the Rubisco matrix towards the tubule network (Meyer et al. 2020). However, whether (1) RBMP1 and RBMP2 recruits Rubisco towards the pyrenoid tubules; (2) Rubisco/pyrenoid recruitment on thylakoid membrane leads to tubule synthesis; or (3) if RBMP1 and RBMP2 are directly involved in tubule synthesis is largely unresolved. RBMP1 was recently identified as a homolog of the Bestrophin protein family and was subsequently renamed as BESTROPHIN-LIKE PROTEIN 4 (BST4). The *bst4* mutants exhibits an intact pyrenoid matrix coupled to the pyrenoid tubule and the pyrenoid tubule itself does not exhibit any visible structural defect in TEM micrographs (Adler et al. 2023). Therefore at least in the case of RBMP1/BST4, this protein alone does not recruit Rubisco towards pyrenoid tubules and is not directly involved in shaping the unique morphology in pyrenoid tubules. However, It is reported that the interaction between phase-separated components and lipid membrane alone is able to drive membrane tubulation due to the compressive stress applied (Yuan et al. 2021). It therefore remains a possibility that, the collective interaction between the Rubisco-matrix and all Rubisco-binding protein on the pyrenoid tubule can lead to the modified pyrenoid tubule morphology.

Due to the high-level of curvature displayed by the pyrenoid tubules, the thylakoid curvature proteins, CURTs, have also been proposed as likely candidates for modulating the unique network morphology of the pyrenoid tubules. In Arabidopsis, the CURT1A protein has been characterised to be important for the maintenance of thylakoid grana structure by their insertion to the grana edge (Armbruster et al. 2013). Further, observation of etiolating Arabidopsis cotyledons found that the CURT proteins are enriched at the transitional region between crystalline prolamellar bodies, a para-crystalline structure of thylakoid membrane precursor, which supports their function in directly modulating thylakoid membrane architecture (Sandoval-Ibáñez et al. 2021; Liang et al. 2022). However, the Chlamydomonas CURT1 (Cre05.g233950) localisation is not shown to be preferentially enriched in the pyrenoid tubules (Wang et al. 2023), suggesting they are unlikely to be responsible for the

tubule network seen in the Chlamydomonas. In short, while multiple hypotheses have been suggested, the biogenesis mechanism of pyrenoid tubules remains largely unresolved. Though it is possible that multiple of these hypotheses work concertedly to yield the unique morphology of the pyrenoid tubules. Regardless, it is clear that more work is needed to elucidate the factors involved in generating this unique membrane structure that is integral to the Chlamydomonas CCM.

### A Rubisco guided view of pyrenoid assembly

When initially characterising the Rubisco-interactor SAGA1, a co-immunoprecipitation experiment was carried out using an antibody raised against the SAGA1 protein to identify its interactors(Meyer et al. 2020). Serendipitously, the immunised peptide was designed as the Rubisco binding domain on EPYC1, which leads to the co-immunoprecipitation of Rubisco as well as many other proteins that similarly carry the RBM such as SAGA2, RBMP1 and RBMP2. It was then tested that the RBMs, through their ability to bind to Rubisco, can direct previously non-pyrenoid proteins to be partitioned into pyrenoid (Meyer et al. 2020). Due to the crucial function of SAGA1 and SAGA2 on reshaping the starch sheath surrounding the pyrenoid, and EPYC1's ability to phase separate Rubisco holoenzymes, a hypothesis was proposed that RBMs drive the assembly of the pyrenoid organelle as a whole. It posits that the presence of RBM-containing proteins on the Rubisco matrix, starch sheath, and pyrenoid tubule links the three compartments together: where EPYC1's RBM driving phase separation of Rubisco to form the Rubisco matrix; then Rubisco matrix-starch association is mediated by the SAGA1 and SAGA2's RBM; followed by the RBMP1 and RBMP2 proteins which anchors the Rubisco matrix on the pyrenoid tubule (Meyer et al. 2020). This model addressed the unexpected observation of a small but sizable Rubisco enrichment at the canonical pyrenoid position in EPYC1-deficient strains (Mackinder et al. 2016). As one could reason that the pyrenoid tubules retains their ability to recruit a limited amount of Rubisco via the RBMs on RBMP proteins. Conversely, when the pyrenoid is removed by the replacement of the native RBCS by that of vascular plant *Spinacia oleracea*, the previously observed Rubisco enrichment is now absent as RBM interaction between pyrenoid tubule, starch sheath and Rubisco holoenzymes are removed altogether by the RBCS replacement (Caspari et al. 2017).

An unresolved question, however, is the sequence of events during pyrenoid assembly. Whether pyrenoid tubules, starch sheath and Rubisco matrix formation occur in separate instances or are co-dependent on their development. Some clues can be drawn based on the pyrenoid-less strain described earlier. First, a pyrenoid tubule network is still present in the pyrenoid-less strains where RBCS is replaced, suggesting that the pyrenoid tubule formation is likely independent from Rubisco matrix formation. On the contrary, starch

granules in the same strains exhibits a morphology that is more akin to the stromal starch granule, therefore matrix formation likely occurs prior to starch granule remodelling. Therefore, it can be hypothesised that during pyrenoid formation, the pyrenoid tubule is formed initially which provides a central point for Rubisco recruitment and their eventual phase separation by EPYC1. This is followed by the recruitment of starch granule towards the Rubisco matrix via the interactions between SAGA1/2 and Rubisco. Here, their proximity to the matrix leads to their remodelling into the pyrenoid starch sheath, resulting in a fully formed pyrenoid.

### Projects aims

This PhD project aims to understand the biogenesis process of the pyrenoid, particularly the two understudied components of the pyrenoid organelle: starch sheath and pyrenoid tubules. Since much of the protein composition on these compartments are largely unknown, I first set off to characterise their composition using proteomic techniques. Previously, the pyrenoid composition has been profiled using the biochemical purification of the pyrenoid which generated a highly informative dataset that contains important proteins in pyrenoid formation such as EPYC1 and SAGA1 etc. However, this dataset contains a large volume of noise especially with the extensive photosystem components identified within the reported pyrenoid proteome (Zhan et al. 2018). Proximity labelling is a relatively novel method that makes use of either a biotin ligase (TurboID (Branon et al. 2018)) or Ascorbate peroxidase (APEX2 (Hwang and Espenshade 2016)) enzyme to generate a labile biotin radical that can biotinylate neighbouring proximal proteins. The resultant spatially defined biotin-labelled proteins can then be purified via streptavidin-pulldown and subsequently analysed through mass spectrometry analysis. There are two advantages of proximity labelling in comparison to biochemical purification of organelles in determining proteomic composition. Firstly, the generation of biotin radicals is strictly dependent on exogenously added biotin substrates. If designed appropriately, this enables the detection of temporal and transient interaction as opposed to a stable interaction that affinity-purification coupled to mass spectrometry (AP-MS) often captures. Secondly, stringent wash steps can be applied during the purification of biotinylated proteins due to the strong biotin-streptavidin interaction aiding the removal of contaminants. Additionally, I reasoned that despite most of the pyrenoid formation factors identified are found to interact with Rubisco, this interaction is likely not a strict requirement. Indeed, as proposed in the earlier section, pyrenoid tubule formation likely occurs independently of Rubisco matrix formation, suggesting that the potential pyrenoid tubule forming proteins do not necessarily interact with Rubisco. Therefore, in **Chapter 2**, a proximity labelling pipeline was developed and was first tested on the pyrenoid matrix. Here I tagged the RBCS2 and EPYC1 protein, the main constituent of the pyrenoid matrix, and two

stromal localised protein RPE1 and PRK1 as a compartment control. By contrasting their biotinylated protein profile, I obtained a high confidence pyrenoid proxiome (**Proxi**mity prote**ome**) where over half of the detected proteins were previously identified pyrenoid proteins. The subsequent validation of this list identified 6 new proteins that are localised within the pyrenoid with a diverse sub-pyrenoid pattern. This work establishes proximity labelling as a valuable technique in Chlamydomonas to investigate protein-protein interaction and highlighted several pyrenoid proteins that warrant further characterisation. Chapter 2 is published at The Plant Cell journal as "A phase-separated $CO_2$-fixing pyrenoid proteome determined by TurboID in *Chlamydomonas reinhardti*" Lau et al., 2023 (doi: https://doi.org/10.1093/plcell/koad131)

In **Chapter 3**, I further push towards identifying novel pyrenoid-assembly critical proteins. To do so, I expressed the TurboID tag targeted towards the understudied sites of starch sheath and thylakoid membrane. Aided by previous work on characterising pyrenoid composition (Mackinder et al. 2017; Meyer et al. 2020), a suite of proteins with defined localisation can already be selected to target these positions. I selected the starch sheath edge localised LCI9 due to its proposed role in regulating starch sheath morphology (Mackinder et al. 2017); the pyrenoid tubule localised FZL and RBMP2, the latter was also proposed to be important for pyrenoid matrix-tubule association (Findinier et al. 2019a; Meyer et al. 2020); and a homogenously thylakoid protein PSAF as a compartment control (Emrich-Mills et al. 2021). Frustrated by the previous cloning method which necessitate gene domestication whenever a target gene contains a Type IIS enzyme recognition site, I modified the plasmid originally required for recombineering cloning (pLM099; (Emrich-Mills et al. 2021)) by replacing the fluorescent protein tag with the TurboID tag. This modification allows us to take advantage of the previously developed recombineering cloning processes which allows the cloning of large and complex genes within Chlamydomonas (Emrich-Mills et al. 2021). The inclusion of the native promoter in this cloning pipeline also potentially avoids disturbing the native cellular functions due to over-expression of tagged proteins. Using a similar workflow as before, I filtered the enrichment data and found particularly interesting insights on the starch-related proteins LCI9. I discovered that LCI9 retains its previously reported interaction with SBE3 but are not found to be associated with the PFK proteins which suggest the PFK-LCI9 interaction might be highly transient in nature (Mackinder et al. 2017). Additionally, the identified proxiome of LCI9 is populated with both starch synthesis and degradation enzymes, its regulation on starch morphology is therefore likely more complex than simply directing these enzymes towards starch granules. Furthermore, I found that LCI9 is proximal with an uncharacterised protein named **P**yrenoid **IN**terface **S**tarch 2 – PINS2. LCI9 is a

homolog of PINS2 and shares a similar domain arrangement, this led to our targeted characterisation of PINS2 and LCI9 (now renamed PINS1) in the following chapter.

In the final **Chapter 4**, I focused on the characterisation of a starch-related protein PINS2 identified in **Chapter 2 and 3**. PINS2 domain arrangement is homologous to the PINS1 protein with both proteins containing a CBM20 domain alongside a coiled-coil domain. I additionally identified another homolog, PINS3, and show that they form a three gene-cluster on the chromosome. I employed CRISPR-Cas9 mediated editing to generate multiple independent mutants of PINS2 and characterise their phenotype through confocal imaging, transmission electron microscopy and growth assay. Collectively, I found that the *pins2* mutant exhibits an aberrant starch sheath formation and fails to recruit LCIB towards the pyrenoid periphery during low $CO_2$ condition, further affirming the starch-related phenotype. I subsequently propose that PINS proteins maintain the starch sheath morphology by inhibiting starch synthesis at the edge of pyrenoid starch plates.

**References**


**Adams S, Maple J, and Møller SG**. Functional conservation of the MIN plastid division homologues of Chlamydomonas reinhardtii. Planta. 2008:**227**(6):1199–1211.

**Adler L, Lau CS, Shaikh KM, van Maldegem KA, Payne-Dwyer AL, Lefoulon C, Girr P, Atkinson N, Barrett J, Emrich-Mills TZ, et al.** The role of BST4 in the pyrenoid of*Chlamydomonas reinhardtii*. bioRxiv. 2023. https://doi.org/10.1101/2023.06.15.545204

**Armbruster U, Labs M, Pribil M, Viola S, Xu W, Scharfenberg M, Hertle AP, Rojahn U, Jensen PE, Rappaport F, et al.** Arabidopsis CURVATURE THYLAKOID1 proteins modify thylakoid architecture by inducing membrane curvature. Plant Cell. 2013:**25**(7):2661–2678.

**Atkinson N, Feike D, Mackinder LCM, Meyer MT, Griffiths H, Jonikas MC, Smith AM, and McCormick AJ**. Introducing an algal carbon-concentrating mechanism into higher plants: location and incorporation of key components. Plant Biotechnol J. 2016:**14**(5):1302–1315.

**Atkinson N, Stringer R, Mitchell SR, Seung D, and McCormick AJ**. SAGA1 and SAGA2 promote starch formation around proto-pyrenoids in Arabidopsis chloroplasts. Proc Natl Acad Sci U S A. 2024:**121**(4). https://doi.org/10.1073/pnas.2311013121

**Barrett J, Girr P, and Mackinder LCM**. Pyrenoids: CO2-fixing phase separated liquid organelles. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2021:**1868**(5):118949.

**Bonente G, Ballottari M, Truong TB, Morosinotto T, Ahn TK, Fleming GR, Niyogi KK, and Bassi R**. Analysis of LhcSR3, a protein essential for feedback DE-excitation in the Green alga Chlamydomonas reinhardtii. PLoS Biol. 2011:**9**(1):e1000577.

**Borkhsenious ON, Mason CB, and Moroney JV**. The intracellular localization of ribulose-1,5-bisphosphate Carboxylase/Oxygenase in chlamydomonas reinhardtii. Plant Physiol. 1998:**116**(4):1585–1591.

**Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, Feldman JL, Perrimon N, and Ting AY**. Efficient proximity labeling in living cells and organisms with TurboID. Nat Biotechnol. 2018:**36**(9):880–887.

**Brueggeman AJ, Gangadharaiah DS, Cserhati MF, Casero D, Weeks DP, and Ladunga I**. Activation of the carbon concentrating mechanism by CO2 deprivation coincides with massive transcriptional restructuring in Chlamydomonas reinhardtii. Plant Cell. 2012:**24**(5):1860–1875.

**Burlacot A**. Quantifying the roles of algal photosynthetic electron pathways: a milestone towards photosynthetic robustness. New Phytol. 2023:**240**(6):2197–2203.

**Burlacot A, Dao O, Auroy P, Cuiné S, Li-Beisson Y, and Peltier G**. Alternative photosynthesis pathways drive the algal CO2-concentrating mechanism. Nature. 2022:**605**(7909):366–371.

**Burlacot A and Peltier G**. Energy crosstalk between photosynthesis and the algal CO2-concentrating mechanisms. Trends Plant Sci. 2023:**28**(7):795–807.

**Burow MD, Chen Z-Y, Mouton TM, and Moroney JV**. Isolation of cDNA clones of genes induced upon transfer of Chlamydomonas reinhardtii cells to low CO2. Plant Mol Biol. 1996:**31**(2):443–448.

**Cameron JC, Wilson SC, Bernstein SL, and Kerfeld CA**. Biogenesis of a bacterial organelle: The carboxysome assembly pathway. Cell. 2013:**155**(5):1131–1140.

**Caspari OD, Meyer MT, Tolleter D, Wittkopp TM, Cunniffe NJ, Lawson T, Grossman AR, and Griffiths H**. Pyrenoid loss in Chlamydomonas reinhardtii causes limitations in CO2 supply, but not thylakoid operating efficiency. J Exp Bot. 2017:**68**(14):3903–3913.

**Cheetham NWH and Tao L**. Variation in crystalline type with amylose content in maize starch granules: an X-ray powder diffraction study. Carbohydr Polym. 1998:**36**(4):277–284.

**Chen AH, Robinson-Mosher A, Savage DF, Silver PA, and Polka JK**. The bacterial carbon-fixing organelle is formed by shell envelopment of preassembled cargo. PLoS One. 2013:**8**(9):e76127.

**Comino N, Cifuente JO, Marina A, Orrantia A, Eguskiza A, and Guerin ME**. Mechanistic insights into the allosteric regulation of bacterial ADP-glucose pyrophosphorylases. J Biol Chem. 2017:**292**(15):6255–6268.

**Courseaux A, George O, Deschamps P, Bompard C, Duchêne T, and Dauvillée D**. BE3 is the major branching enzyme isoform required for amylopectin synthesis in Chlamydomonas reinhardtii. Front Plant Sci. 2023:**14**. https://doi.org/10.3389/fpls.2023.1201386

**Danila FR, Thakur V, Chatterjee J, Bala S, Coe RA, Acebron K, Furbank RT, von Caemmerer S, and Quick WP**. Bundle sheath suberisation is required for C4 photosynthesis in a Setaria viridis mutant. Commun Biol. 2021:**4**(1). https://doi.org/10.1038/s42003-021-01772-4

**Deschamps P, Haferkamp I, d'Hulst C, Neuhaus HE, and Ball SG**. The relocation of starch metabolism to chloroplasts: when, why and how. Trends Plant Sci. 2008:**13**(11):574–582.

**Duanmu D, Miller AR, Horken KM, Weeks DP, and Spalding MH**. Knockdown of limiting-$CO_2$–induced gene *HLA3* decreases $HCO_3^-$ transport and photosynthetic Ci affinity in *Chlamydomonas reinhardtii*. Proc Natl Acad Sci U S A. 2009a:**106**(14):5990–5995.

**Duanmu D, Wang Y, and Spalding MH**. Thylakoid lumen carbonic anhydrase (*CAH3*) mutation suppresses air-dier phenotype of *LCIB* mutant in *Chlamydomonas reinhardtii*. Plant Physiol. 2009b:**149**(2):929–937.

**Eisenhut M, Ruth W, Haimovich M, Bauwe H, Kaplan A, and Hagemann M**. The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants. Proc Natl Acad Sci U S A. 2008:**105**(44):17199–17204.

**Emrich-Mills TZ, Yates G, Barrett J, Girr P, Grouneva I, Lau CS, Walker CE, Kwok TK, Davey JW, Johnson MP, et al.** A recombineering pipeline to clone large and complex genes in Chlamydomonas. Plant Cell. 2021:**33**(4):1161–1181.

**Engel BD, Schaffer M, Kuhn Cuellar L, Villa E, Plitzko JM, and Baumeister W**. Native architecture of the Chlamydomonas chloroplast revealed by in situ cryo-electron tomography. Elife. 2015:**4**. https://doi.org/10.7554/elife.04889

**Ermakova M, Danila FR, Furbank RT, and von Caemmerer S**. On the road to $C_4$ rice: advances and perspectives. Plant J. 2020:**101**(4):940–950.

**Falkowski PG and Raven JA**. Aquatic Photosynthesis 2nd ed. (Princeton University Press: Princeton, NJ).

**Fang W, Si Y, Douglass S, Casero D, Merchant SS, Pellegrini M, Ladunga I, Liu P, and Spalding MH**. Transcriptome-wide changes in Chlamydomonas reinhardtii gene expression regulated by carbon dioxide and the CO2-concentrating mechanism regulator CIA5/CCM1. Plant Cell. 2012:**24**(5):1876–1893.

**Faulkner M, Szabó I, Weetman SL, Sicard F, Huber RG, Bond PJ, Rosta E, and Liu L-N**. Molecular simulations unravel the molecular principles that mediate selective permeability of carboxysome shell protein. Sci Rep. 2020:**10**(1). https://doi.org/10.1038/s41598-020-74536-5

**Findinier J, Delevoye C, and Cohen MM**. The dynamin-like protein Fzl promotes thylakoid fusion and resistance to light stress in Chlamydomonas reinhardtii. PLoS Genet. 2019a:**15**(3):e1008047.

**Findinier J, Laurent S, Duchêne T, Roussel X, Lancelon-Pin C, Cuiné S, Putaux J-L, Li-Beisson Y, D'Hulst C, Wattebled F, et al.** Deletion of BSG1 in Chlamydomonas reinhardtii leads to abnormal starch granule size and morphology. Sci Rep. 2019b:**9**(1):1–13.

**Förster B, Rourke LM, Weerasooriya HN, Pabuayon ICM, Rolland V, Au EK, Bala S, Bajsa-Hirschel J, Kaines S, Kasili RW, et al.** The *Chlamydomonas reinhardtii* chloroplast envelope protein LCIA transports bicarbonate *in planta*. J Exp Bot. 2023:**74**(12):3651–3666.

**Freeman Rosenzweig ES, Xu B, Kuhn Cuellar L, Martinez-Sanchez A, Schaffer M, Strauss M, Cartwright HN, Ronceray P, Plitzko JM, Förster F, et al.** The Eukaryotic CO2-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization. Cell. 2017:**171**(1):148-162.e19.

**Gantt E and Conti SF**. THE ULTRASTRUCTURE OF *PORPHYRIDIUM CRUENTUM* . J Cell Biol. 1965:**26**(2):365–381.

**Gao H, Wang Y, Fei X, Wright DA, and Spalding MH**. Expression activation and functional analysis of HLA3, a putative inorganic carbon transporter in *Chlamydomonas reinhardtii*. Plant J. 2015:**82**(1):1–11.

**Gee CW and Niyogi KK**. The carbonic anhydrase CAH1 is an essential component of the carbon-concentrating mechanism in *Nannochloropsis oceanica*. Proc Natl Acad Sci U S A. 2017:**114**(17):4537–4542.

**Goodenough UW**. CHLOROPLAST DIVISION AND PYRENOID FORMATION IN *CHLAMYDOMONAS REINHARDI*[1]. J Phycol. 1970:**6**(1):1–6.

**Griffiths DJ**. The pyrenoid and its role in algal metabolism. Sci Prog. 1980:**66**(264):537–553.

**Grimaud F, Rogniaux H, James MG, Myers AM, and Planchot V**. Proteome and phosphoproteome analysis of starch granule-associated proteins from normal maize and mutants affected in starch biosynthesis. J Exp Bot. 2008:**59**(12):3395–3406.

**Hagemann M, Kern R, Maurino VG, Hanson DT, Weber APM, Sage RF, and Bauwe H**. Evolution of photorespiration from cyanobacteria to land plants, considering protein phylogenies and acquisition of carbon concentrating mechanisms. J Exp Bot. 2016:**67**(10):2963–2976.

**Han S, Tang R, Anderson LK, Woerner TE, and Pei Z-M**. A cell surface receptor mediates extracellular Ca2+ sensing in guard cells. Nature. 2003:**425**(6954):196–200.

**He S, Chou H-T, Matthies D, Wunder T, Meyer MT, Atkinson N, Martinez-Sanchez A, Jeffrey PD, Port SA, Patena W, et al.** The structural basis of Rubisco phase separation in the pyrenoid. Nat Plants. 2020:**6**(12):1480–1490.

**Herdean A, Teardo E, Nilsson AK, Pfeil BE, Johansson ON, Ünnep R, Nagy G, Zsiros O, Dana S, Solymosi K, et al.** A voltage-dependent chloride channel fine-tunes photosynthesis in plants. Nat Commun. 2016:**7**(1). https://doi.org/10.1038/ncomms11654

**Hwang J and Espenshade PJ**. Proximity-dependent biotin labelling in yeast using the engineered ascorbate peroxidase APEX2. Biochem J. 2016:**473**(16):2463–2469.

**Im CS and Grossman AR**. Identification and regulation of high light-induced genes in *Chlamydomonas reinhardtii*. Plant J. 2002:**30**(3):301–313.

**Itakura AK, Chan KX, Atkinson N, Pallesen L, Wang L, Reeves G, Patena W, Caspari O, Roth R, Goodenough U, et al.** A Rubisco-binding protein is required for normal pyrenoid number and starch sheath morphology in *Chlamydomonas reinhardtii*. Proc Natl Acad Sci U S A. 2019:**116**(37):18445–18454.

**Izumo A, Fujiwara S, Sakurai T, Ball SG, Ishii Y, Ono H, Yoshida M, Fujita N, Nakamura Y, Buléon A, et al.** Effects of granule-bound starch synthase I-defective mutation on the morphology and structure of pyrenoidal starch in Chlamydomonas. Plant Sci. 2011:**180**(2):238–245.

**Jin S, Sun J, Wunder T, Tang D, Cousins AB, Sze SK, Mueller-Cajar O, and Gao Y-G**. Structural insights into the LCIB protein family reveals a new group of β-carbonic anhydrases. Proc Natl Acad Sci U S A. 2016:**113**(51):14716–14721.

**Kaneko Y, Danev R, Nagayama K, and Nakamoto H**. Intact carboxysomes in a cyanobacterial cell visualized by Hilbert differential contrast transmission electron microscopy. J Bacteriol. 2006:**188**(2):805–808.

**Karlsson J**. A novel alpha -type carbonic anhydrase associated with the thylakoid membrane in Chlamydomonas reinhardtii is required for growth at ambient CO2. EMBO J. 1998:**17**(5):1208–1216.

**Kasili RW, Rai AK, and Moroney JV**. LCIB functions as a carbonic anhydrase: evidence from yeast and Arabidopsis carbonic anhydrase knockout mutants. Photosynth Res. 2023. https://doi.org/10.1007/s11120-023-01005-1

**Kerfeld CA and Melnicki MR**. Assembly, function and evolution of cyanobacterial carboxysomes. Curr Opin Plant Biol. 2016:**31**:66–75.

**Kerfeld CA, Sawaya MR, Tanaka S, Nguyen CV, Phillips M, Beeby M, and Yeates TO**. Protein structures forming the shell of primitive bacterial organelles. Science. 2005:**309**(5736):936–938.

**Kono A, Chou T-H, Radhakrishnan A, Bolla JR, Sankar K, Shome S, Su C-C, Jernigan RL, Robinson CV, Yu EW, et al.** Structure and function of LCI1: a plasma membrane $CO_2$ channel in the *Chlamydomonas* $CO_2$ concentrating mechanism. Plant J. 2020:**102**(6):1107–1126.

**Kono A and Spalding MH**. LCI1, a Chlamydomonas reinhardtii plasma membrane protein, functions in active CO2 uptake under low CO2. Plant J. 2020:**102**(6):1127–1141.

**Küken A, Sommer F, Yaneva-Roder L, Mackinder LC, Höhne M, Geimer S, Jonikas MC, Schroda M, Stitt M, Nikoloski Z, et al.** Effects of microcompartmentation on flux distribution and metabolic pools in Chlamydomonas reinhardtii chloroplasts. Elife. 2018:**7**. https://doi.org/10.7554/eLife.37960

**Liang Z, Yeung W-T, Ma J, Mai KKK, Liu Z, Chong Y-LF, Cai X, and Kang B-H**. Electron tomography of prolamellar bodies and their transformation into grana thylakoids in cryofixed Arabidopsis cotyledons. Plant Cell. 2022:**34**(10):3830–3843.

**Lin L, Guo D, Zhao L, Zhang X, Wang J, Zhang F, and Wei C**. Comparative structure of starches from high-amylose maize inbred lines and their hybrids. Food Hydrocoll. 2016:**52**:19–28.

**Long BM, Badger MR, Whitney SM, and Price GD**. Analysis of carboxysomes from Synechococcus PCC7942 reveals multiple rubisco complexes with carboxysomal proteins CcmM and CcaA. J Biol Chem. 2007:**282**(40):29323–29335.

**Lucker B and Kramer DM**. Regulation of cyclic electron flow in Chlamydomonas reinhardtii under fluctuating carbon availability. Photosynth Res. 2013:**117**(1–3):449–459.

**Mackinder LCM**. The Chlamydomonas CO2 -concentrating mechanism and its potential for engineering photosynthesis in plants. New Phytol. 2018:**217**(1):54–61.

**Mackinder LCM, Chen C, Leib RD, Patena W, Blum SR, Rodman M, Ramundo S, Adams CM, and Jonikas MC**. A Spatial Interactome Reveals the Protein

Organization of the Algal CO2-Concentrating Mechanism. Cell. 2017:**171**(1):133-147.e14.

**Mackinder LCM, Meyer MT, Mettler-Altmann T, Chen VK, Mitchell MC, Caspari O, Freeman Rosenzweig ES, Pallesen L, Reeves G, Itakura A, et al.** A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. Proc Natl Acad Sci U S A. 2016:**113**(21):5958–5963.

**Mahinthichaichan P, Morris DM, Wang Y, Jensen GJ, and Tajkhorshid E**. Selective permeability of carboxysome shell pores to anionic molecules. J Phys Chem B. 2018:**122**(39):9110–9118.

**Mariscal V, Moulin P, Orsel M, Miller AJ, Fernández E, and Galván A**. Differential regulation of the Chlamydomonas Nar1 gene family by carbon and nitrogen. Protist. 2006:**157**(4):421–433.

**Meyer MT, Genkov T, Skepper JN, Jouhet J, Mitchell MC, Spreitzer RJ, and Griffiths H**. Rubisco small-subunit α-helices control pyrenoid formation in *Chlamydomonas*. Proc Natl Acad Sci U S A. 2012:**109**(47):19474–19479.

**Meyer MT, Itakura AK, Patena W, Wang L, He S, Emrich-Mills T, Lau CS, Yates G, Mackinder LCM, and Jonikas MC**. Assembly of the algal $CO_2$-fixing organelle, the pyrenoid, is guided by a Rubisco-binding motif. Sci Adv. 2020:**6**(46). https://doi.org/10.1126/sciadv.abd2408

**Meyer MT, Whittaker C, and Griffiths H**. The algal pyrenoid: key unanswered questions. J Exp Bot. 2017:**68**(14):3739–3749.

**Miura K, Yamano T, Yoshioka S, Kohinata T, Inoue Y, Taniguchi F, Asamizu E, Nakamura Y, Tabata S, Yamato KT, et al.** Expression profiling-based identification of CO2-responsive genes regulated by CCM1 controlling a carbon-concentrating mechanism in *Chlamydomonas reinhardtii*. Plant Physiol. 2004:**135**(3):1595–1607.

**Moroney JV, Husic HD, and Tolbert NE**. Effect of carbonic anhydrase inhibitors on inorganic carbon accumulation by *Chlamydomonas reinhardtii*. Plant Physiol. 1985:**79**(1):177–183.

**Moroney JV, Husic HD, Tolbert NE, Kitayama M, Manuel LJ, and Togasaki RK**. Isolation and characterization of a mutant of *Chlamydomonas reinhardtii* deficient in the $CO_2$ concentrating mechanism. Plant Physiol. 1989:**89**(3):897–903.

**Moroney JV, Tolbert NE, and Sears BB**. Complementation analysis of the inorganic carbon concentrating mechanism of Chlamydomonas reinhardtii. Mol Gen Genet. 1986:**204**(2):199–203.

**Mukherjee A, Lau CS, Walker CE, Rai AK, Prejean CI, Yates G, Emrich-Mills T, Lemoine SG, Vinyard DJ, Mackinder LCM, et al.** Thylakoid localized bestrophin-like proteins are essential for the $CO_2$ concentrating mechanism of *Chlamydomonas reinhardtii*. Proc Natl Acad Sci U S A. 2019:**116**(34):16915–16920.

**Nishi A, Nakamura Y, Tanaka N, and Satoh H**. Biochemical and genetic analysis of the effects of amylose-extender mutation in rice endosperm. Plant Physiology. 2001:**127**(2):459.

**Ohnishi N, Mukherjee B, Tsujikawa T, Yanase M, Nakano H, Moroney JV, and Fukuzawa H**. Expression of a low CO2–inducible protein, LCI1, increases inorganic

carbon uptake in the Green alga *Chlamydomonas reinhardtii*. Plant Cell. 2010:**22**(9):3105–3117.

**Omata T, Price GD, Badger MR, Okamura M, Gohta S, and Ogawa T**. Identification of an ATP-binding cassette transporter involved in bicarbonate uptake in the cyanobacterium *Synechococcus* sp. strain PCC 7942. Proc Natl Acad Sci U S A. 1999:**96**(23):13571–13576.

**Prasad R, Gupta SK, Shabnam N, Oliveira CYB, Nema AK, Ansari FA, and Bux F**. Role of microalgae in global CO2 sequestration: Physiological mechanism, recent development, challenges, and future prospective. Sustainability. 2021:**13**(23):13061.

**Preiser AL, Banerjee A, Weise SE, Renna L, Brandizzi F, and Sharkey TD**. Phosphoglucoisomerase is an important regulatory enzyme in partitioning carbon out of the Calvin-Benson cycle. Front Plant Sci. 2020:**11**. https://doi.org/10.3389/fpls.2020.580726

**Qu Z and Hartzell HC**. Bestrophin $Cl^-$ channels are highly permeable to $HCO_3^-$. Am J Physiol Cell Physiol. 2008:**294**(6):C1371–C1377.

**Rae BD, Long BM, Badger MR, and Price GD**. Functions, compositions, and evolution of the two types of carboxysomes: Polyhedral microcompartments that facilitate $CO_2$ fixation in Cyanobacteria and some Proteobacteria. Microbiol Mol Biol Rev. 2013:**77**(3):357–379.

**Ramazanov Z, Rawat M, Henk MC, Mason CB, Matthews SW, and Moroney JV**. The induction of the $CO_2$ -concentrating mechanism is correlated with the formation of the starch sheath around the pyrenoid of Chlamydomonas reinhardtii. Planta. 1994:**195**(2):210–216.

**Rickaby REM and Eason Hubbard MR**. Upper ocean oxygenation, evolution of RuBisCO and the Phanerozoic succession of phytoplankton. Free Radic Biol Med. 2019:**140**:295–304.

**Roach T**. LHCSR3-type NPQ prevents photoinhibition and slowed growth under fluctuating light in Chlamydomonas reinhardtii. Plants. 2020:**9**(11):1604.

**Rottet S, Förster B, Hee WY, Rourke LM, Price GD, and Long BM**. Engineered accumulation of bicarbonate in plant chloroplasts: Known knowns and known unknowns. Front Plant Sci. 2021:**12**. https://doi.org/10.3389/fpls.2021.727118

**Samborska B and Kimber MS**. A dodecameric CcmK2 structure suggests β-carboxysomal shell facets have a double-layered organization. Structure. 2012:**20**(8):1353–1362.

**Sandoval-Ibáñez O, Sharma A, Bykowski M, Borràs-Gas G, Behrendorff JBYH, Mellor S, Qvortrup K, Verdonk JC, Bock R, Kowalewska Ł, et al.** Curvature thylakoid 1 proteins modulate prolamellar body morphology and promote organized thylakoid biogenesis in*Arabidopsis thaliana*. Proc Natl Acad Sci U S A. 2021:**118**(42). https://doi.org/10.1073/pnas.2113934118

**Seung D**. Amylose in starch: towards an understanding of biosynthesis, structure and function. New Phytol. 2020:**228**(5):1490–1504.

**Seung D, Boudet J, Monroe J, Schreier TB, David LC, Abt M, Lu K-J, Zanella M, and Zeeman SC**. Homologs of PROTEIN TARGETING TO STARCH control starch granule initiation in Arabidopsis leaves. Plant Cell. 2017:**29**(7):1657–1677.

**Seung D, Soyk S, Coiro M, Maier BA, Eicke S, and Zeeman SC**. PROTEIN TARGETING TO STARCH is required for localising GRANULE-BOUND STARCH SYNTHASE to starch granules and for normal amylose synthesis in Arabidopsis. PLoS Biol. 2015:**13**(2):e1002080.

**Shi X and Bloom A**. Photorespiration: The futile cycle? Plants. 2021:**10**(5):908.

**Shibata M, Katoh H, Sonoda M, Ohkawa H, Shimoyama M, Fukuzawa H, Kaplan A, and Ogawa T**. Genes essential to sodium-dependent bicarbonate transport in Cyanobacteria. J Biol Chem. 2002:**277**(21):18658–18664.

**Shikanai T and Yamamoto H**. Contribution of cyclic and pseudo-cyclic electron transport to the formation of proton motive force in chloroplasts. Mol Plant. 2017:**10**(1):20–29.

**Shimamura D, Ikeuchi T, Tsuji Y, Fukuzawa H, and Yamano T**. Periplasmic carbonic anhydrase CAH1 contributes to high inorganic carbon affinity in*Chlamydomonas reinhardtii*. bioRxiv. 2024. https://doi.org/10.1101/2024.03.04.583368

**Shimamura D, Yamano T, Niikawa Y, Hu D, and Fukuzawa H**. A pyrenoid-localized protein SAGA1 is necessary for Ca2+-binding protein CAS-dependent expression of nuclear genes encoding inorganic carbon transporters in Chlamydomonas reinhardtii. Photosynth Res. 2023. https://doi.org/10.1007/s11120-022-00996-7

**Smith SM, Fulton DC, Chia T, Thorneycroft D, Chapple A, Dunstan H, Hylton C, Zeeman SC, and Smith AM**. Diurnal changes in the transcriptome encoding enzymes of starch metabolism provide evidence for both transcriptional and posttranscriptional regulation of starch metabolism in Arabidopsis leaves. Plant Physiol. 2004:**136**(1):2687–2699.

**Sommer M, Sutter M, Gupta S, Kirst H, Turmo A, Lechno-Yossef S, Burton RL, Saechao C, Sloan NB, Cheng X, et al.** Heterohexamers formed by CcmK3 and CcmK4 increase the complexity of beta carboxysome shells. Plant Physiol. 2019:**179**(1):156–167.

**Spalding MH, Spreitzer RJ, and Ogren WL**. Carbonic anhydrase-deficient mutant of *Chlamydomonas reinhardii* requires elevated carbon dioxide concentration for photoautotrophic growth. Plant Physiol. 1983:**73**(2):268–272.

**Steensma AK, Shachar-Hill Y, and Walker BJ**. The carbon-concentrating mechanism of the extremophilic red microalga Cyanidioschyzon merolae. Photosynth Res. 2023:**156**(2):247–264.

**Tanaka S, Kerfeld CA, Sawaya MR, Cai F, Heinhorst S, Cannon GC, and Yeates TO**. Atomic-level models of the bacterial carboxysome shell. Science. 2008:**319**(5866):1083–1086.

**TerBush AD, Yoshida Y, and Osteryoung KW**. FtsZ in chloroplast division: structure, function and evolution. Curr Opin Cell Biol. 2013:**25**(4):461–470.

**Tetlow IJ and Bertoft E**. A review of starch biosynthesis in relation to the building block-backbone model. Int J Mol Sci. 2020:**21**(19):7011.

**Tiessen A, Hendriks JHM, Stitt M, Branscheid A, Gibon Y, Farré EM, and Geigenberger P**. Starch synthesis in potato tubers is regulated by post-translational redox modification of ADP-glucose pyrophosphorylase. Plant Cell. 2002:**14**(9):2191–2213.

**Toyokawa C, Yamano T, and Fukuzawa H**. Pyrenoid Starch Sheath Is Required for LCIB Localization and the CO2-Concentrating Mechanism in Green Algae. Plant Physiol. 2020:**182**(4):1883–1893.

**Van K and Spalding MH**. Periplasmic carbonic anhydrase structural gene (*Cah1*) mutant in *Chlamydomonas reinhardtii*1. Plant Physiol. 1999:**120**(3):757–764.

**Wang C, Sun B, Zhang X, Huang X, Zhang M, Guo H, Chen X, Huang F, Chen T, Mi H, et al.** Structural mechanism of the active bicarbonate transporter from cyanobacteria. Nat Plants. 2019a:**5**(11):1184–1193.

**Wang H, Yan X, Aigner H, Bracher A, Nguyen ND, Hee WY, Long BM, Price GD, Hartl FU, and Hayer-Hartl M**. Rubisco condensate formation by CcmM in β-carboxysome biogenesis. Nature. 2019b:**566**(7742):131–135.

**Wang L, Patena W, Van Baalen KA, Xie Y, Singer ER, Gavrilenko S, Warren-Williams M, Han L, Harrigan HR, Hartz LD, et al.** A chloroplast protein atlas reveals punctate structures and spatial organization of biosynthetic pathways. Cell. 2023:**186**(16):3499-3518.e14.

**Wang L, Yamano T, Takane S, Niikawa Y, Toyokawa C, Ozawa S-I, Tokutsu R, Takahashi Y, Minagawa J, Kanesaki Y, et al.** Chloroplast-mediated regulation of $CO_2$ -concentrating mechanism by $Ca^{2+}$ -binding protein CAS in the green alga *Chlamydomonas reinhardtii*. Proc Natl Acad Sci U S A. 2016:**113**(44):12586–12591.

**Wang Y and Spalding MH**. An inorganic carbon transport system responsible for acclimation specific to air levels of $CO_2$ in *Chlamydomonas reinhardtii*. Proc Natl Acad Sci U S A. 2006:**103**(26):10110–10115.

**Wang Y and Spalding MH**. Acclimation to very low $CO_2$: Contribution of limiting $CO_2$ inducible proteins, LCIB and LCIA, to inorganic carbon uptake in *Chlamydomonas reinhardtii*. Plant Physiol. 2014:**166**(4):2040–2050.

**Wang Y, Stessman DJ, and Spalding MH**. The CO2 concentrating mechanism and photosynthetic carbon assimilation in limiting CO2 : how Chlamydomonas works against the gradient. Plant J. 2015:**82**(3):429–448.

**Wang Y, Sun Z, Horken KM, Im C-S, Xiang Y, Grossman AR, and Weeks DP**. Analyses of CIA5, the master regulator of the carbon-concentrating mechanism in *Chlamydomonas reinhardtii*, and its control of gene expression. Can J Bot. 2005:**83**(7):765–779.

**Wunder T, Cheng SLH, Lai S-K, Li H-Y, and Mueller-Cajar O**. The phase separation underlying the pyrenoid-based microalgal Rubisco supercharger. Nat Commun. 2018:**9**(1):5076.

**Yamano T, Toyokawa C, Shimamura D, Matsuoka T, and Fukuzawa H**. CO2-dependent migration and relocation of LCIB, a pyrenoid-peripheral protein in *Chlamydomonas reinhardtii*. Plant Physiol. 2022:**188**(2):1081–1094.

**Yoshioka S, Taniguchi F, Miura K, Inoue T, Yamano T, and Fukuzawa H**. The novel Myb transcription factor LCR1 regulates the CO2-responsive gene *Cah1*, encoding a periplasmic carbonic anhydrase in *Chlamydomonas reinhardtii* [W]. Plant Cell. 2004:**16**(6):1466–1477.

**Yuan F, Alimohamadi H, Bakka B, Trementozzi AN, Day KJ, Fawzi NL, Rangamani P, and Stachowiak JC**. Membrane bending by protein phase separation. Proc Natl Acad Sci U S A. 2021:**118**(11). https://doi.org/10.1073/pnas.2017435118

**Zang K, Wang H, Hartl FU, and Hayer-Hartl M**. Scaffolding protein CcmM directs multiprotein phase separation in β-carboxysome biogenesis. Nat Struct Mol Biol. 2021:**28**(11):909–922.

**Zhan Y, Marchand CH, Maes A, Mauries A, Sun Y, Dhaliwal JS, Uniacke J, Arragain S, Jiang H, Gold ND, et al.** Pyrenoid functions revealed by proteomics in Chlamydomonas reinhardtii. PLoS One. 2018:**13**(2):e0185039.

**Zhang F, Yao X, Sun S, Wang L, Liu W, Jiang X, and Wang J**. Effects of mesotrione on oxidative stress, subcellular structure, and membrane integrity in Chlorella vulgaris. Chemosphere. 2020:**247**(125668):125668.

**Zhang N, Mattoon EM, McHargue W, Venn B, Zimmer D, Pecani K, Jeong J, Anderson CM, Chen C, Berry JC, et al.** Systems-wide analysis revealed shared and unique responses to moderate and acute high temperatures in the green alga Chlamydomonas reinhardtii. Commun Biol. 2022:**5**(1). https://doi.org/10.1038/s42003-022-03359-z

# Chapter 2 – A phase-separated $CO_2$-fixing pyrenoid proteome determined by TurboID in *Chlamydomonas reinhardtii*

## Declaration of Authorship

All experimental work presented below was carried out by Chun Sing Lau with the following exception:

Adam Dowle performed the LC-MS/MS experiment and the associated quantitation.
Philipp Girr generated the plasmids encoding STR16-mNeon for confocal imaging. Other confocal imaging experiments are performed by Chun Sing Lau

All figures are assembled by Chun Sing Lau

Chapter 2 is published as: "A phase-separated $CO_2$-fixing pyrenoid proteome determined by TurboID in *Chlamydomonas reinhardtii*"
Lau, C.S., Dowle, A., Thomas, G.H., Girr, P., and Mackinder, L.C.M.
The Plant Cell, Volume 35, Issue 9, September 2023, Pages 3260–3279,
https://doi.org/10.1093/plcell/koad131

## A phase-separated $CO_2$-fixing pyrenoid proteome determined by TurboID in *Chlamydomonas reinhardtii*

### Authors

Chun Sing Lau[1], Adam Dowle[1], Gavin H Thomas[1], Philipp Girr[1], Luke C M Mackinder[1]

[1]The University of York, Centre of Novel Agricultural Products, Department of Biology, United Kingdom

### Abstract

Phase separation underpins many biologically important cellular events such as RNA metabolism, signalling, and $CO_2$ fixation. However, determining the composition of a phase-separated organelle is often challenging due to its sensitivity to environmental conditions, which limits the application of traditional proteomic techniques like organellar purification or affinity purification mass spectrometry to understand their composition. In *Chlamydomonas reinhardtii*, Rubisco is condensed into a crucial phase-separated organelle called the pyrenoid that improves photosynthetic performance by supplying Rubisco with elevated concentrations of $CO_2$. Here, we developed a TurboID-based proximity labelling technique in which proximal proteins in Chlamydomonas chloroplasts are labelled by biotin radicals generated from the TurboID-tagged protein. By fusing 2 core pyrenoid components with the TurboID tag, we generated a high-confidence pyrenoid proxiome that contains most known pyrenoid proteins, in addition to new pyrenoid candidates. Fluorescence protein tagging of 7

previously uncharacterized TurboID-identified proteins showed that 6 localized to a range of subpyrenoid regions. The resulting proxiome also suggests new secondary functions for the pyrenoid in RNA-associated processes and redox-sensitive iron–sulfur cluster metabolism. This developed pipeline can be used to investigate a broad range of biological processes in Chlamydomonas, especially at a temporally resolved suborganellar resolution.

## Introduction

Nearly all algae contain a microcompartment in their chloroplast called the pyrenoid, which is estimated to be responsible for ~30% of global $CO_2$ fixation (Mackinder et al. 2016). The pyrenoid of the model green alga *Chlamydomonas* (*Chlamydomonas reinhardtii*) is a 1- to 2-µm biomolecular condensate of the principal $CO_2$-fixing enzyme Rubisco. It is formed through liquid–liquid phase separation (LLPS) of Rubisco mediated by ESSENTIAL PYRENOID COMPONENT 1 (EPYC1), which harbours 5 evenly spaced Rubisco-binding motifs (RBMs) interspaced by disordered sequences (Mackinder et al. 2016; Freeman Rosenzweig et al. 2017; Wunder et al. 2018; He et al. 2020). The deletion of EPYC1, or the reciprocal binding site of EPYC1 on Rubisco, abolishes pyrenoid formation. Correct pyrenoid assembly is essential for a functional $CO_2$-concentrating mechanism (CCM) (Mackinder et al. 2016) that works to saturate Rubisco with $CO_2$ to minimize energetically costly photorespiration, thereby improving photosynthetic efficiency (Wang et al. 2015; Fei et al. 2022). In the face of growing food security issues, the engineering of a pyrenoid-based CCM into major C3 crop plants such as rice (*Oryza sativa*), soybean (*Glycine max*), and wheat (*Triticum aestivum*) is regarded as a promising strategy for yield improvement, with prospects of increasing food production by up to 60% (Ray et al. 2013; Long et al. 2019). Recent work reconstituted a proto-pyrenoid in the land plant Arabidopsis (*Arabidopsis thaliana*) (Atkinson et al. 2020). However, additional structural components, such as those needed to traverse thylakoid membranes and establish a $CO_2$ diffusion barrier will be required for efficient function (Fei et al. 2022). Many of the proteins underpinning these additional structural requirements are unknown, making a deep understanding of the structural organization and molecular function of the pyrenoid critical.

Previous pyrenoid proteomes have been achieved via organelle purification (Mackinder et al. 2016; Zhan et al. 2018) and affinity purification followed by mass spectrometry (AP-MS) (Mackinder et al. 2017); however, these methods have limitations. While multiple robust methods, like AP-MS, exist to identify strong protein–protein interactions, the ability to identify weak and transient interactions in vivo is limited. At a larger spatial scale, subcellular fractionation followed by protein purification and MS is prone to cross-contamination (Christopher et al. 2021). Biomolecular condensates, like the pyrenoid, fall into a class of subcellular structures whose proteomes are challenging to accurately determine as they are

typically dynamic, involving weak and transient interactions that are highly sensitive to small changes in the surrounding environment, can vary considerably in size, and are not always clearly spatially defined due to the absence of an encapsulating membrane (Hyman et al. 2014; Choi et al. 2020; Barrett et al. 2021). Recently developed proximity labeling methods such as APEX2 and TurboID (Lam et al. 2015; Branon et al. 2018) are particularly poised to determine the transient interactions and proteomes of biomolecular condensates (Bracha et al. 2019). APEX2 and TurboID use an enzyme tag that drives biotinylation of neighboring proteins in vivo. In APEX2, an engineered ascorbate peroxidase converts biotin-phenol to biotin-phenoxyl radicals; with TurboID, an engineered biotin ligase generates biotin-5´-AMP radicals from biotin and ATP (Roux et al. 2012; Branon et al. 2018). These labile radicals spontaneously biotinylate the surface of exposed residues from proteins in close proximity. This reaction gives rise to a localized biotinylation event that is spatially restricted to 10 to 40 nm (Kim et al. 2014, 2016) by the diffusion of the radical from the enzyme tag. This in vivo biotinylation method bypasses the need to purify proteins in their native association, with the high affinity of the biotin tag to streptavidin beads enabling the removal of background contaminants via harsh wash conditions. Proximity labeling thus results in the identification of strong, weak, and transient interactions, in addition to noninteracting proximal proteins. However, since its development, proximity labeling has seen limited application in phase-separated systems (Youn et al. 2018; Zhou and Zou 2021) and has yet to be established in plastids or the alga *Chlamydomonas*.

In this study, we attempted to identify those proteins that were missed by AP-MS and pyrenoid purification by developing a pyrenoid-based proximity labelling methodology. Using TurboID-based proximity labelling, we identify a complementary and robust pyrenoid "proxiome." Our pyrenoid proxiome contains most previously known pyrenoid proteins and has identified multiple new pyrenoid components that show distinct subpyrenoid localizations, as determined via fluorescence tagging. The ability to identify core proteins involved in pyrenoid phase separation highlights the strength of proximity labelling for investigating biomolecular condensate composition and formation. Furthermore, our method establishes proximity labelling in plastids and the leading model algal system, *Chlamydomonas*.

## Results

### Development of proximity labeling in *Chlamydomonas*

We set out to establish proximity labeling in the LLPS pyrenoid within the Chlamydomonas chloroplast (Fig. 1A). TurboID has been established in Arabidopsis (Zhang et al. 2019; Mair and Bergmann 2022) and APEX2 in cyanobacteria (Dahlgren et al. 2021) and diatoms

(Turnšek et al. 2021). To determine which approach is best suited for Chlamydomonas, we designed constructs to test both APEX2 and TurboID (Supplemental Data Set 1). We designed expression constructs to be compatible with the Chlamydomonas modular cloning (MoClo) framework (Crozet et al. 2018) to enable community adoption and compatibility with a broad range of promoters, terminators, and selection markers.

Figure 1

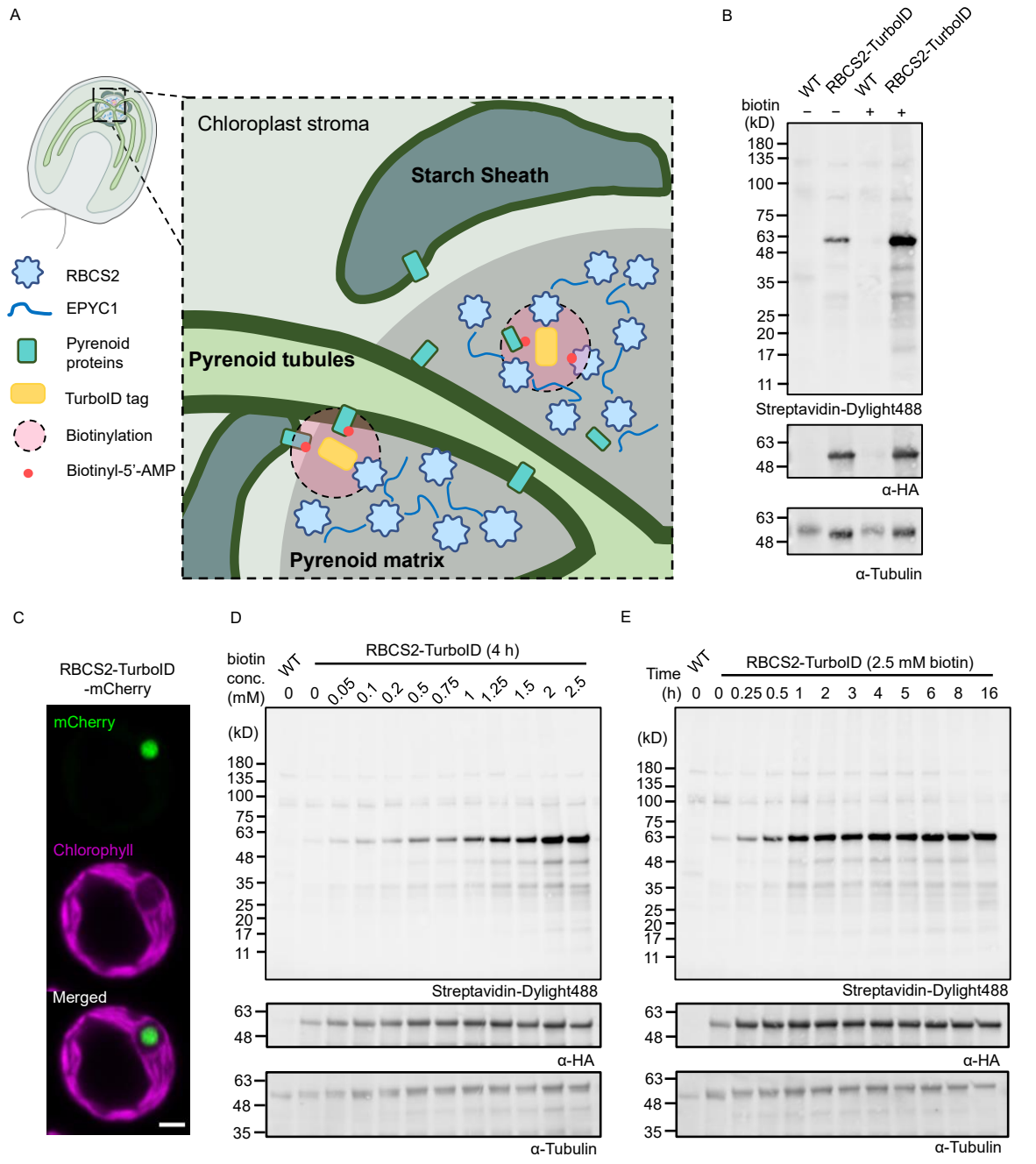



**Figure 1.** Establishment and optimization of TurboID labeling in the Chlamydomonas chloroplast using RBCS2-TurboID lines. **A)** Schematic representation of the *Chlamydomonas* pyrenoid and RBCS2-TurboID. The pyrenoid matrix is surrounded by a starch sheath and traversed by pyrenoid

tubules. The RBCS2-TurboID fusion protein is targeted to the pyrenoid matrix; upon addition of the biotin substrate, short-lived biotin radicals (round red dots) diffuse from the TurboID tag and spontaneously biotinylate neighbouring pyrenoid proteins. **B)** Biotinylation signals of strains transformed with the RBCS2-TurboID construct and the untagged background (WT) were assessed by immunoblotting whole-cell lysate with a streptavidin conjugate. Anti-tubulin was used as a loading control, with anti-HA used to probe for abundance of the fusion protein. **C)** Confocal imaging of RBCS2-TurboID-mCherry. Green and magenta signals represent the mCherry and chlorophyll autofluorescence respectively. Scale bar is 2 µm. **D to E)** RBCS2-TurboID labelling efficiency was determined by labelling cells across a biotin concentration gradient (0 to 2,500 µM) for 4 h **D)** or across a time range (0 to 16 h) with 2.5 mM biotin substrate **E)**.

We initially chose the Rubisco small subunit 2 (RBCS2, encoded by Cre02.g120150) as our bait due to (i) the central role of Rubisco in pyrenoid LLPS (Meyer et al. 2012; Wunder et al. 2018); (ii) previous data showing that tagging exogenous RBCS does not affect CCM functionality (Freeman Rosenzweig et al. 2017); and (iii) the availability of known interacting partners for downstream validation (Mackinder et al. 2017; Meyer et al. 2020). We thus fused either the APEX2 or the TurboID tag to the C-terminus of RBCS2 and placed the encoding cassette under the control of the well-established PSAD promoter/terminator pair previously used for fluorescence protein tagging of a broad range of pyrenoid components including RBCS2 (Mackinder et al. 2017). We transformed all constructs individually via electroporation into the widely used wild-type (WT) strain CC-4533 (Li et al. 2016, 2019). We screened hygromycin-resistant colonies for genomic insertion of the RBCS2 fusion construct via PCR and then for protein accumulation by immunoblotting against the C-terminal epitope tag (Supplemental Figs. S1A and S2A). We named the resulting strains harboring each construct RBCS2-APEX2 and RBCS2-TurboID.

We confirmed the correct localization of RBCS2-APEX2 to the pyrenoid by immunofluorescence against the 3xFlag tag at the C-terminus of APEX2 (Supplemental Fig. S1B). To validate the activity of RBCS2-APEX2, we incubated RBCS2-APEX2 strain A2 (Supplemental Fig. S1A) with the biotin-phenol substrate, which showed a subtle yet different biotinylation pattern from that of the untagged WT background, especially when activated with higher $H_2O_2$ concentration (Supplemental Fig. S1C). This observation led us to pursue a preliminary labeling experiment followed by MS of affinity-purified biotinylated proteins. Analysis of these data showed minimal enrichment for Rubisco or known pyrenoid components (Supplemental Fig. S1D). However, when assessing APEX2 peroxidase activity using Amplex Red, we detected higher peroxidase activity in RBCS2-APEX2 than in its untagged counterpart, suggesting that the fusion protein is functional (Supplemental Fig. S1E). We tentatively conclude that biotin-phenol has limited cellular permeability resulting in poor labeling. This poor permeability agrees with previous reports in budding yeast

(*Saccharomyces cerevisiae*), where cell wall modification was required to facilitate biotin-phenol uptake (Hwang and Espenshade 2016; Li et al. 2020). The failure of APEX2 to work in Chlamydomonas was also reported by Kreis et al. (2022).

By contrast, initial tests of RBCS2-TurboID showed clear increased biotinylation in comparison to WT with the addition of the biotin substrate (Fig. 1B). We observed a pronounced band at ~50 kD that likely corresponds to either the self-biotinylation of the RBCS2-TurboID fusion protein (55 kD) or the Rubisco large subunit (55 kD) (Fig. 1B). We also observed a weak biotinylation signal in the absence of external biotin addition, indicating that naturally occurring biotin is present in the chloroplast, as suggested by the presence of endogenously biotinylated chloroplast proteins (Li-Beisson et al. 2015).

After demonstrating TurboID activity, we assessed the localization of the fusion protein by generating a RBCS2-TurboID-mCherry fusion. Confocal imaging confirmed its pyrenoid localization, with the mCherry signal forming a single punctum at the canonical pyrenoid position characterized by an absence of chlorophyll fluorescence signal (Fig. 1C). We next optimized the concentration of the biotin substrate and labeling time (Fig. 1, D and E). To this end, we grew cells photoautotrophically with air-level $CO_2$ supplementation to induce the CCM, which leads to nearly all Rubisco being condensed into the pyrenoid (Borkhsenious et al. 1998). We then incubated these cells with a range of biotin concentrations (0.1 to 2.5 mM) over different time periods (1 to 16 h). We determined that biotin labeling occurs in a substrate- (Fig. 1D) and time- (Fig. 1E) dependent manner. In contrast to land plants where labeling saturation can be achieved with 50 µM biotin (Mair et al. 2019; Wurzinger et al. 2022), labeling in Chlamydomonas appears to saturate at the much higher biotin concentration of 2.5 mM. This result is in line with Kreis et al. (2022) with the use of a 1 mM concentration. To maximize labeling, we performed all later experiments using a final concentration of 2.5 mM biotin. In agreement with previous reports (Mair et al. 2019; Zhang et al. 2019), we similarly observed the rapid activity by TurboID, which allowed labeling to approach saturation after ~1 h (Fig. 1E).

**RBCS2-TurboID labels Rubisco interactors and pyrenoid proteins**

We established a pipeline for streptavidin affinity purification and protein identification by liquid chromatography–tandem MS (LC-MS/MS) (Fig. 2A; Materials and methods). Due to the relatively high levels of background biotinylation, we set out to further optimize labeling time in a pilot experiment. Accordingly, we incubated RBCS2-TurboID and the untagged WT strains with 2.5 mM biotin across a range of durations (1, 2, 4, and 8 h). We then subjected proteins extracted from the labelled cells to affinity purification with streptavidin magnetic beads. We detected a total of 918 proteins by LC-MS/MS across all samples. Initial results

showed a strong enrichment for core pyrenoid localized proteins, including RBCS1, RbcL, EPYC1, STARCH GRANULES ABNORMAL 2 (SAGA2), RUBISCO-BINDING MEMBRANE PROTEIN 1 (RBMP1), and RBMP2 when compared to WT cells not expressing RBCS2-TurboID (Fig. 2B and Supplemental Data Set 2). Using the detected proteins, we manually curated 4 benchmark protein sets with known localizations from the literature, namely, pyrenoid-specific proteins (P; 18 proteins), proteins found in the pyrenoid and the stroma (PS; 10 proteins), proteins found in the stroma but excluded from the pyrenoid (S; 15 proteins), and nonchloroplast proteins (NC; 13 proteins) (Fig. 2C and Supplemental Data Set 3). We used these benchmark proteins to calculate the enrichment threshold used to assess significant pyrenoid enrichment by applying a receiver–operator characteristic (ROC) analysis (Branon et al. 2018); Fig. 2D). For the ROC analysis, we adopted a stringent threshold by considering true positive proteins as exclusively pyrenoid-localized (P) proteins. It should be noted that a portion of the pyrenoid-localized proteins used for ROC analysis does not partition within the LLPS pyrenoid matrix but localizes to the starch plate or the pyrenoid tubules. However, we reasoned that their close association to the pyrenoid would still support their labelling by RBCS2-TurboID.

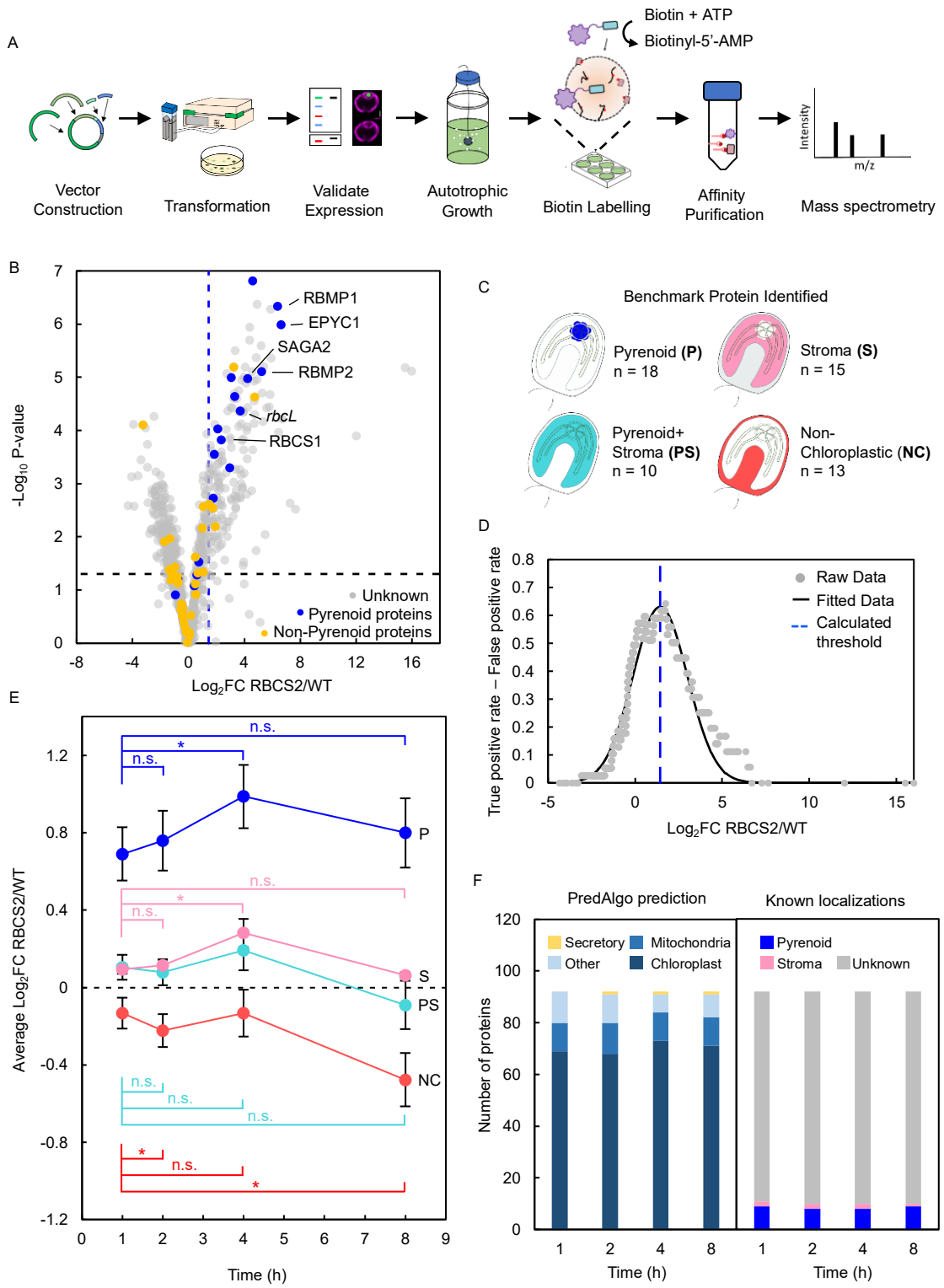



**Figure 2.** TurboID pipeline development and optimization of labelling time. **A)** Schematic representation of the developed TurboID pipeline. **B)** Volcano plot representing $Log_2$ FC between protein abundance in RBCS2-TurboID and WT. Proteins are coloured according to their localization: unknown (gray), pyrenoid proteins (blue), and other localizations including chloroplast stroma,

pyrenoid + stroma, and nonchloroplastic (yellow). The $Log_2$ FC threshold (dashed blue line) was calculated via the ROC analysis where only pyrenoid proteins are considered true positives. $-Log_{10}$P-value was used to represent statistical significance from the 1-way ANOVA test carried out on the difference in abundance between RBCS2-TurboID and WT. P-value of <0.05 was used as a threshold. **C)** Benchmark proteins detected from the RBCS2-TurboID sample; a complete list of benchmark proteins used is given in Supplemental Data Set 3. **D)** Trade-off between the true-positive rate and false-positive rates plotted against the Log2 FC value. A Gaussian function was fitted to the experimental data to determine a maximum, which was used as the enrichment threshold used in **B)**. **E)** $Log_2$ FC of RBCS2-TurboID according to localisation category in **C)** calculated at each labelling time point. Statistical significance was tested between time points within each class of benchmark proteins by a 2-way repeated measures ANOVA. *: $P < 0.01$; n.s.: nonsignificant comparison (Supplemental Data Set 8). **F)** PredAlgo-predicted localization and benchmark protein categories of the top 10% enriched proteins from RBCS2-TurboID at each labelling time point.

We then investigated protein labelling by RBCS2-TurboID at each time point for the different benchmark sets (Fig. 2E). We established that pyrenoid-localized proteins (P) consistently show the highest labelling across all time points, with both Pyrenoid proteins and Stromal proteins exhibiting a statistically significant increase in labelling from 1 to 4 h, while nonchloroplastic proteins remained stable. Interestingly, all benchmark proteins appear to decrease in labelling at the 8-h time point. This decrease is due to an increase in biotinylated protein abundance in untagged WT, rather than lower labelling by RBCS2-TurboID (Supplemental Data Set 2). While further testing on finer time points will be required to establish the true saturation point in biotin labelling, our data suggest that protein labelling by RBCS2-TurboID begins to approach saturation around 4 h. When we compared the top 10% of enriched proteins across the 4 time points, we detected consistent agreement with their predicted cellular localization, consistent enrichment for pyrenoid-localized proteins (Fig. 2F) and a >72% overlap in protein identity (Supplemental Data Set 2).

Collectively, most pyrenoid-localized proteins can be enriched within the first hour; however, increasing incubation time leads to increased biotinylation. Excluding the 8-h time point due to its increased background abundance, the largest differences between pyrenoid proteins, pyrenoid-excluded stromal proteins and nonchloroplast proteins occur at 4 h. We thus opted for 4-h incubations for later experiments. We hypothesize that the rapid labeling dynamics of pyrenoid proteins within the first hour and the slower increase in labeling of stromal proteins can be explained by the LLPS properties of the pyrenoid where Rubisco is present in both the condensed phase (pyrenoid) and dilute phase (stroma). The high concentration of Rubisco in the condensed phase enables rapid labeling of proximal pyrenoid proteins. However, as Rubisco is also in the dilute phase at a much lower concentration, stromal proteins are biotinylated at a slower rate. This idea is further supported by experimental

studies that show, under similar growth conditions used for our experiments, that ~90% of Rubisco is in the pyrenoid with the rest in the stroma (Borkhsenious et al. 1998; Mackinder et al. 2016).

**Stromal-TurboID controls enable a refined pyrenoid proteome**

Although our current approach enabled enrichment of pyrenoid proteins, we wished to refine the pyrenoid proteome by trying to distinguish between pyrenoid-specific proteins and proteins that are found within the pyrenoid and the stroma and to remove the bias of increased labelling of abundant background proteins—a typical challenge in proximity labelling studies (Han et al. 2018). To achieve this goal, we developed 2 chloroplast stromal controls and an additional pyrenoid-specific TurboID strain. For stromal controls, we identified 2 Calvin-cycle enzymes, RIBULOSE EPIMERASE 1 (RPE1; encoded by Cre12.g511900) and PHOSPHORIBULOKINASE 1 (PRK1, encoded by Cre12.g554800), which are abundant and localize to the chloroplast stroma but are excluded from the pyrenoid matrix (Fig. 3A) (Küken et al. 2018). We chose the Rubisco linker protein EPYC1 as an additional pyrenoid-specific protein due to its abundance and functional importance for the LLPS of Rubisco to form the pyrenoid (Mackinder et al. 2016). We assembled these new constructs, RPE1-TurboID, PRK1-TurboID, and EPYC1-TurboID, with the TurboID cloned in frame at the 3´ end of each coding sequence. We introduced each construct into CC-4533 as above and assessed the activity of the resulting positive strains alongside the RBCS2-TurboID strain (Figs. 3B, S2, A and B, and S3). While we saw evidence for EPYC1-TurboID protein degradation with increased biotin incubation time, the overall biotinylation signal does not appear to be perturbed (Supplemental Fig. S2B). We therefore continued with EPYC1-TurboID with a 4-h biotin incubation time in our MS experiments.

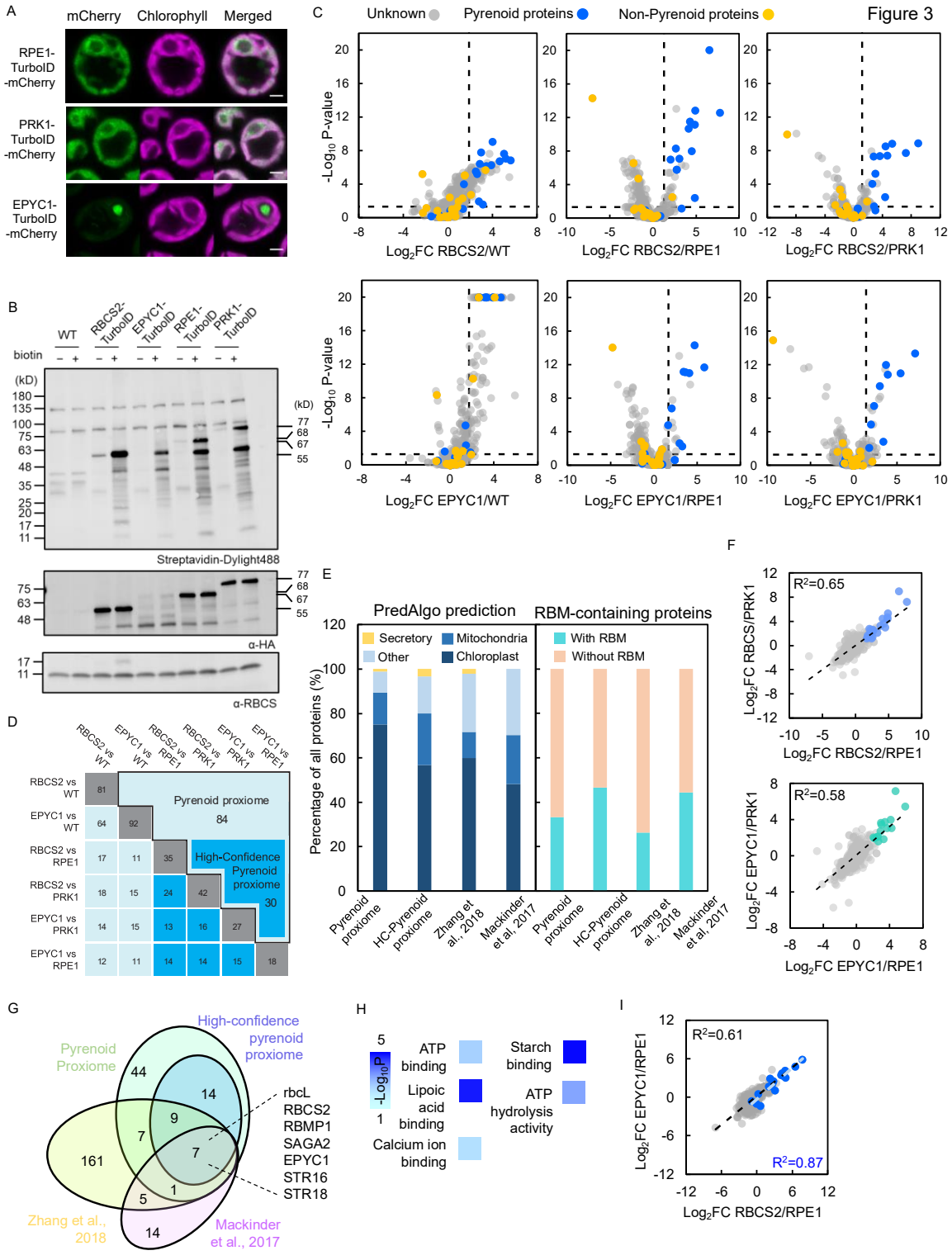



**Figure 3.** Determining the pyrenoid proteome using proximity labeling. **A)** Localization of the mCherry fusions of RPE1-TurboID, PRK1-TurboID, and EPYC1-TurboID. Green and magenta signals represent mCherry and chlorophyll autofluorescence, respectively. Scale bar is 2 µm. **B)** Labeling activity of RBCS2-TurboID, EPYC1-TurboID, RPE1-TurboID, and PRK1-TurboID strains, as determined in the absence (−) or presence (+) of 2.5 mM biotin for 4 h. Biotinylation was visualized via immunoblotting whole-cell lysate with a streptavidin conjugate. Abundance of RBCS2-TurboID (55

kD), EPYC1-TurboID (68 kD), RPE1-TurboID (67 kD), and PRK1-TurboID (77 kD) was probed by anti-HA. Anti-RBCS was used as a loading control. **C)** Volcano plots representing the $Log_2$ FC of RBCS2-TurboID and EPYC1-TurboID compared to WT and stromal controls. Pyrenoid proteins (blue dots) and non-pyrenoid proteins (yellow dots) were used to calculate the enrichment thresholds (vertical dashed line); the values are as follows: RBCS2/WT (1.88); RBCS2/RPE1 (1.31); RBCS2/PRK1 (1.14); EPYC1/WT (1.74); EPYC1/RPE1 (1.67); and EPYC1/PRK1 (1.42). Statistical significance for each pairwise comparison was calculated using the PEAKSQ method, a significance P-value cutoff of <0.05 was used (horizontal dashed line). The maximum $-Log_{10}$P-value computed by PEAKSQ was 20. **D)** Overlap matrix of identified proteins that are above the enrichment threshold in each treatment group. Bolded border highlights the overall pyrenoid proxiome, while the dark blue shaded box denotes the HC-pyrenoid proxiome. For both the pyrenoid proxiome and HC-pyrenoid proxiome, proteins had to be above the threshold in two or more comparisons. **E)** Predicted localization obtained from PredAlgo (Tardif et al. 2012) and percentage of RBM-containing proteins (Meyer et al. 2020) in the pyrenoid proxiome, the HC-pyrenoid proxiome, and previous published pyrenoid proteomes (Mackinder et al. 2017; Zhan et al. 2018). **F)** Comparison of Log2 FC in RBCS2-TurboID and EPYC1-TurboID between the 2 stromal controls. Statistically significant proteins that passed the ROC enrichment threshold are colored. **G)** Venn diagram showing the overlap between the pyrenoid proxiome and HC-pyrenoid proxiome (Mackinder et al. 2017; Zhan et al. 2018). **H)** GO enrichment analysis of the HC-pyrenoid proxiome (n = 30) using the PANTHER GO Complete Molecular Function data set. Significance as $-Log_{10}$P-value calculated from Fisher's exact test is presented in a color gradient. Only the GO terms of the most specific subclass that were represented by 2 or more proteins are shown. **I)** Comparison of protein enrichment between RBCS2-TurboID and EPYC1-TurboID. Blue dots represent known pyrenoid proteins. The black and blue dashed line represents the calculated trendline using all proteins or known pyrenoid proteins, respectively.

To ensure optimal conditions for identifying the pyrenoid proteome, we grew all expression strains photoautotrophically in 0.04% (v/v) $CO_2$ where nearly all of Rubisco is recruited to the pyrenoid and the CCM is fully induced (Mackinder 2018). Labeling was allowed to proceed for 4 h before we enriched for the resulting biotinylated proteins with streptavidin beads (see Materials and methods). Samples in triplicate were tandem mass tag (TMT) labeled to enable a relative quantification and comparison of protein abundance between each strain (Supplemental Data Set 4). We identified a total of 831 proteins derived from 5,227 peptides, with each protein containing at least 2 unique peptides. We calculated the $Log_2$ fold-change (FC) in reporter ion intensity between the pyrenoid-specific TurboID strains (RBCS2-TurboID and EPYC1-TurboID) and controls (WT, RPE1-TurboID, and PRK1-TurboID). We then determined the enrichment of pyrenoid proteins in each comparison. In agreement with our previous pilot experiment, we observed that pyrenoid proteins are predominantly enriched by the pyrenoid-specific TurboID strains across all comparison groups (Fig. 3C, blue dots).

To calculate the enrichment threshold used to assess significant pyrenoid enrichment, we applied the ROC analysis as in Fig. 2, C and D, and a significance threshold of $P < 0.05$ calculated by the PEAKSQ significance test (Cox and Mann 2008). We applied this analysis across all 6 comparison groups (Fig. 3C). This analysis yielded 141 unique proteins across the 6 groups (Supplemental Data Set 5). To remove out possible non-pyrenoid localized proteins, we only considered as true pyrenoid components those identified proteins that were consistently above the enrichment threshold in at least 2 of the comparison groupings. We obtained a final set of 84 unique proteins that we termed the “pyrenoid proxiome” (Fig. 3D, black bordered box). The pyrenoid proxiome contains 14 out of 19 known pyrenoid components detected in our data set and is highly enriched for proteins that are predicted to be targeted to the chloroplast (Fig. 3E).

Next, we set out to see if comparison against stromal control strains improves distinction between pyrenoid proteins and stromal proteins relative to a WT control. We first tested if there were any major differences between our 2 stromal controls. Plotting the Log2 FC of RBCS2-TurboID/RPE1-TurboID versus that of RBCS2-TurboID/PRK1-TurboID showed a strong correlation ($R^2 = 0.65$; Fig. 3F), suggesting that both controls give similar results and that their similar stromal localization is the main driver of protein labelling. We obtained a similar result when comparing EPYC1 against the 2 stromal controls ($R^2 = 0.58$; Fig. 3F). We next determined the difference between mean $Log_2$ FC of known pyrenoid and stromal proteins in each comparison pair (i.e. RBCS2-TurboID vs. WT, RBCS2-TurboID vs. RPE1-TurboID, and so on). Indeed, the difference between mean $Log_2$ FC of pyrenoid and stromal proteins was most evident in the stromal control comparisons (Supplemental Fig. S4). This result is further supported by our observation that proteins peripheral to the pyrenoid Rubisco-EPYC1 matrix but not in it, such as LOW-$CO_2$-INDUCIBLE PROTEIN B (LCIB), LCIC, STARCH SYNTHASE 2 (STA2), and STARCH BRANCHING ENZYME 3 (SBE3) (Yamano et al. 2010; Mackinder et al. 2017), are not enriched when stromal-specific TurboID strains are used as controls in place of WT. Our data here indicate that using the stromal controls gives a robust proteome of the Rubisco matrix. Taking proteins that are only seen above the threshold in 2 or more comparisons with stromal controls gives us 30 proteins (Supplemental Data Set 6). We named this set the “high-confidence pyrenoid proxiome” (HC-pyrenoid proxiome) (Fig. 3D). Compared to the pyrenoid proxiome, the HC-pyrenoid proxiome contains most known pyrenoid proteins (11/14) found in the former candidate pools. Similar to the abovementioned changes, proteins excluded from the HC-pyrenoid proxiome are either peripheral to the pyrenoid (LCIB and SBE3) or thylakoid membrane proteins (RBMP2). There was also a higher representation of RBM-containing proteins in the HC-pyrenoid proxiome (13/30, ~43%) than in the pyrenoid proxiome (28/84, ~28.5%). Our

results here give further support for the notion that the use of a compartment control yields a much more precise pyrenoid proxiome. We also evaluated our pyrenoid proxiome against published pyrenoid proteomic data obtained from either pyrenoid purification followed by MS (Zhan et al. 2018) or RBCS1/2 and EPYC1 AP-MS (Mackinder et al. 2017). We determined that 24/84 proteins within the pyrenoid proxiome and 16/30 of the HC-pyrenoid proteome overlap with at least 1 of the published data sets (Fig. 3G). Overall, 7 proteins are present in all 4 data sets, and 5 are known pyrenoid-localized proteins. Taking RBM-containing proteins as a proxy for pyrenoid localization, the HC-pyrenoid proxiome shows the highest fraction of RBM proteins (Fig. 3E). The similarity between our obtained proxiome and previously published data supports proximity labeling as a highly comparable technique to investigate organelle composition.

Gene ontology (GO) term enrichment analysis of the HC-pyrenoid proxiome indicated that these proteins can be functionally grouped into a small number of biological processes (Fig. 3H). These included lipoic acid binding, which represents sulfur-related compounds (GO: GO:0031405), carbohydrate-related processes like alpha-amylase activity and starch-binding (GO: GO:0004556 and GO:2001070, respectively) and ATP-binding groups (GO:0005524). We identified multiple proteins in the HC-pyrenoid proxiome, notably the proteins encoded by Cre06.g269650, Cre03.g158050, and SAGA1 that contain a starch-binding domain alongside a variety of functional domains. This similarity suggests that the matrix–starch interface might act as a specialized site for specific structural or biological functions. A broader analysis of the pyrenoid proxiome also reveals that multiple proteins contain iron–sulfur (Fe-S)-binding domains (encoded by Cre05.g240850, Cre13.g592200, and Cre02.g093650) or have RNA-related functions (encoded by Cre10.g440050, Cre10.g435800, Cre09.g393358, and Cre13.g578650). Tentatively, enrichment of these proteins in the pyrenoid proxiome suggests that the pyrenoid might take on other roles in addition to carbon fixation.

**RBCS-TurboID and EPYC1-TurboID generate comparable pyrenoid proteomes**

Rubisco and EPYC1 are the 2 major components of the pyrenoid. Their interactions with each other are both essential for phase separation and pyrenoid formation. However, an AP-MS study using both RBCS2 (and RBCS1) and EPYC1 as baits identified multiple distinct interacting partners as well as a shared set of interactors (Mackinder et al. 2017). Since the majority of pyrenoid proteins we used as a benchmark in this study were previously characterized due to their interactions with RBCS, it is difficult to ascertain whether the use of RBCS2-TurboID preferentially labelled Rubisco interactors or the broader pyrenoid proteome. We reasoned that by comparing proteins obtained from RBCS2-TurboID against those with EPYC1-TurboID, we might be able to distinguish between these 2 possibilities

and more broadly determine if proximity labelling of proteins in a dynamic molecular condensate preferentially labels the proteome of the condensate or the direct interactors of the bait. A comparison of EPYC1-TurboID and RBCS2-TurboID's respective FC against the stromal controls showed a strong correlation ($R^2$ = 0.61); this was considerably strengthened when focusing on known pyrenoid proteins ($R^2$ = 0.87; Fig. 3I, blue dots). We also found 12 out of 15 proteins identified in the EPYC1-TurboID relative to stromal controls (RPE1 and PRK1) in the RBCS2-TurboID relative to stromal controls analysis (Fig. 3D and Supplemental Data Set 5). In conclusion, irrespective of bait, using a mobile protein of the phase-separated pyrenoid yields a high-confidence proteome of the biomolecular condensate.

**Proximity labeling identifies new pyrenoid proteins**

To validate our pyrenoid proteome, we chose 7 proteins lacking localization data from the preliminary data and our initial RBCS2-TurboID versus WT comparison for fluorescence tagging (Fig. 4A). We primarily selected these proteins based on either occurrence in previous interactome/pull-down data sets (SULFURTRANSFERASE 16 [STR16; Cre13.g573250], STR18 [Cre16.g663150], and ATP-BINDING CASSETTE FAMILY F 6 [ABCF6; Cre06.g271850]; see Fig. 3G) or domain homology to known pyrenoid proteins (Cre03.g172700, Cre09.g394510, and Cre17.g720450). We cloned the open reading frame plus ~2,000-bp upstream of each target gene in-frame with the sequence encoding the fluorescent proteins Venus or mScarlet-I by recombineering to retain their native promoter (Emrich-Mills et al. 2021). We then transformed each construct into WT Chlamydomonas. Six of the 7 tagged proteins showed a primarily pyrenoid localization, with a broad range of subpyrenoid localization patterns (Figs. 4B and S5). Their localization patterns and their domain annotations provide novel insights into pyrenoid function and formation.

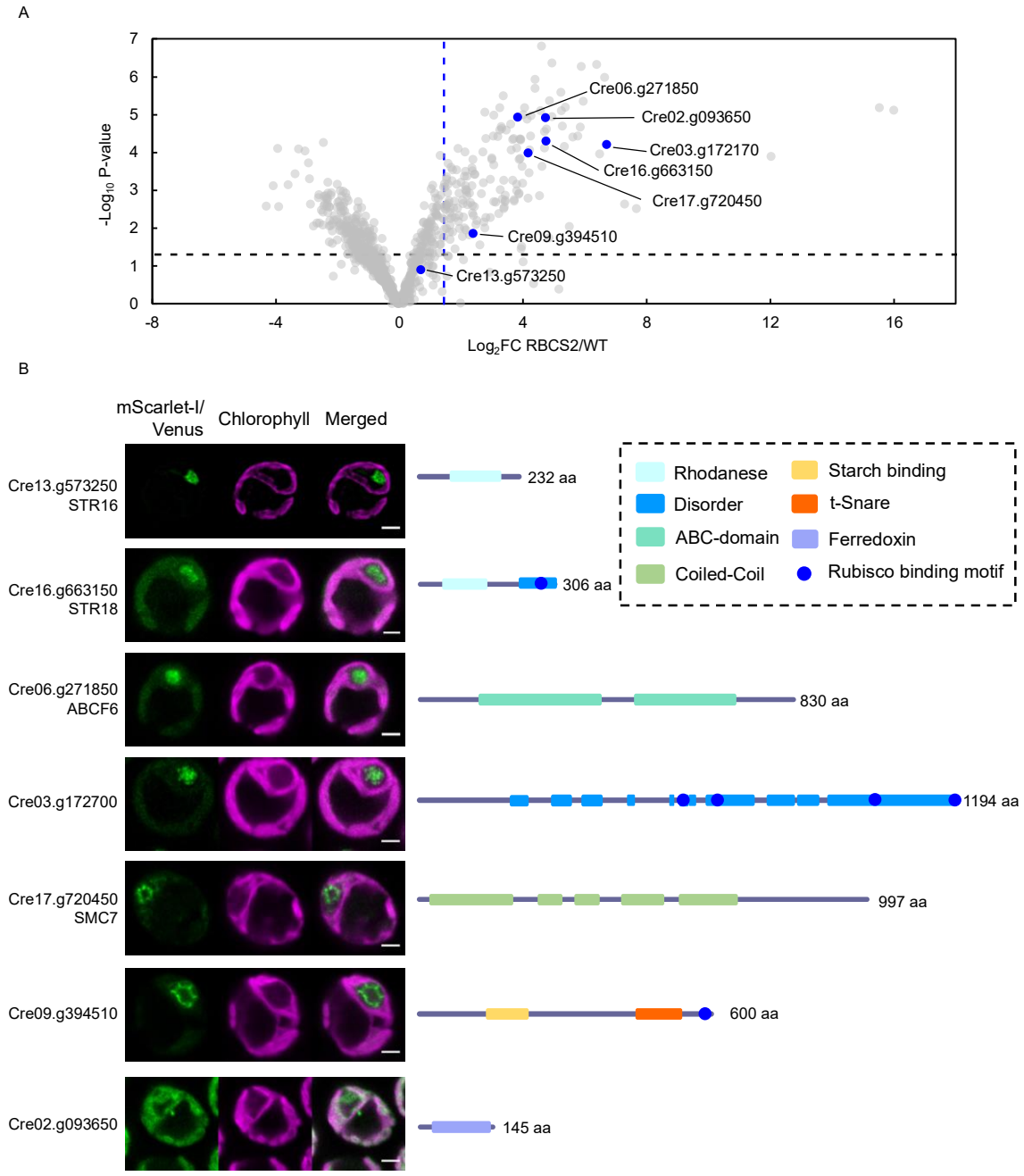



**Figure 4.** Proximity labeling identifies new pyrenoid proteins. **A)** The volcano plot in Fig. 2B was reproduced here to highlight the proteins that were chosen for localization (blue dots). **B)** Confocal imaging of the chosen proteins. The respective coding regions were cloned in-frame with Venus or mScarlet-I under their native promoter sequence. Green and magenta signals denote the fluorescence channel and chlorophyll autofluorescence, respectively. Scale bar is 2 µm. Schematic overview of structural prediction from PSI-pred and conserved domains are highlighted next to the confocal images. aa, amino acid.

STR16, STR18, and ABCF6 showed a localization pattern consistent to the pyrenoid matrix, which is supported by their lack of a predicted transmembrane or starch-binding domain.

STR16 and STR18 contain a rhodanese (thiosulfate sulfurtransferase) domain like the previously identified pyrenoid proteins CALCIUM SENSING RECEPTOR 1 (CAS1) and RBMP2. In contrast to STR16 and STR18, CAS1 and RBMP2 lack a critical cysteine in their active site and thus are presumably catalytically inactive. Rhodanese domains have been implicated in an array of functions including disulfide bond formation (Chng et al. 2012) and Fe-S cluster biosynthesis (Bonomi et al. 1977). The latter is particularly interesting as multiple proteins in the pyrenoid proxiome contain an Fe-S cluster domain such as the proteins encoded by Cre02.g093650, Cre08.g365692, and Cre15.g643600 (Supplemental Data Set 5). ABCF6 is predicted to be a member of the ABCF family, which has been shown to regulate protein translation via binding to ribosomes (Boël et al. 2014). The AlphaFold modeling of ABCF6 presents a structure consistent with its ABCF annotation, with the presence of the canonical arm and linker domains (Supplemental Fig. S6; UniProt Consortium 2021; (Jumper et al. 2021)). A fluorescently tagged version of the protein encoded by Cre03.g172700 formed distinct puncta within the pyrenoid matrix (Figs. 4B and S5) unlike the more homogenous signal observed for matrix proteins such as RBCS2. This subpyrenoid localization suggests that it may be associated with pyrenoid tubules. While PSI-pred structural prediction suggests that the protein encoded by Cre03.g172700 is predominantly disordered, AlphaFold prediction suggests that its C-terminus is composed of a central long alpha-helix surrounded by multiple shorter helices interspaced with disordered sequences that contain 4 RBMs (Figs. 4B and S6). The disordered sequences and RBMs combined might allow the protein encoded by Cre03.g172700 to act as a potential pyrenoid tether that recruits Rubisco to the pyrenoid tubules in a similar fashion as the previously hypothesized function of RBMP1 and RBMP2 (Meyer et al. 2020). Unlike the other proteins that localize to the pyrenoid matrix, STRUCTURAL MAINTENANCE OF CHROMOSOMES 7 (SMC7, encoded by Cre17.g720450) and the protein encoded by Cre09.g394510 are found at the edge of the pyrenoid matrix, with SMC7 forming discrete puncta surrounding the matrix while the protein encoded by Cre09.g394510 appears to line the starch–matrix interface. These proteins show a similar localization pattern as SAGA1, which occupies the starch–matrix–tubule interface. SMC7 lacks the signature ATP-binding and hinge domain important for its predicted function in chromatin condensation (Harvey et al. 2002) and only contains the conserved coiled-coil domain. This structure arrangement mirrors that of SAGA1 and SAGA2 (Itakura et al. 2019) that were also annotated as SMC components and suggests that SMC7 might function in a similar manner. The protein encoded by Cre09.g394510 contains a N-terminal CBM20 starch-binding domain and a t-SNARE domain at its C-terminus, the latter known to mediate vesicle fusion (Han et al. 2017). This observation suggests that the protein encoded by Cre09.g394510 may be involved in membrane remodelling of the pyrenoid tubules, as they are structurally reorganized from

thylakoid sheets to pyrenoid tubules as they traverse gaps within the starch sheath (Engel et al. 2015). Collectively, these new pyrenoid proteins represent exciting candidates for further investigation into pyrenoid formation and function.

**Changes in the pyrenoid proteome in response to CO2**

When a CCM is not required such as at high $CO_2$, the pyrenoid partially dissolves, with ~50% of Rubisco leaving the pyrenoid into the surrounding stroma (Borkhsenious et al. 1998). In addition, the starch sheath breaks down, and stromal starch content increases (Kuchitsu et al. 1988). However, at a transcriptional and protein abundance level, matrix pyrenoid proteins show a broad range of responses (Brueggeman et al. 2012; Fang et al. 2012; Arias et al. 2020). To explore if the pyrenoid composition changes in response to $CO_2$, we compared RBCS2-TurboID strains grown at high and low $CO_2$ (Fig. 5, A and B, and Supplemental Data Set 7). Many previously known pyrenoid proteins and proteins in our HC-pyrenoid proxiome were not preferentially enriched across $CO_2$ conditions, indicating that the vast majority of the pyrenoid proteome is not $CO_2$ responsive. However, a small number of proteins showed a >2 FC, with 20.5% (7/34) enriched at low $CO_2$ and 2.9% (1/34) enriched at high $CO_2$. Three of the low $CO_2$ enriched proteins, SAGA1, LCI9, and AMA3 (ALPHA AMYLASE 3), are associated with starch binding/metabolism. LCI9 was previously localized to the starch plate interfaces and proposed to play a role in starch metabolism (Mackinder et al. 2017). AMA3 is an alpha amylase also involved in starch hydrolysis (Gargouri et al. 2015), and mutants in SAGA1 have a severe starch structural defect (Itakura et al. 2019). Collectively, these results support the major remodelling of starch to form the starch sheath under low $CO_2$ conditions.

**Possible role of phase separation in protein recruitment to the pyrenoid matrix**

The deletion of EPYC1 leads to abolishment of the pyrenoid and CCM due to the failure to condense Rubisco into the pyrenoid (Mackinder et al. 2016). Confident that the HC-pyrenoid proxiome is labeled by RBCS2-TurboID, we explored how labeling changed when Rubisco was not condensed into the pyrenoid. To this end, we selected RBCS2-TurboID strains in WT and the *epyc1* mutant that accumulate the tagged protein to comparable levels (Fig. 5A). We determined that a large number of known pyrenoid proteins and proteins within the HC-pyrenoid proxiome are enriched in WT when compared to *epyc1* (Fig. 5C and Supplemental Data Set 7), indicating that phase separation either results in more efficient labeling or that phase separation is required for close proximity to Rubisco. However, a subset (11/25) of proteins in the HC-pyrenoid proxiome showed very little enrichment ($Log_2$ FC < 0.5) upon Rubisco condensation, suggesting that these proteins may directly interact with Rubisco independently of pyrenoid presence. The differences seen are unlikely due to changes in

protein abundance between WT and *epyc1* as these remained highly comparable (Supplemental Fig. S7). Unexpectedly, many proteins containing RBMs required the presence of the pyrenoid to be enriched (top right quadrant of Fig. 5C) indicating that the weak binding affinity (Kd ~3 mM; (He et al. 2020)) of RBMs may not be sufficient to allow Rubisco-RBM complex formation prior to Rubisco condensation by EPYC1.

Figure 5

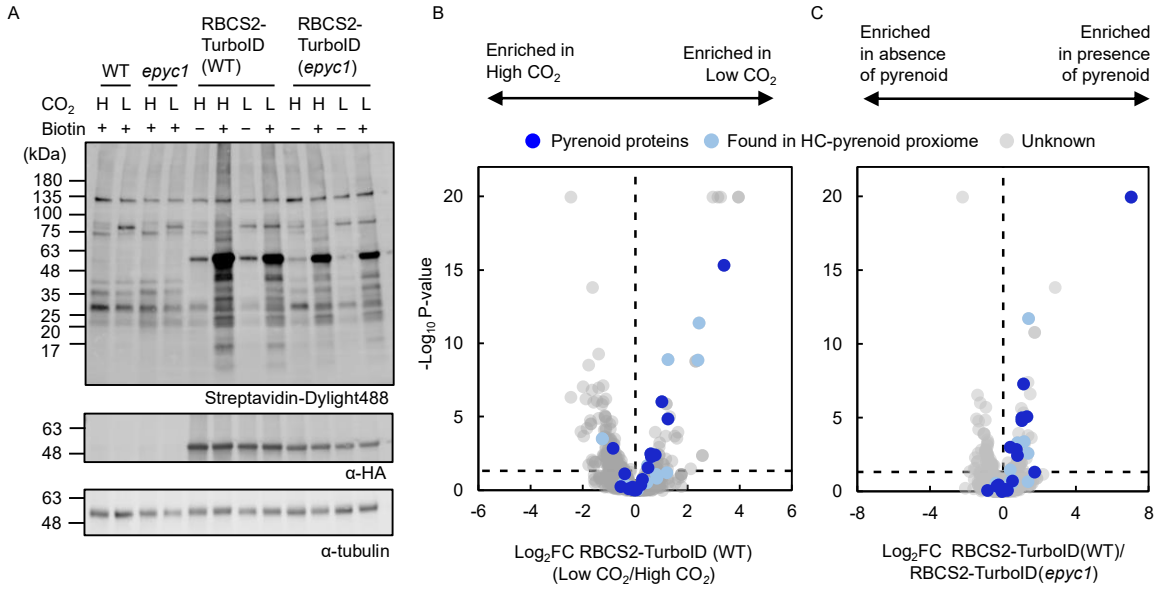



**Figure 5.** Proximity labelling suggests that the pyrenoid proteome has a subtle response to changes in $CO_2$ and phase separation. **A)** Protein labelling of RBCS2-TurboID in WT and epyc1 as well as their corresponding untagged background were tested under different $CO_2$ conditions. Respective strains were grown photoautotrophically and supplemented with 3% $CO_2$ (H) or 0.04% $CO_2$ (L). Harvested cells were incubated with 2.5 mM biotin for 4 h. Labeling was visualized by immunoblotting the whole cell lysate against streptavidin. Anti-HA was used to probe for RBCS2-TurboID abundance and anti-tubulin was used as a loading control. **B, C)** Volcano plots representing the $Log_2$ FC of RBCS2-TurboID in low $CO_2$ versus high $CO_2$ **B)** or RBCS2-TurboID in the WT background compared to RBCS2-TurboID in the *epyc1* mutant **C)**. Known pyrenoid proteins and the HC-pyrenoid proxiome are coloured dark blue and light blue, respectively, while unknowns are coloured in gray. Statistical significance for each pairwise comparison was calculated using the PEAKSQ method, a significance cutoff for $P < 0.05$ was used (horizontal dashed line).

## Discussion

We established TurboID-based proximity labelling in the chloroplast of the model green alga *C. reinhardtii*. Proximity labelling has proven powerful in unravelling a broad range of cellular functions and suborganelle composition in a diverse range of organisms including plants (Zhang et al. 2019; Mair and Bergmann 2022), diatoms (Turnšek et al. 2021), and

cyanobacteria (Dahlgren et al. 2021). However, until now, it had not been established in plastids or Chlamydomonas. In parallel to our work, 2 other studies give a snapshot of the diversity of possible applications of TurboID in both plant (Wurzinger et al. 2022) and algal plastids (our study and (Kreis et al. 2023)). The independently determined similar biotin concentrations and incubation time for labelling in the *Chlamydomonas* chloroplast by our work and the work by (Kreis et al. 2023)) highlight the reproducibility and robustness of the method.

Once established, we applied TurboID to determine the protein composition of the phase-separated pyrenoid. We identified a “pyrenoid proxiome” containing 84 proteins. A large number of previously localized pyrenoid proteins (67%) from the literature were present in our pyrenoid proxiome. However, it did miss several previously classified pyrenoid proteins. A deeper analysis of these missing proteins indicated that they were primarily located within specific pyrenoid subcompartments where they may remain inaccessible by the matrix generated biotin radicals. For example, CARBONIC ANHYDRASE 3 (CAH3), a pyrenoid tubule lumen protein, was not among the identified proteins in our proxiome most likely due to the limited penetration of biotin radicals across membranes (Rhee et al. 2013).

By including robust stromal controls for proteins that are adjacent to the pyrenoid but do not partition into the matrix, we established a “HC-pyrenoid proxiome” containing 30 proteins. This protein set excluded multiple proteins classified as pyrenoid proteins that are found at the pyrenoid periphery but do not partition into the matrix. These proteins included LCIB, LCI9, LCIC, and SBE3. These data along with the identification of nearly all known matrix proteins and proteins with RBMs that are at the matrix interface (i.e. SAGA1, BST4, and RBMP2) give us high confidence in this data set.

GO term enrichment analysis of the HC-pyrenoid proxiome and a broader analysis of the pyrenoid proxiome showed the enrichment of proteins in a small number of biochemical functions and pathways, suggesting that the pyrenoid plays additional roles to $CO_2$ concentration. Three groups that stood out were RNA-binding/translation proteins, Fe-S-containing proteins, and starch-binding proteins. Biomolecular condensates are regularly associated with RNA sequestration and processing (Banani et al. 2017). This association allows cells to respond in a timely manner in face of cellular stress. In Chlamydomonas, the photosynthetic machinery is translated at a specialized position adjacent to the pyrenoid called the translation zone (or T-zone; (Sun et al. 2019)). Under light and oxidative stress, the mRNA of the core photosystem II component PsbA becomes enriched within the pyrenoid matrix (Uniacke and Zerges 2008; Zhan et al. 2015), which suggests that the pyrenoid recruits RNA as a stress response. However, the molecular basis and function of

this mRNA sequestration remains unclear. In this study, we identified multiple RNA-associated proteins within the pyrenoid proxiome (proteins encoded by Cre10.g440050, Cre10.g435800, Cre09.g393358, and Cre13.g578650). We also localized a new ribosome-associated protein, ABCF6, to the pyrenoid. An Escherichia coli homolog of ABCF6, EttA, was demonstrated to prevent translation by its binding to 70S ribosomes in a ATP/ADP ratio-dependent manner (Boël et al. 2014). The localization of ABCF6 to the pyrenoid further supports a role for the pyrenoid in RNA metabolism, by either sequestering chloroplast ribosomes in the pyrenoid or partitioning ABCF6 away from chloroplast ribosomes under certain environmental conditions.

Fe-S protein assembly and activity is typically sensitive to molecular $O_2$ (Boyd et al. 2014). It was intriguing to see that the pyrenoid was enriched for both Fe-S assembly and Fe-S-containing proteins. A proposed, but unconfirmed, function of the pyrenoid to enhance $CO_2$ fixation is to minimize the presence of $O_2$ to increase the $CO_2$: $O_2$ ratio at the active site of Rubisco. A reduced $O_2$ environment could also favour other $O_2$-sensitive biological reactions. We found that the rhodanese domain-containing proteins STR16 and STR18 are localized to the pyrenoid; rhodanese domains are linked to the biogenesis of Fe-S clusters (Rydz et al. 2021). Pyrenoid localization might allow them to be shielded from the oxygenic environment outside the pyrenoid matrix, allowing these oxygen-sensitive reactions to be carried out. Alternatively, rhodanese has also been suggested to participate in reactive oxygen species (ROS) scavenging via the production of reactive sulfur species (Wang et al. 2021). Since ROS have also been found to drive pyrenoid formation (Neofotis et al. 2021), the presence of rhodanese domain-containing proteins in the pyrenoid suggests that the pyrenoid itself is involved with ROS metabolism or redox signaling.

The pyrenoid starch sheath is proposed to act as a diffusion barrier that limits $CO_2$ diffusion away from the pyrenoid matrix. Recent evidence has suggested that this matrix–starch association is critical for the organization of many pyrenoid components. The deletion of the gene encoding the starch-binding protein SAGA1 results in the formation of multiple pyrenoids with altered starch sheath and pyrenoid tubule morphology (Itakura et al. 2019). Additionally, the knockout of ISOAMYLASE 1 (ISA1) that abolishes the pyrenoid starch sheath results in the CCM-essential carbonic anhydrase LCIB to mislocalize as an aggregate at the basal region of the pyrenoid (Toyokawa et al. 2020), in contrast to its typical pyrenoid peripheral localization. Together, starch-binding proteins are crucial to the functioning of the pyrenoid in CCM-related functions. In this work, we localized an additional protein (encoded by Cre09.g394510 and contained a starch-binding CBM20 domain) to the pyrenoid. This protein contains an additional t-SNARE functional domain and has a similar domain arrangement to SAGA1 and LCI9, which also share a similar localization pattern

(Mackinder et al. 2017). Investigating the role of these proteins in pyrenoid structural organization and function may provide insights into pyrenoid assembly needed for future engineering of a functional pyrenoid into land plants (Adler et al. 2022).

Once we had determined a HC-pyrenoid proteome, we explored the change in the proxiome of Rubisco at low or high $CO_2$ and with (WT) or without (*epyc1*) phase separation. Surprisingly, most proteins appeared to be present in the pyrenoid under both $CO_2$ conditions, indicating that the core proteome of the pyrenoid is relatively stable. However, a subset involved in starch metabolism was predominantly enriched under low $CO_2$ when starch needs to be remodelled to form a $CO_2$ leakage barrier. By using the *epyc1* mutant, we explored how labelling by RBCS2-TurboID differs when Rubisco condensation into the pyrenoid is disrupted. Most HC-pyrenoid proxiome components were enriched by Rubisco condensation, indicating that they are brought into closer proximity upon pyrenoid formation. However, a subset showed very little change, suggesting that they may already be interacting with Rubisco independently of pyrenoid assembly. For both the high versus low $CO_2$ and WT versus *epyc1* comparisons, it should be noted that the 4 h incubation time of labelled strains could have led to translational changes resulting in compounding data between absolute protein amounts and partitioning into the pyrenoid. In addition, the partial dissolution of the pyrenoid during high $CO_2$ also resulted in a higher proportion of RBCS2-TurboID in the dilute phase. This in turn potentially increases labelling of proteins that have not yet partitioned into the pyrenoid. In the future, shorter labelling times may help further refine the pyrenoid proteome under varying conditions.

Proximity labeling has been underutilized for understanding phase-separated proteomes that are highly dynamic and thus are challenging to purify (Hubstenberger et al. 2017). The presence and exchange of bait proteins between the condensed phase and dilute phase might result in reduced specificity of RBCS2/EPYC1-TurboID over time and labelling outside of the condensate. To counteract this issue, we found that the use of abundant soluble controls that are excluded from the pyrenoid allowed the determination of a highly refined pyrenoid proteome. Future experiments using proximity labelling, specifically to determine the proteomes of biomolecular condensates, should include carefully chosen controls.

To make TurboID easily accessible for other laboratories using Chlamydomonas, we based our constructs on the MoClo Golden Gate cloning framework that enables TurboID to be used with a broad range of parts (Crozet et al. 2018) and easily fused to proteins that are already within this framework. To enable easy adoption of this powerful method, all developed vectors and lines were deposited at the Chlamydomonas Resource Center.

## Materials and methods

### Construction of APEX2/TurboID vectors in C. reinhardtii

Construction of APEX2/TurboID-expression cassettes for *Chlamydomonas* was designed using the MoClo system Chlamydomonas MoClo toolkit (Crozet et al. 2018). Golden Gate-compatible syntaxes were added to synthesize parts encoding the APEX2/TurboID enzyme and target proteins (RBCS2/EPYC1) or via PCR using CC-4533 genomic DNA for RPE1/PRK1 (see Supplemental Data Set 9 for all primer sequences used). Due to the low complexity and high repeat nature of EPYC1, the EPYC1 coding sequence was synthesized in 4 parts as a Level-1 construct, while the RBCS2 coding sequence was synthesized as 2 parts to avoid a detected sequence repeat. The APEX2 and TurboID tag sequences (Branon et al. 2018; Ganapathy et al. 2018) were codon optimized for Chlamydomonas (Nakamura et al. 2000) with the RBCS2i2 (Cre02.g120150) and LHCBM1i2 (Cre01.g066917) introns inserted at ~500-bp increments to improve protein production (Baier et al. 2018). The coding sequence of the tags was similarly synthesized as Level-1 parts. Together, the Level-1 and PCR-amplified target genes, APEX2/TurboID tag, and a sequence encoding a small flexible linker (GSGSTSGSGS) were assembled to a Level-0 product occupying the B3-B4 MoClo position using the pUAP1 backbone such that the target genes are expressed with the sequence encoding the enzyme tag at their 3′ end, bridged by the small flexible linker. The Level-1 cassette was then assembled using the target gene-TurboID/APEX2 fusion part, the PSAD promoter/terminator pair, and either a tandem HA/Flag tag epitope at the 3′ end of the construct for labelling experiments or a sequence encoding mCherry for localization. The resultant Level-2 expression module consists of the target gene-TurboID fusion cassette and an antibiotic resistance cassette for selection. To enable accessible use of TurboID-based proximity labelling in the Golden Gate cloning pipeline, the identical TurboID coding sequence with the flexible linker was also cloned into a Level-0 part occupying the B4 MoClo position. Sequences for all developed vectors are in Supplemental Data Set 1. All vectors and strains are deposited at the Chlamydomonas Resource Center (https://www.chlamycollection.org).

### Chlamydomonas growth and transformation

Chlamydomonas cultures were maintained on TAP medium with revised Hunter's trace elements (Kropat et al. 2011). For biotin labeling experiments, cells were grown photoautotrophically in Tris phosphate (TP) medium at ~21 °C under LED lights (Valoya C65 LEDs with AP673L spectrum) at ~50 µmol photons $m^{-2}$ $s^{-1}$. Assembled plasmids were linearized with I-SceI (for fluorescent tagging plasmids) or BsaI (for proximity labeling

plasmids) and transformed into Chlamydomonas via electroporation according to (Mackinder et al. 2017).

**Protein extraction and immunoblotting**

For immunoblotting, cells were grown photoautotrophically to mid-log phase and were harvested by centrifugation 17,900 × g for 5 min at 4 °C. Cell pellets were resuspended in lysis buffer (25 mM Tris–HCl pH 7.4, 300 mM NaCl, 1 mM DTT, 5 mM MgCl2, 0.1 mM PMSF, 1× EDTA-free protease inhibitor [Roche], 0.1% [w/v] SDS, 0.5% [w/v] deoxycholic acid, and 1% [v/v] Triton X-100) before snap-freezing in liquid nitrogen. The cell suspensions were lysed by 5 freeze/thaw cycles and centrifuged at 17,900 × g for 10 min at 4 °C. The resulting supernatants were used as protein samples in later experiments and stored at −70 °C if not used immediately. For immunoblotting, boiled protein samples were resolved by SDS–PAGE and transferred to a PVDF membrane via a semidry transfer system. Membrane was blocked with 3% (w/v) BSA in Tris-buffered saline with 0.1% (v/v) Tween 20 (TBST) and probed with antibodies accordingly. Antibodies were diluted in TBST as follows: Streptavidin Dylight-488 conjugate (1:4,000, Fisher Scientific #21832); anti-HA (1:1,000, Fisher Scientific 26183); anti-Flag (1:1,000, Sigma #F1804); and anti-tubulin (1:2,000, Sigma #T6074).

**Biotin labeling and streptavidin affinity purification**

All 3 TurboID-labeling experiments were performed similarly. The starter culture of TurboID expression strains and WT were grown to mid-log phase in TAP medium. They were used to inoculate 400 mL of TP medium supplied with elevated $CO_2$ (3% [v/v] $CO_2$) until mid-log phase and then transferred to air-level $CO_2$ (0.04% [v/v] $CO_2$) for ~2 d or maintained at 3% (v/v) $CO_2$ as indicated. Cells were harvested by centrifugation 1,500 × g for 5 min at room temperature. They were then resuspended in fresh TP medium in a 6-well cell culture plate to an $OD_{750}$ of 2.5. Then, 100 mM biotin stock in DMSO was added to the cell suspension to a final concentration of 2.5 mM to initiate the labelling reaction. Biotin labelling was allowed to proceed for 1 to 8 h in the pilot experiment or for 4 h in the later experiments on an orbital shaker. Biotin-labelled cells were harvested by centrifugation 21,300 × g, 2 min at 4 °C and rinsed 3 times with ice-cold TP medium. Cell pellets were snap frozen in liquid nitrogen and stored at −70 °C until streptavidin affinity purification.

For APEX2 labelling, the RBCS2-APEX2 expression cells were grown and harvested to an $OD_{750}$ of 2.5 as mentioned above. Biotin-phenol at a final concentration of 2.5 mM was added to the harvested cell suspension from a 250 mM biotin-phenol stock in DMSO. Biotin-phenol incubation was performed for 2 h on an orbital shaker at 20, 30, or 37 °C. The $H_2O_2$ activator at 2 mM concentration was spiked into the suspension to initiate biotin labelling for 2 min. The reaction was then quenched by addition of an ice-cold quencher solution (10 mM

sodium ascorbate, 5 mM Trolox, and 10 mM sodium azide in PBS, pH 7.4) and pelleted by centrifugation 21,300 × g for 1 min at 4 °C and stored at −70 °C until streptavidin affinity purification.

Protein extraction was carried out as described above. Prior to streptavidin affinity pull-down, free biotin was removed from protein samples using a Zeba Spin Desalting column (#89891, Thermo Fisher) using lysis buffer. To determine protein concentration, a small aliquot (50 µL) of the desalted protein was diluted 10 times in water, and concentration was measured using a Pierce BCA protein assay kit (#23225, Thermo Fisher) as per the manufacturer's instructions. For streptavidin affinity purification, a total of 1.75 mg of protein was used with 50 µL of Pierce Streptavidin Magnetic Beads (88816; Thermo Fisher) equilibrated with lysis buffer. The bead suspension was incubated at 4 °C overnight on a rotor wheel. Beads were then washed twice with lysis buffer for 5 min: once with 1 M KCl for 2 min; once with 0.1 M $NaCO_3$ for 1 min; once with 4 M urea in 50 mM triethylammonium bicarbonate, pH 8.5 (TEAB) for 1 min; once with 6 M urea in 50 mM TEAB for 1 min; and twice with 50 mM TEAB buffer for 5 min. Washed beads were frozen at −70 °C until submitted for mass spectrometry.

**LC-MS/MS and analysis of APEX2 and TurboID pilot studies**

**APEX2 digestion**

For the APEX2 experiments, streptavidin beads were eluted by boiling with 2× Laemmli loading buffer (Biorad, 161 to 0737) containing 20 mM DTT and 2 mM biotin. The eluate was then run on a 4% to 15% Tris-glycine gel (Biorad, #4561084) for 30 min at 50 V. Gel slices were then fixed according to (Mackinder et al. 2017)). In-gel tryptic digestion was performed after reduction with 10 mM dithioerythritol and 50 mM S-carbamidomethylation with iodoacetamide. Gel pieces were washed 2 times with aqueous 50% (v/v) acetonitrile containing 25 mM ammonium bicarbonate and then once with acetonitrile and dried in a vacuum concentrator for 20 min. A 500-ng aliquot of sequencing-grade trypsin (Promega) was added prior to incubation at 37 °C for 16 h.

**TurboID digestion**

For the TurboID pilot experiment, on-bead digestion was performed after reduction with 10 mM tris(2-carboxyethyl)phosphine and alkylation with 10 mM iodoacetamide in 50 mM TEAB containing 0.01% (w/v) ProteaseMAX surfactant (Promega). A 500-ng aliquot of sequencing-grade trypsin (Promega) was added prior to incubation at 37 °C for 16 h.

**LC-MS/MS acquisition of APEX2 and TurboID pilot experiments**

Resulting peptides were resuspended in aqueous 0.1% (v/v) trifluoroacetic acid and then loaded onto an mClass nanoflow UPLC system (Waters) equipped with a nanoEaze M/Z Symmetry 100-Å C18 and 5-µm trap column (180 µm × 20 mm, Waters) and a PepMap, 2-µm, 100-Å, and C18 EasyNano nanocapillary column (75 mm × 500 mm, Thermo). The trap wash solvent was aqueous 0.05% (v/v) trifluoroacetic acid and the trapping flow rate was 15 µL/min. The trap was washed for 5 min before switching the flow to the capillary column. Separation used gradient elution of 2 solvents: solvent A, aqueous 0.1% (v/v) formic acid; and solvent B, acetonitrile containing 0.1% (v/v) formic acid. The flow rate for the capillary column was 300 nL/min, and the column temperature was 40 °C. The linear multistep gradient profile was 3% to 10% B over 7 min, 10% to 35% B over 80 min, and 35% to 99% B over 10 min and then proceeded to wash with 99% solvent B for 8 min. The column was returned to initial conditions and reequilibrated for 15 min before subsequent injections. The nanoLC system was interfaced with an Orbitrap Fusion Tribrid mass spectrometer (Thermo) with an EasyNano ionization source (Thermo). Positive ESI-MS and $MS^2$ spectra were acquired using Xcalibur software (version 4.0, Thermo). Instrument source settings were ion spray voltage, 1,900 V; sweep gas, 0 Arb; ion transfer tube temperature; and 275 °C. $MS^1$ spectra were acquired in the Orbitrap with the following: 120,000 resolution, scan range: m/z 375 to 1,500; AGC target, $4e^5$; and max fill time, 100 ms. Data-dependent acquisition was performed in top speed mode using a 1-s cycle, selecting the most intense precursors with charge states >1. Easy-IC was used for internal calibration. Dynamic exclusion was performed for 50-s postprecursor selection, and a minimum threshold for fragmentation was set to $5 \times 10^3$. $MS^2$ spectra were acquired in the linear ion trap with the following: scan rate, turbo; quadrupole isolation, 1.6 m/z; activation type, HCD; activation energy, 32%; AGC target, $5 \times 10^3$; first mass, 110 m/z; and max fill time, 100 ms. Acquisitions were arranged by Xcalibur to inject ions for all available parallelizable time.

**Spectral counting APEX2**

Peak lists in Thermo.raw format were converted to .mgf using MSConvert (version 3.0, ProteoWizard) before submitting to database searching against 19,716 Chlamydomonas protein sequences appended with common proteomic contaminants. Mascot Daemon (version 2.6.0, Matrix Science) was used to submit the search to a locally running copy of the Mascot program (Matrix Science Ltd., version 2.7.0). Mascot was searched with a fragment ion mass tolerance of 0.50 D and a parent ion tolerance of 3.0 ppm. O-124 of pyrrolysine, j-16 of leucine/isoleucine indecision, and carbamidomethyl of cysteine were specified in Mascot as fixed modifications. Oxidation of methionine was specified in Mascot as a variable modification. Scaffold (version Scaffold_5.2.0, Proteome Software Inc., Portland, OR, USA) was used to validate MS/MS-based peptide and protein identifications.

Peptide identifications were accepted if they could be established at >84.0% probability to achieve a false discovery rate (FDR) of 1.0% or less by the percolator posterior error probability calculation. Protein identifications were accepted if they could be established at >6.0% probability to achieve an FDR of <1.0% and contained at least 2 identified peptides. Quantitative value of total spectra was used to calculate the $Log_2$ FC between RBCS2-APEX2 and WT samples, and the Student's t-test derived P-value was $-Log_{10}$ transformed before presented.

**Precursor intensity-based relative quantification TurboID pilot**

Peak lists in .raw format were imported into Progenesis QI (version 2.2., Waters) and LC-MS runs aligned to the common sample pool. Precursor ion intensities were normalized against total intensity for each acquisition. A combined peak list was exported in .mgf format for database searching against 19,716 Chlamydomonas protein sequences appended with common proteomic contaminants. Mascot Daemon (version 2.6.0, Matrix Science) was used to submit the search to a locally running copy of the Mascot program (Matrix Science Ltd., version 2.7.0). Search criteria specified were as follows: enzyme, trypsin; max missed cleavages, 1; fixed modifications, carbamidomethyl (C); variable modifications, oxidation (M); peptide tolerance, 3 ppm; MS/MS tolerance, 0.5 D; and instrument, ESI-TRAP. Peptide identifications were passed through the percolator algorithm to achieve a 1% FDR assessed against a reverse database and individual matches filtered to require minimum expect score of 0.05. The Mascot .XML result file was imported into Progenesis QI and peptide identifications associated with precursor peak areas matched between runs. Relative protein abundance was calculated using precursor ion areas from nonconflicting unique peptides. Accepted protein quantifications were set to require a minimum of 2 unique peptide sequences. Missing values were then replaced by the minimal value detected from each bait. The FC in the RBCS2-TurboID versus WT comparison was calculated on the sum of relative protein abundance at all time points and was $Log_2$ transformed. Statistical testing was performed in Progenesis QI from ArcSinh-normalized peptide abundances and the ANOVA-derived P-values was $-Log_{10}$ transformed and presented.

**LC-MS/MS and analysis of TMT-labeled TurboID experiments**

**TurboID digestion and TMT labelling**

For the TurboID experiments in Figs. 3 and 4, on-bead digestion was performed after reduction with 10 mM tris(2-carboxyethyl)phosphine and alkylation with 50 mM methyl methanethiosulfonatein in 50 mM TEAB. A 500-ng aliquot of sequencing-grade trypsin (Promega) was added prior to incubation at 37 °C for 16 h. Postdigestion, the peptide-containing supernatants were removed from the beads for TMT labeling. Peptides were

labelled with TMTPro 16-plex reagents (Thermo Fisher) as detailed in the manufacturer's protocol. Postlabeling samples were combined and dried in a vacuum concentrator before reconstituting in 100-mL $H_2O$.

**LC-MS/MS acquisition of TMT-labeled TurboID experiment**

Peptides were fractionated by high pH reversed phase C18 HPLC. Samples were loaded onto an Agilent 1260 II HPLC system equipped with a Waters XBridge 3.5-µm, C18 column (2.1 mm × 150 mm, Thermo). Separation used gradient elution of 2 solvents: solvent A, aqueous 0.1% (v/v) ammonium hydroxide; and solvent B, acetonitrile containing 0.1% (v/v) ammonium hydroxide. The flow rate for the capillary column was 200 mL/min, and the column temperature was 40 °C. The linear multistep gradient profile for the elution was 5% to 35% B over 20 min and 35% to 80% B over 5 min; the gradient was followed by washing with 80% (v/v) solvent B for 5 min before returning to initial conditions and reequilibrating for 7 min prior to subsequent injections. Eluate was collected at 1-min intervals into LoBind Eppendorf tubes. Peptide elution was monitored by UV absorbance at 215 and 280 nm. Fractions were pooled across the UV elution profile to give 12 fractions for LC-MS/MS acquisition. Peptide fractions were dried in a vacuum concentrator before reconstituting in 20 mL aqueous 0.1% (v/v) trifluoroacetic acid.

TMT-labeled peptides fractions were loaded onto an mClass nanoflow UPLC system (Waters) equipped with a nanoEaze M/Z Symmetry 100-Å C18 and 5-µm trap column (180 µm × 20 mm, Waters) and a PepMap, 2-µm, 100-Å, and C18 EasyNano nanocapillary column (75 mm × 500 mm, Thermo). The trap wash solvent was aqueous 0.05% (v/v) trifluoroacetic acid, and the trapping flow rate was 15 µL/min. The trap was washed for 5 min before switching the flow to the capillary column. Separation used gradient elution of 2 solvents: solvent A, aqueous 0.1% (v/v) formic acid; and solvent B, acetonitrile containing 0.1% (v/v) formic acid. The flow rate for the capillary column was 330 nL/min, and the column temperature was 40 °C. The linear multistep gradient profile was 2.5% to 10% B over 10 min, 10% to 35% B over 75 min, and 35% to 99% B over 15 min before proceeded to wash with 99% solvent B for 5 min. The column was returned to initial conditions and reequilibrated for 15 min before subsequent injections. The nanoLC system was interfaced to an Orbitrap Fusion hybrid mass spectrometer (Thermo) with an EasyNano ionization source (Thermo). Positive ESI-MS, MS2, and MS3 spectra were acquired using Xcalibur software (version 4.0, Thermo). Instrument source settings were as follows: ion spray voltage, 2,100 V; sweep gas, 0 Arb; and ion transfer tube temperature, 275 °C. MS1 spectra were acquired in the Orbitrap with: 120,000 resolution, scan range; m/z 380 to 1,500; AGC target, 2 × 105; and max fill time, 50 ms. Data-dependent acquisition was performed in top speed mode using a 4-s cycle, selecting the most intense precursors with charge states 2 to

6. Dynamic exclusion was performed for 50-s postprecursor selection, and a minimum threshold for fragmentation was set at 3 × 104. MS2 spectra were acquired in the linear ion trap with: scan rate, turbo; quadrupole isolation, 1.2 m/z; activation type, CID; activation energy, 35%; AGC target, 1 × 104; first mass, 120 m/z; and max fill time, 35 ms. MS3 spectra were acquired in multinotch synchronous precursor mode (SPS3), selecting the 5 most intense MS2 fragment ions between 400 and 1,000 m/z. SPS3 spectra were measured in the Orbitrap mass analyzer using 50,000 resolution; quadrupole isolation, 1 m/z; activation type, HCD; collision energy, 65%; scan range, m/z 110 to 500; AGC target, 4 × 105; and max fill time, 10 ms. Acquisitions were arranged by Xcalibur to inject ions for all available parallelizable time.

**Protein identification and TMT label intensity quantification**

Peak lists in .raw format were imported into PEAKS StudioX Pro (version 10.6 Bioinformatics Solutions Inc.) for peak picking, database searching, and relative quantification. MS2 peak lists were searched against 19,716 Chlamydomonas protein sequences appended with common proteomic contaminants. Search criteria specified were as follows: enzyme, trypsin; max missed cleavages, 1; fixed modifications, TMT16plex (K- and N-term peptide); variable modifications, oxidation (M); peptide tolerance, 3 ppm; MS/MS tolerance, 0.5 D; and instrument, ESI-TRAP. Peptide identifications were filtered to achieve a 1% peptide spectral match FDR as assessed empirically against a reversed database search. Protein identifications were further filtered to require a minimum of 2 unique peptides per protein. TMT reporter ion intensities acting as markers of relative intersample peptide abundance were extracted from $MS^3$ spectra for quantitative comparison. Protein level quantification significance used ANOVA for multiway comparison and the PEAKSQ significance test for pairwise comparisons. In both cases, the null hypothesis was that individual protein abundance was equal between groups. Normalization of label intensity was then carried out using the global ratio derived from total intensity of all labels. The FCs between comparison groups were calculated based on their normalized TMT reporter ion intensities. Proteins that were not detected in all replicates for an individual bait were removed from calculation. Missing values were then replaced by the minimal value detected from each bait. Significance was determined via PEAKSQ test represented as −Log10P value.

**Recombineering cloning for localization**

Cloning of fluorescent protein-tagged constructs was performed as previously described (Emrich-Mills et al. 2021). Briefly, homology arms to target genes at the 5′ of the native promoter and 3′ UTR were added to destination vectors via PCR. Homology arms of Cre13.g573250 were cloned into the pLM162-mScarlet-I backbone. Homology arms of

Cre16.g663150, Cre06.g271850, Cre03.g172700, Cre17.g720450, Cre09.g394510, and Cre02.g093650 were cloned into the pLM099-Venus backbone. Amplified backbones were transformed by electroporation into E. coli containing a bacterial artificial chromosome and RecA vector, which drives the recombination event. The resulting plasmids were selected on LB agar plates containing kanamycin and junctions confirmed by sequencing.

**Imaging of fluorescently tagged lines**

For imaging of fluorescently tagged lines, photoautotrophically grown cells were immobilized on 1.5% (w/v) low-melting point agarose in TP medium. Indirect immunofluorescence of RBCS2-APEX2 was performed according to (Uniacke et al. 2011) with the following modifications: cells were fixed with 3.7% (w/v) formaldehyde solution in PBS for 30 min at room temperature. Anti-Flag antibody (F1804; Sigma-Aldrich) at 1:1,000 dilution in PBS containing 1% (w/v) BSA was used as primary antibody. Anti-Mouse Alexa Fluor plus 555 (A32727; Invitrogen) was used as the secondary antibody at 1:1,000 dilution. Labeled cells were then kept in the dark prior to imaging. Images were taken using a Zeiss LSM880 microscope with the Airyscan module or a Zeiss Elyra7 Lattice SIM. Excitation and emission filters of fluorophore and chlorophyll autofluorescence were set as follows: mVenus (excitation: 514 nm; emission: 520 to 550 nm); chlorophyll (excitation: 633; emission: 610 to 650 nm); and mCherry/mScarlet-I/Alexa Fluor plus 555 (excitation: 561 nm; emission: 580 to 600 nm).

**Amplex Red assay**

Amplex UltraRed assay for RBCS2-APEX2 peroxidase activity was carried out according to the manufacturer's manual. Briefly, Amplex Red reagent (Fisher Scientific; Invitrogen Amplex UltraRed Reagent #10737474) was dissolved in DMSO to a 10 mM stock. RBCS2-APEX2 and the untagged WT strains were grown photoautotrophically and split into triplicates. Cells were then chilled on ice for 5 min before resuspending in 200 µL of reaction buffer (50 M Amplex Red, 2 mM $H_2O_2$ in PBS, and pH 7.4). The reaction was carried out on ice for 15 min. Resorufin fluorescence measurement was performed using a Clariostar Plus Microplate reader using the following excitation and emission settings: resorufin (excitation: 535 to 555 nm; emission: 580 to 620 nm) and chlorophyll autofluorescence (excitation: 610 to 630 nm; emission: 660 to 695 nm).

**Accession numbers**

Sequence data from this article can be found in Phytozome, the Plant Comparative Genomics portal of the Department of Energy's Joint Genome Institute, under the following accession numbers: Cre02.g120150: RBCS2; Cre10.g436550: EPYC1; Cre12.g511900: RPE1; Cre12.g554800: PRK1; Cre13.g573250: STR16; Cre16.g663150: STR18;

Cre06.g271850: ABCF6; Cre03.g172700; Cre17.g720450: SMC7; Cre09.g394510; and Cre02.g093650. Proteomic data are deposited in MassIVE: https://doi.org/doi:10.25345/C5057D306 with ProteomeXchange identifier: PXD041970.


## Acknowledgments

The authors would like to thank the University of York Biosciences Technology Facility for confocal microscopy access and support and Masa Onishi for constructive discussions during the establishment of TurboID.


## Author contributions

L.C.M.M. guided and supervised the project; C.S.L. designed and performed the biotin labeling experiments. A.D. performed LC-MS/MS analysis. C.S.L. and P.G. performed the fluorescent protein tagging and confocal imaging. C.S.L. and L.C.M.M. analyzed the data and wrote the manuscript with contributions from P.G., G.H.T., and A.D. All authors discussed the results and commented on the manuscript.

## Supplemental data

The following materials are available in the online version of this article.

Supplemental Figure S1. APEX2 does not efficiently label pyrenoid proteins in the Chlamydomonas chloroplast.

Supplemental Figure S2. Screening strains for TurboID accumulation and activity.

Supplemental Figure S3. Confocal imaging of mCherry tagged TurboID strains.

Supplemental Figure S4. Enrichment of differentially localized proteins using different controls.

Supplemental Figure S5. Additional images of the TurboID identified pyrenoid proteins.

Supplemental Figure S6. AlphaFold-modeled structures for the proteins encoded by Cre06.g271850 (ABCF6) and Cre03.g172700.

Supplemental Figure S7. Comparison of protein abundance between WT and epyc1 background strains expressing RBCS2-TurboID after streptavidin affinity purification.

Supplemental Data Set 1. A list of all plasmids used and developed in this study.

Supplemental Data Set 2. Pilot mass spectrometry analysis of RBCS2-TurboID samples against WT controls.

Supplemental Data Set 3. Localization of proteins based on fluorescence tagging/immunofluorescence obtained from the literature.

Supplemental Data Set 4. TMT mass spectrometry analysis of pyrenoid baits against stromal and WT controls.

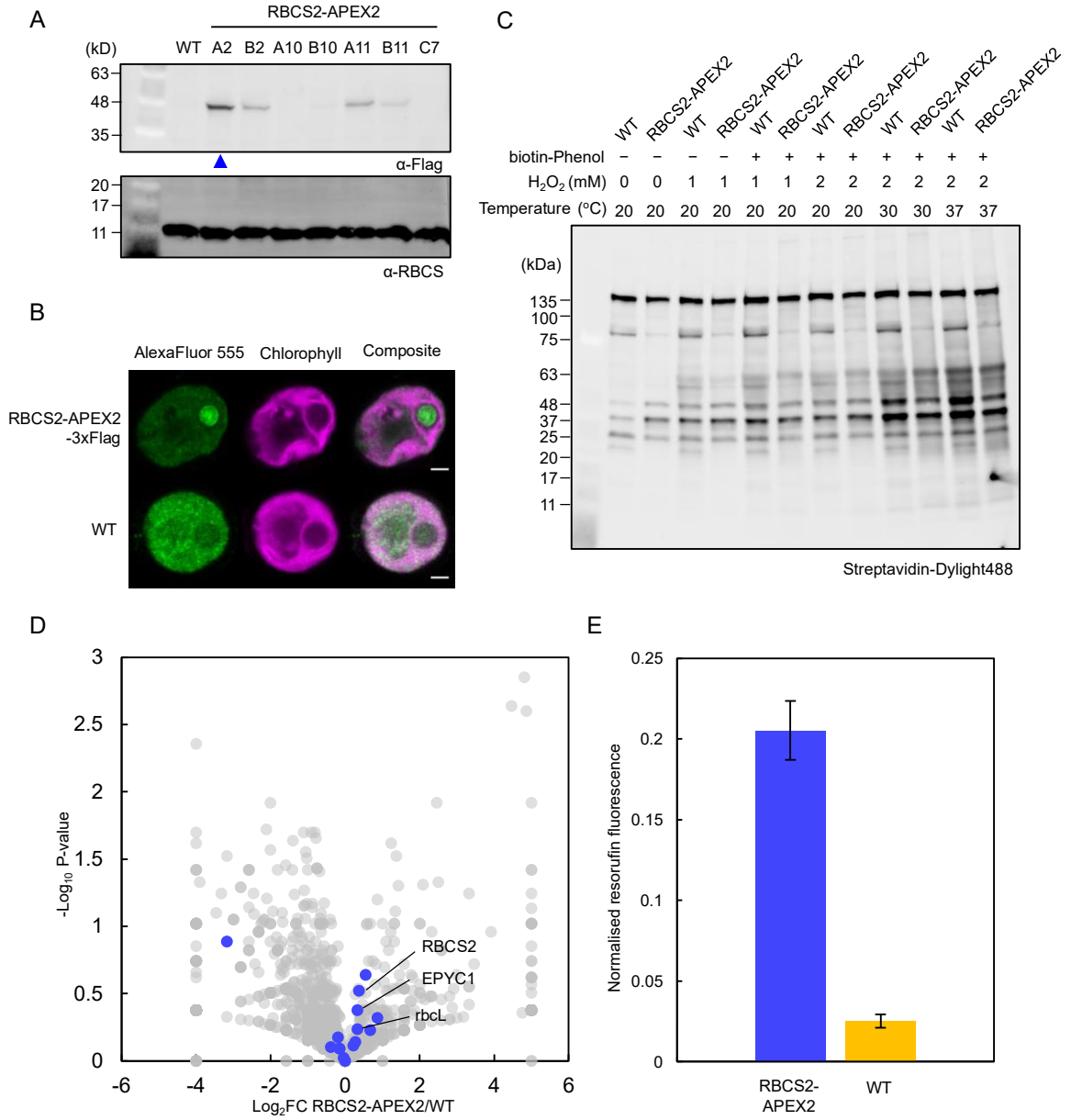
Supplemental Data Set 5. Pyrenoid proxiome.

Supplemental Data Set 6. HC-pyrenoid proxiome.

Supplemental Data Set 7. TMT mass spectrometry analysis of RBCS2-TurboID at different CO2 levels and in the epyc1 mutant.

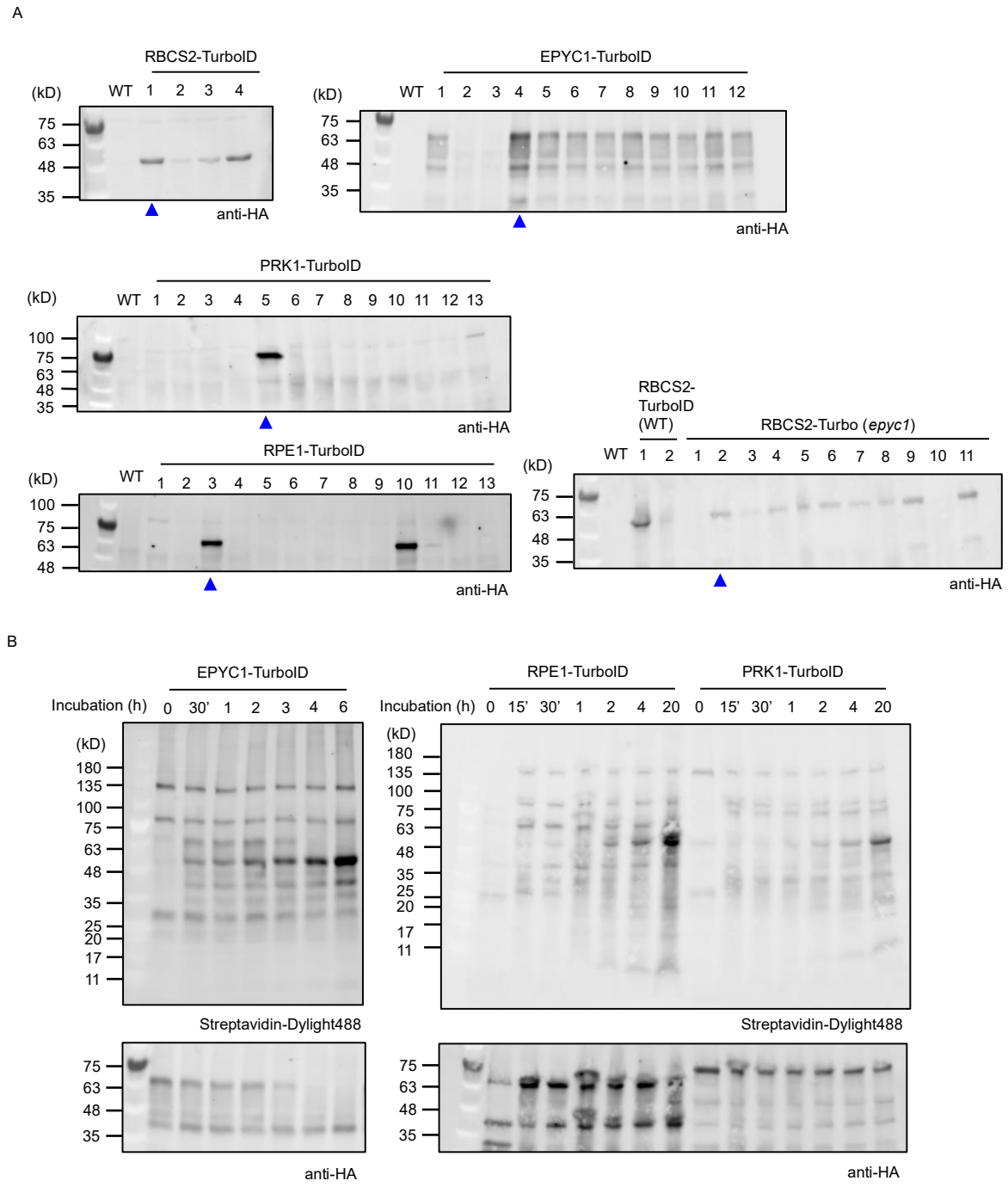
Supplemental Data Set 8. The statistical summary of results shown in Fig. 2E.

Supplemental Data Set 9. Oligo sequences used.

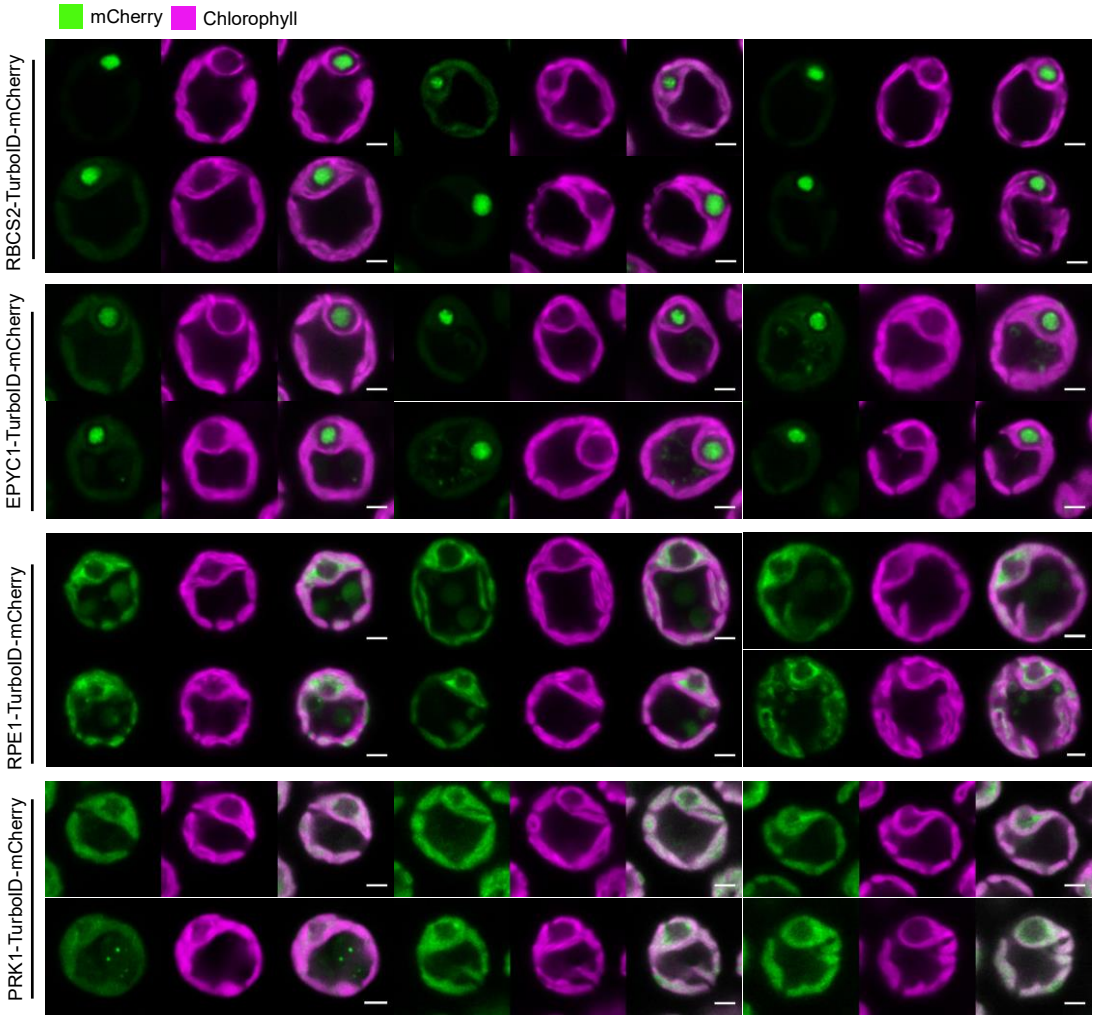



**Supplemental Figure S1.** APEX2 does not efficiently label pyrenoid proteins in the Chlamydomonas chloroplast. (Supports Figure 1.) **A,** Abundance of RBCS2-APEX2 in *Chlamydomonas* CC-4533 transformants was verified by immunoblotting whole cell lysates with anti-Flag antibody. Anti-RBCS was used as a loading control. Blue arrow denotes the strain chosen for later labeling experiments. **B,** Localization of the RBCS2-APEX2 fusion protein as determined by immunofluorescence using anti-Flag antibody. Green and Magenta signals denote Alexa Fluor 555 and chlorophyll fluorescence, respectively. Non-specific binding of the anti-Flag antibody results in background signal in WT cells. Scale bars, 2 µm. **C,** Labelling efficiency of RBCS2-APEX2 was tested by incubating expressing strains with 2.5 mM biotin-phenol substrate for 2 h. $H_2O_2$ activator was added at different

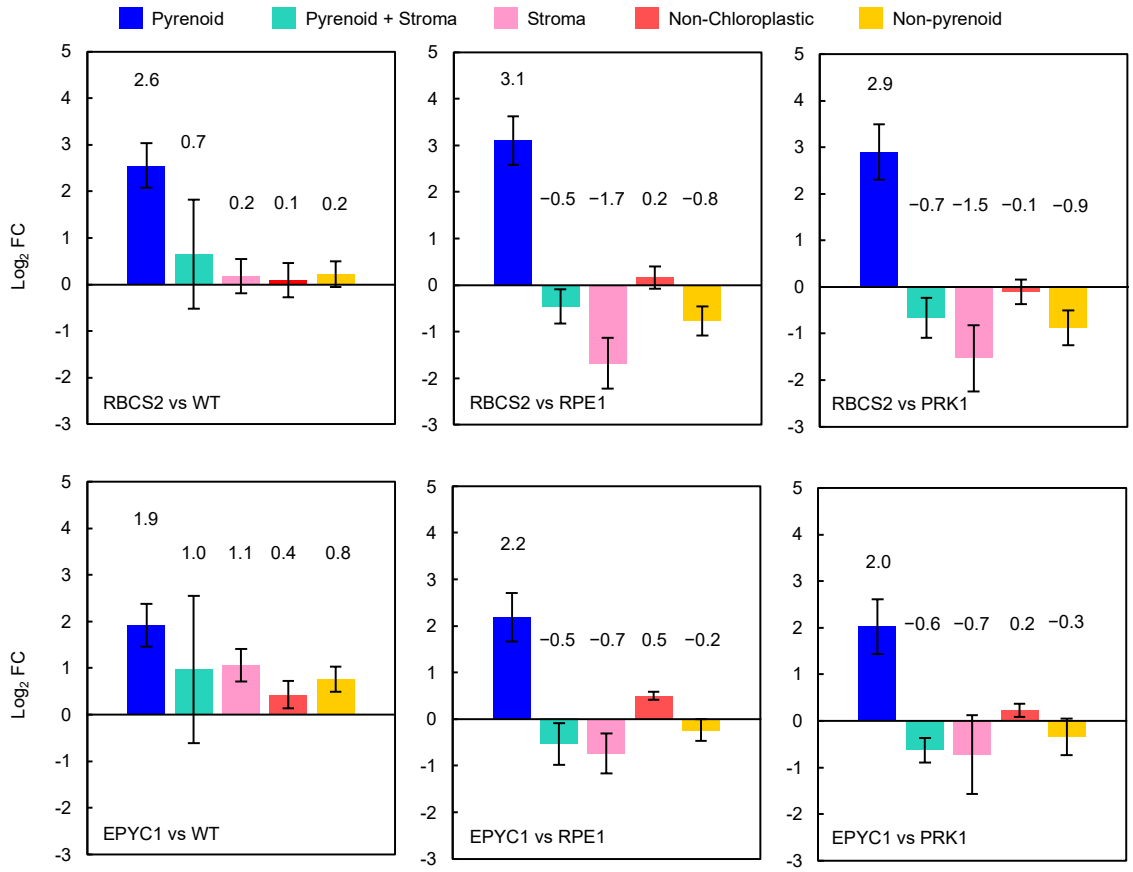
concentrations (0–2 mM) and activation was carried out at a range of temperatures (20–37°C). Biotin labeling was visualized by immunoblotting whole cell lysates against streptavidin. **D,** Volcano plot representing the $Log_2$ fold change of spectral counts from RBCS2-APEX2 compared to untagged WT labeling experiments. Gray and dark blue dots represent detected proteins and known pyrenoid proteins, respectively. Significance was determined via t-test. Proteins detected only in WT are set to –4 $Log_2$ FC and proteins only detected in RBCS2-APEX2 are set to 5 $Log_2$ FC. **E,** Amplex-red Assay was carried out to determine the peroxidase activity of RBCS2-APEX2 strains. Untagged WT and RBCS2-APEX2 strains were incubated with Amplex-red reagent and activated with $H_2O_2$ at 1 mM. Resorufin fluorescence emission (excitation: 535–555 nm; emission: 580–620 nm) was normalized against chlorophyll autofluorescence (excitation: 610–630 nm; emission: 660–695 nm) (n=6). Error bars indicate standard error.



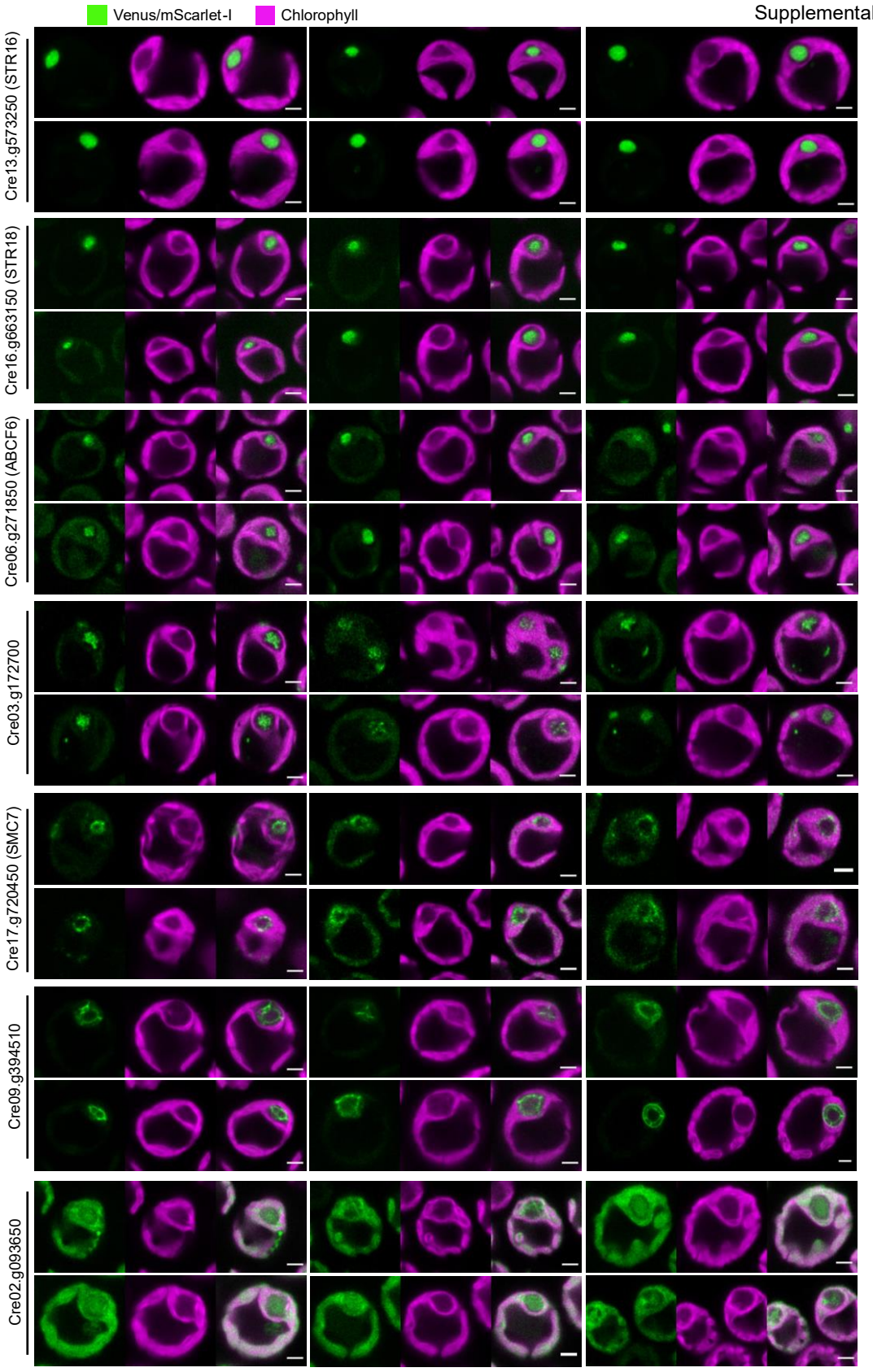

**Supplemental Figure S2**. Screening strains for TurboID accumulation and activity. (Supports Figure 1, 2, 3 and 5.) A, *Chlamydomonas* cells transformed with *TurboID*-tag plasmids were grown in TAP medium. Protein production was assessed via immunoblotting whole cell lysate from picked strains with anti-HA antibody. Blue arrows denote the strains used in labelling experiments in Figure 1, 2, 3 and 5. B, Labelling activity of the various TurboID-tagged lines was assessed by incubating them in 2.5 mM Biotin for a range of durations (0–20 h). Anti-HA antibody was used to assess protein accumulation.




**Supplemental Figure S3.** Confocal imaging of mCherry tagged TurboID strains. (Supports Figure 1 and 3.) RBCS2, EPYC1, RPE1 and PRK1 were cloned in-frame with a TurboID-mCherry sequence at their 3′ ends. Images shown are representative images from two independently transformed strains. Green and magenta signals denote mCherry fluorescence and autofluorescence from chlorophyll, respectively. Scale bars, 2 μm.

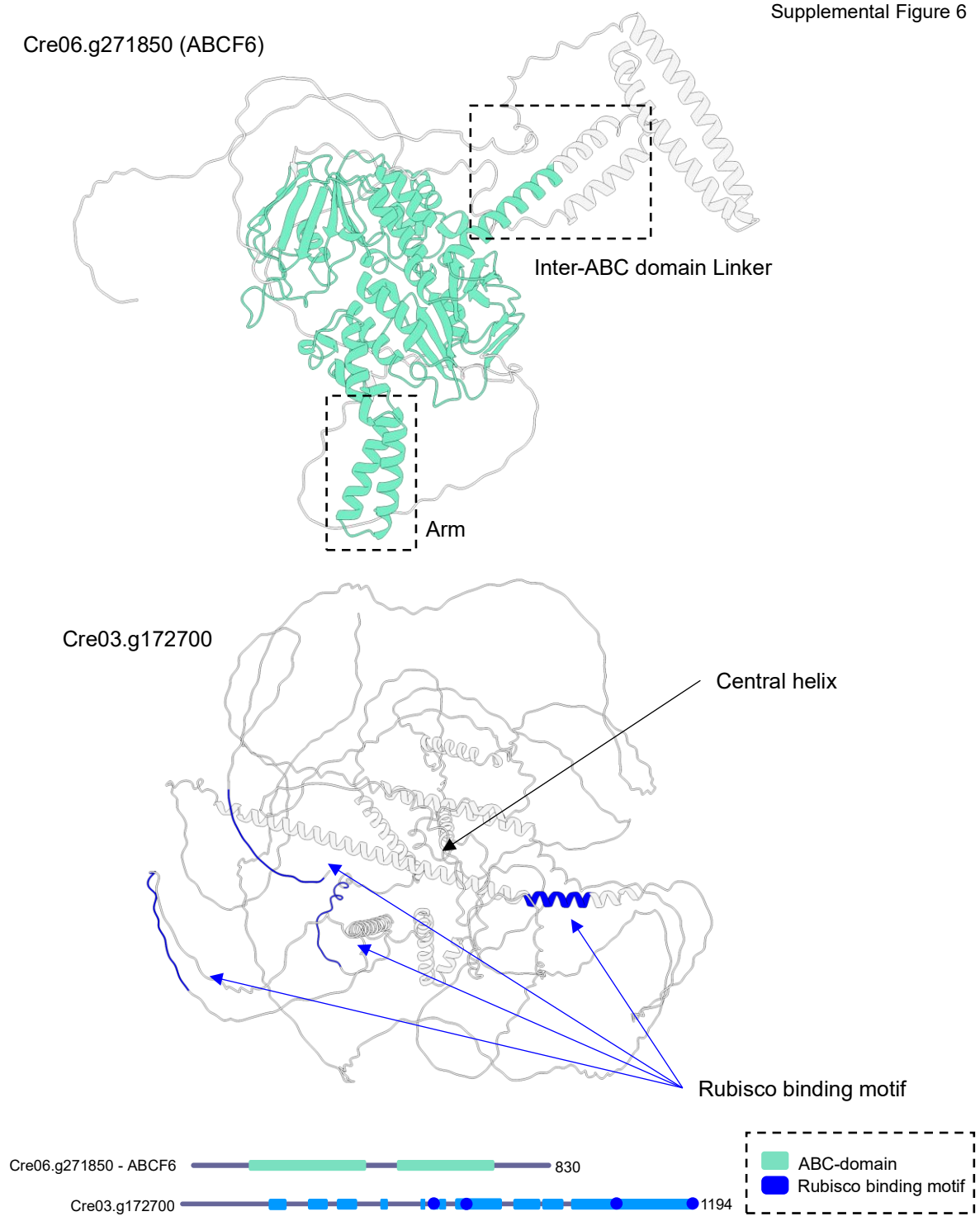



**Supplemental Figure S4.** Enrichment of differentially localized proteins using different **controls.** (Support Figure 3.) Averaged $Log_2$ FC for each comparison group was calculated according to their localization classification, a final category “non-pyrenoid” was created by combining all the non-pyrenoid proteins. Pyrenoids (n=16-19), pyrenoid+stroma (n=3-5) and Non-chloroplastic (n=10). The benchmark proteins used here are shown in Supplemental Data Set S3. Error bars indicate standard error.



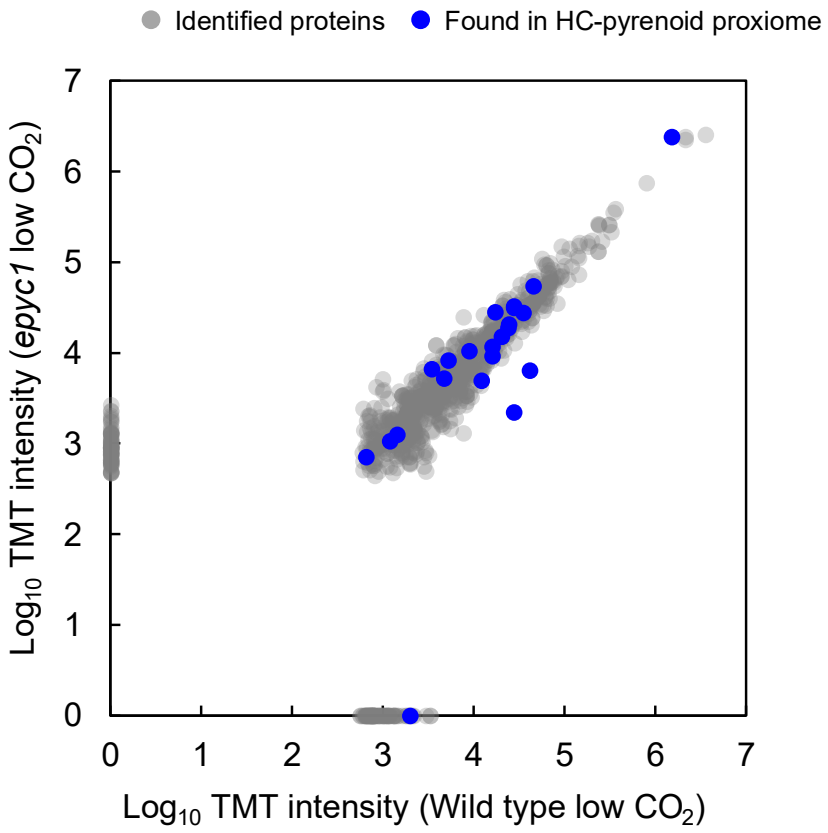

**Supplemental Figure S5.** Additional images of the TurboID identified pyrenoid proteins. (Supports Figure 4.) Two independently transformed strains were imaged for each gene; representative images

shown here. Green and magenta signals denote Venus or Scarlet fluorescence and chlorophyll autofluorescence, respectively. Scale bars, 2 μm.




**Supplemental Figure S6.** AlphaFold-modeled structures for the proteins encoded by Cre06.g271750 (ABCF6) and Cre03.g172700. (Supports Figure 4.) The protein models (A0A2K3DX68 for Cre03.g172700 and A0A2K3DNE2 for Cre06.g271850) were obtained from the Uniparc archive (UniProt Consortium, 2021; Jumper et al., 2021). Dashed rectangles on ABCF6 denotes either the inter-ABC domain linker commonly found on antibiotic resistance-conferring ABCF protein where the linker is hypothesized to interacts with Ribosome-bound antibiotics (Murina et al., 2019) or the L1 ribosome-binding arms. The four Rubisco binding motifs on Cre03.g172700 are colored in blue (Blue

arrows). The black arrow denotes the central alpha helix. Schematic diagrams and color scheme of the two proteins from Figure 4B is presented below.




**Supplemental Figure S7.** Comparison of protein abundance between wildtype and *epyc1* background strains expressing RBCS2-TurboID after streptavidin-affinity purification. (Supports Figure 5.) The Log-transformed TMT intensity values of the normalised intensity form wildtype and *epyc1* grown in low $CO_2$ are compared. Comparison between the two background strains suggests there are only minor difference between the overall protein profile.

**Supplemental Data Set S1.** A list of all plasmids used and developed in this study.

| Plasmid number | Name | Backbone | Level | E. coli antibiotic resistance | Chlamydomonas antibiotic | Source | Part (1) | Part (2) | Part (3) | Part (4) | Part (5) | Part (6) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| - | pCM0-016 | - | 0 | Chloramphenicol | - | Crozet et al., 2018 | - | - | - | - | - | - |
| - | pCM0-017 | - | 0 | Chloramphenicol | - | Crozet et al., 2018 | - | - | - | - | - | - |
| - | pCM0-073 | - | 0 | Chloramphenicol | - | Crozet et al., 2018 | - | - | - | - | - | - |
| - | pCM0-074 | - | 0 | Chloramphenicol | - | Crozet et al., 2018 | - | - | - | - | - | - |
| - | pCM0-100 | - | 0 | Chloramphenicol | - | Crozet et al., 2018 | - | - | - | - | - | - |
| - | pCM0-108 | - | 0 | Chloramphenicol | - | Crozet et al., 2018 | - | - | - | - | - | - |
| - | pCM0-114 | - | 0 | Chloramphenicol | - | Crozet et al., 2018 | - | - | - | - | - | - |
| - | pCM0-094 | - | 0 | Chloramphenicol | - | Crozet et al., 2018 | - | - | - | - | - | - |
| - | pCM0-115 | - | 0 | Chloramphenicol | - | Crozet et al., 2018 | - | - | - | - | - | - |
| - | pAGM8031 | - | - | Spectinomycin | - | Engler et al., 2014 | - | - | - | - | - | - |
| - | pICH47742 | - | - | Ampicillin | - | Engler et al., 2014 | - | - | - | - | - | - |
| - | pICH47732 | - | - | Ampicillin | - | Engler et al., 2014 | - | - | - | - | - | - |
| - | pUAP1 | - | - | Chloramphenicol | - | Patron et al., 2015 | - | - | - | - | - | - |
| - | pICH50881 | - | - | Ampicillin | - | Engler et al., 2014 | - | - | - | - | - | - |
| pLM1386 | EPYC1-1 | pTwist Amp High | -1 | ampicillin | - | This study | - | - | - | - | - | - |
| pLM1387 | EPYC1-2 | pTwist Amp High | -1 | ampicillin | - | This study | - | - | - | - | - | - |
| pLM1388 | EPYC1-3 | pTwist Amp High | -1 | ampicillin | - | This study | - | - | - | - | - | - |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| pLM1389 | EPYC1-4 | pTwist Amp High | -1 | ampicillin | - | This study | - | - | - | - | - | - |
| pLM1390 | RBCS2-1 | pTwist Amp High | -1 | ampicillin | - | This study | - | - | - | - | - | - |
| pLM1391 | RBCS2-2 | pTwist Amp High | -1 | ampicillin | - | This study | - | - | - | - | - | - |
| pLM1392 | TurboID | pTwist Amp High | -1 | ampicillin | - | This study | - | - | - | - | - | - |
| pLM1393 | APEX2 | pTwist Amp High | -1 | ampicillin | - | This study | - | - | - | - | - | - |
| pLM1394 | pUAP1-TurboID | pUAP1 | 0 | Chloramphenicol | - | This study | Flexible linker (by PCR) | TurboID | - | - | - | - |
| pLM891 | pUAP1-EPYC1-TurboID | pUAP1 | 0 | Chloramphenicol | - | This study | EPYC1-1 | EPYC1-2 | EPYC1-3 | EPYC1-4 | Flexible linker (by PCR) | TurboID |
| pLM883 | pUAP1-RBCS2-TurboID | pUAP1 | 0 | Chloramphenicol | - | This study | RBCS2-1 | RBCS2-2 | Flexible linker (by PCR) | TurboID | - | - |
| pLM1395 | pUAP1-RPE1-TurboID | pUAP1 | 0 | Chloramphenicol | - | This study | RPE1 (by PCR) | Flexible linker (by PCR) | TurboID | - | - | - |
| pLM1396 | pUAP1-PRK1-TurboID | pUAP1 | 0 | Chloramphenicol | - | This study | PRK1 (by PCR) | Flexible linker (by PCR) | TurboID | - | - | - |
| pLM879 | pUAP1-RBCS2-APEX2 | pUAP1 | 0 | Chloramphenicol | - | This study | RBCS2-1 | RBCS2-2 | Flexible linker (by PCR) | APEX2 | - | - |
| pLM899 | L1-RBCS2-TurboID | pICH47742 | 1 | ampicillin | - | This study | pCM0-016 | pUAP1-RBCS2-TurboID | pCM0-100 | pCM0-114 | - | - |
| pLM834 | L1-EPYC1-TurboID | pICH47742 | 1 | ampicillin | - | This study | pCM0-016 | pUAP1-EPYC1-TurboID | pCM0-100 | pCM0-114 | - | - |
| pLM1397 | L1-RPE1-TurboID | pICH47742 | 1 | ampicillin | - | This study | pCM0-016 | pUAP1-RPE1-TurboID | pCM0-100 | pCM0-114 | - | - |
| pLM1398 | L1-PRK1-TurboID | pICH47742 | 1 | ampicillin | - | This study | pCM0-016 | pUAP1-PRK1-TurboID | pCM0-100 | pCM0-114 | - | - |
| pLM895 | L1-RBCS2-APEX2 | pICH47742 | 1 | ampicillin | - | This study | pCM0-016 | pUAP1-RBCS2-APEX2 | pCM0-100 | pCM0-114 | - | - |
| pLM902 | L1-RBCS2-TurboID-mCherry | pICH47742 | 1 | ampicillin | - | This study | pCM0-016 | pUAP1-RBCS2-TurboID | pCM0-108 | pCM0-114 | - | - |
| pLM1399 | L1-EPYC1-TurboID-mCherry | pICH47742 | 1 | ampicillin | - | This study | pCM0-016 | pUAP1-EPYC1-TurboID | pCM0-108 | pCM0-114 | - | - |
| pLM1400 | L1-RPE1-TurboID-mCherry | pICH47742 | 1 | ampicillin | - | This study | pCM0-016 | pUAP1-RPE1-TurboID | pCM0-108 | pCM0-114 | - | - |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| pLM1401 | L1-PRK1-TurboID-mCherry | pICH47742 | 1 | ampicillin | - | This study | pCM0-016 | pUAP1-PRK1-TurboID | pCM0-108 | pCM0-114 | - | - |
| pLM630 | L1-Paro | pICH47732 | 1 | ampicillin | Paromomycin | This study | pCM0-017 | pCM0-073 | pCM0-115 | - | - | - |
| pLM631 | L1-Hygro | pICH47732 | 1 | ampicillin | Hygromycin | This study | pCM0-017 | pCM0-074 | pCM0-115 | - | - | - |
| pLM911 | L2-RBCS2-TurboID-Hygro | pAGM8031 | 2 | Spectinomycin | Hygromycin | This study | L1-Hygro | L1-RBCS2-TurboID | pICH50881 | - | - | - |
| pLM1402 | L2-EPYC1-TurboID-Paro | pAGM8031 | 2 | Spectinomycin | Paromomycin | This study | L1-Paro | L1-EPYC1-TurboID | pICH50881 | - | - | - |
| pLM1403 | L2-RPE1-TurboID-Paro | pAGM8031 | 2 | Spectinomycin | Paromomycin | This study | L1-Paro | L1-RPE1-TurboID | pICH50881 | - | - | - |
| pLM1404 | L2-PRK1-TurboID-Paro | pAGM8031 | 2 | Spectinomycin | Paromomycin | This study | L1-Paro | L1-PRK1-TurboID | pICH50881 | - | - | - |
| pLM913 | L2-RBCS2-TurboID-mCherry-Hygro | pAGM8031 | 2 | Spectinomycin | Hygromycin | This study | L1-Hygro | L1-RBCS2-TurboID-mCherry | pICH50881 | - | - | - |
| pLM1405 | L2-EPYC1-TurboID-mCherry-Paro | pAGM8031 | 2 | Spectinomycin | Paromomycin | This study | L1-Paro | L1-EPYC1-TurboID-mCherry | pICH50881 | - | - | - |
| pLM1406 | L2-RPE1-TurboID-mCherry-Paro | pAGM8031 | 2 | Spectinomycin | Paromomycin | This study | L1-Paro | L1-RPE1-TurboID-mCherry | pICH50881 | - | - | - |
| pLM1407 | L2-PRK1-TurboID-mCherry-Paro | pAGM8031 | 2 | Spectinomycin | Paromomycin | This study | L1-Paro | L1-PRK1-TurboID-mCherry | pICH50881 | - | - | - |
| pLM879 | L2-RBCS2-APEX2-Hygro | pAGM8031 | 2 | Spectinomycin | Hygromycin | This study | L1-Hygro | L1-RBCS2-APEX2 | pICH50881 | - | - | - |
| pLM992 | pLM160-Cre13.g573250_mScarle | pLM162 | Recombineering | Kanamycin | Hygromycin | This study | - | - | - | - | - | - |
| pLM156 | pLM099-Cre16.g663150_Venus | pLM099 | Recombineering | Kanamycin | Paromomycin | This study | - | - | - | - | - | - |
| pLM165 | pLM099-Cre06.g271850_Venus | pLM099 | Recombineering | Kanamycin | Paromomycin | This study | - | - | - | - | - | - |
| pLM1408 | pLM099-Cre03.g172700_Venus | pLM099 | Recombineering | Kanamycin | Paromomycin | This study | - | - | - | - | - | - |
| pLM1409 | pLM099-Cre17.g720450_Venus | pLM099 | Recombineering | Kanamycin | Paromomycin | This study | - | - | - | - | - | - |
| pLM547 | pLM099-Cre09.g394510_Venus | pLM099 | Recombineering | Kanamycin | Paromomycin | This study | - | - | - | - | - | - |
| pLM1411 | pLM099-Cre02.g093650_Venus | pLM099 | Recombineering | Kanamycin | Paromomycin | This study | - | - | - | - | - | - |

**Supplemental Data Set S2.** Pilot mass spectrometry analysis of RBCS2-TurboID samples against wild type controls. Gene annotations and autodefinition obtained from Phytozome were used to annotate all proteins detected. Localization predictions are marked under "PredAlgo" and were predicted by submitting each protein sequence to the PredAlgo localization prediction webserver (Tardiff et al., 2012). The symbols used denote the following, C: Chloroplast; M: Mitochondria; SP: Signal Peptide and O: Others. Classification is based on the localization obtained from the literature according to Supplemental Data Set 3. This allows assignment of true and false positive proteins. The $Log_2$ FC of detected proteins at each timepoint was calculated. The average $Log_2$ FC used in Figure 2 was calculated as a comparison between the sum of all intensity from RBCS2-TurboID against wild type. An one-way ANOVA test was carried out for the binary comparison [Anova (p)], and was transformed to a $-Log_{10}P$ value. **This material is too large to include in the thesis and is instead available online at https://doi.org/10.1093/plcell/koad131.**

**Supplemental Data Set S3.** Localization of proteins based on fluorescence tagging/immunofluorescence obtained from the literature.

| Gene Identifier | Gene name | Source | Localization | Classification | TP | FP |
|---|---|---|---|---|---|---|
| Cre02.g120100 | RBCS1 | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Pyrenoid matrix | Pyrenoid | 1 | 0 |
| Cre02.g120150 | RBCS2 | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Pyrenoid matrix | Pyrenoid | 1 | 0 |
| Cre04.g229300 | RCA1 | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Pyrenoid matrix | Pyrenoid | 1 | 0 |
| Cre06.g261750 | RBMP1 | Meyer et al., 2020 (DOI: 10.1126/sciadv.abd2408) | Pyrenoid tubule | Pyrenoid | 1 | 0 |
| Cre06.g307500 | LCIC | Yamano et al., 2010 (DOI: 10.1093/pcp/pcq105) | Pyrenoid periphery | Pyrenoid | 1 | 0 |
| Cre08.g362900 | PSBP4 | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Pyrenoid periphery | Pyrenoid | 1 | 0 |
| Cre09.g394473 | LCI9 | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Pyrenoid periphery | Pyrenoid | 1 | 0 |
| Cre09.g394621 | SAGA2 | Meyer et al., 2020 (DOI: 10.1126/sciadv.abd2408) | Pyrenoid periphery | Pyrenoid | 1 | 0 |
| Cre09.g415700 | CAH3 | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Pyrenoid tubule | Pyrenoid | 1 | 0 |
| Cre09.g416850 | RBMP2 | Meyer et al., 2020 (DOI: 10.1126/sciadv.abd2408) | Pyrenoid tubule | Pyrenoid | 1 | 0 |
| Cre10.g435800 | CSP41B | Meyer et al., 2020 (DOI: 10.1126/sciadv.abd2408) | Pyrenoid matrix | Pyrenoid | 1 | 0 |
| Cre10.g436550 | EPYC1 | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Pyrenoid matrix | Pyrenoid | 1 | 0 |
| Cre10.g440050 | CSP41A | Meyer et al., 2020 (DOI: 10.1126/sciadv.abd2408) | Pyrenoid matrix | Pyrenoid | 1 | 0 |
| Cre10.g444700 | SBE3 | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Pyrenoid periphery | Pyrenoid | 1 | 0 |
| Cre10.g452800 | LCIB | Yamano et al., 2010 (DOI: 10.1093/pcp/pcq105) | Pyrenoid periphery | Pyrenoid | 1 | 0 |
| Cre11.g467712 | SAGA1 | Meyer et al., 2020 (DOI: 10.1126/sciadv.abd2408) | Pyrenoid periphery | Pyrenoid | 1 | 0 |
| Cre12.g497300 | CAS1 | Wang et al., 2016 (https://doi.org/10.1073/pnas.160651911) | Pyrenoid tubule | Pyrenoid | 1 | 0 |
| Cre17.g718950 | RCA2 | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Pyrenoid matrix | Pyrenoid | 1 | 0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Cre17.g721500 | STA2 | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Starch sheath | Pyrenoid | 1 | 0 |
| Cre01.g030900 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre01.g051500 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre02.g078507 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre02.g143450 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre03.g146167 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre03.g179800 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre03.g183850 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre03.g189800 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Thylakoidal | Pyrenoid + stroma | 0 | 1 |
| Cre03.g191250 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre03.g199535 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Thylakoidal | Pyrenoid + stroma | 0 | 1 |
| Cre06.g251150 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Thylakoidal | Pyrenoid + stroma | 0 | 1 |
| Cre06.g259900 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Thylakoidal | Pyrenoid + stroma | 0 | 1 |
| Cre06.g261000 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Thylakoidal | Pyrenoid + stroma | 0 | 1 |
| Cre06.g283750 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Thylakoidal | Pyrenoid + stroma | 0 | 1 |
| Cre08.g372450 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre09.g416800 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre10.g439350 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre12.g494850 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre12.g519300 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Thylakoidal | Pyrenoid + stroma | 0 | 1 |
| Cre12.g560950 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre14.g626700 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre16.g651050 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre16.g652800 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre16.g658400 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre16.g662150 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre17.g724300 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre17.g740950 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Pyrenoid + stroma | 0 | 1 |
| Cre07.g330250 | PSAH | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Thylakoidal | Pyrenoid + stroma | 0 | 1 |
| Cre01.g010900 | GAP3 | Kuken et al., 2018 (DOI: 10.7554/eLife.37960) | Stromal | Stroma | 0 | 1 |
| Cre01.g037850 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Stroma | 0 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Cre01.g049000 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Stroma | 0 | 1 |
| Cre02.g093450 | FBA3 | Kuken et al., 2018 (DOI: 10.7554/eLife.37960) | Stromal | Stroma | 0 | 1 |
| Cre02.g144800 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Stroma | 0 | 1 |
| Cre03.g148000 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Chloroplast envelope | Stroma | 0 | 1 |
| Cre03.g156600 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Stroma | 0 | 1 |
| Cre03.g181500 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Stroma | 0 | 1 |
| Cre03.g185550 | SBP1 | Kuken et al., 2018 (DOI: 10.7554/eLife.37960) | Stromal | Stroma | 0 | 1 |
| Cre03.g187450 | RPI1 | Kuken et al., 2018 (DOI: 10.7554/eLife.37960) | Stromal | Stroma | 0 | 1 |
| Cre05.g234550 | FBA2 | Kuken et al., 2018 (DOI: 10.7554/eLife.37960) | Stromal | Stroma | 0 | 1 |
| Cre06.g255300 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Chloroplast envelope | Stroma | 0 | 1 |
| Cre06.g260450 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Chloroplast envelope | Stroma | 0 | 1 |
| Cre06.g285401 | FAP355, HU | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Stroma | 0 | 1 |
| Cre06.g306300 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Stroma | 0 | 1 |
| Cre06.g309000 | LCIA | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Chloroplast envelope | Stroma | 0 | 1 |
| Cre07.g334800 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Stroma | 0 | 1 |
| Cre07.g335700 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Stroma | 0 | 1 |
| Cre07.g348600 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Chloroplast envelope | Stroma | 0 | 1 |
| Cre08.g379650 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Chloroplast envelope | Stroma | 0 | 1 |
| Cre09.g389615 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Stroma | 0 | 1 |
| Cre09.g389689 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Stroma | 0 | 1 |
| Cre11.g467759 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Chloroplast envelope | Stroma | 0 | 1 |
| Cre11.g467770 | PGK1 | Kuken et al., 2018 (DOI: 10.7554/eLife.37960) | Stromal | Stroma | 0 | 1 |
| Cre12.g485150 | GAP1 | Kuken et al., 2018 (DOI: 10.7554/eLife.37960) | Stromal | Stroma | 0 | 1 |
| Cre12.g511900 | RPE1 | Kuken et al., 2018 (DOI: 10.7554/eLife.37960) | Stromal | Stroma | 0 | 1 |
| Cre12.g524950 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Stroma | 0 | 1 |
| Cre12.g554800 | PRK1 | Kuken et al., 2018 (DOI: 10.7554/eLife.37960) | Stromal | Stroma | 0 | 1 |
| Cre13.g577100 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Stroma | 0 | 1 |
| Cre13.g584850 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Stromal | Stroma | 0 | 1 |
| Cre13.g592400 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Thylakoidal | Stroma | 0 | 1 |
| Cre16.g663400 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Thylakoidal | Stroma | 0 | 1 |
| Cre16.g663450 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Thylakoidal | Stroma | 0 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Cre17.g744547 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Chloroplast envelope | Stroma | 0 | 1 |
| Cre10.g423500 | HMO1 | Duanmu et al., 2013 (https://doi.org/10.1073/pnas.1222375110) | Stromal | Stroma | 0 | 1 |
| Cre13.g587100 | PCYA | Duanmu et al., 2013 (https://doi.org/10.1073/pnas.1222375110) | Thylakoidal | Stroma | 0 | 1 |
| Cre02.g080700 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre02.g088200 | RB60 | Levitan et al., 2005 (https://doi.org/10.1073/pnas.0500676102) | Mitochondria | Non chloroplastic | 0 | 1 |
| Cre02.g097800 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre02.g097800 | HLA3 | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Plasma membrane | Non chloroplastic | 0 | 1 |
| Cre02.g108950 | LF1 | Nguyen et al., 2005 (doi: 10.1534/genetics.104.027615) | Flagella | Non chloroplastic | 0 | 1 |
| Cre02.g145950 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Plasma membrane | Non chloroplastic | 0 | 1 |
| Cre03.g146287 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Mitochondria | Non chloroplastic | 0 | 1 |
| Cre03.g162580 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre03.g162800 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre03.g162800 | LCI1 | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Plasma membrane | Non chloroplastic | 0 | 1 |
| Cre03.g164600 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Plasma membrane | Non chloroplastic | 0 | 1 |
| Cre03.g171350 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre03.g176833 | NAP1 | Onishi et al., 2015 (https://doi.org/10.1534/genetics.115.184663) | Flagella | Non chloroplastic | 0 | 1 |
| Cre03.g181200 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Mitochondria | Non chloroplastic | 0 | 1 |
| Cre03.g181600 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre03.g202400 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Mitochondria | Non chloroplastic | 0 | 1 |
| Cre03.g204577 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre03.g210961 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Plasma membrane | Non chloroplastic | 0 | 1 |
| Cre04.g222750 | CCP2 | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Mitochondria | Non chloroplastic | 0 | 1 |
| Cre04.g223050 | CAH2 | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre04.g223300 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Mitochondria | Non chloroplastic | 0 | 1 |
| Cre05.g245950 | DRP1 | Arakaki et al., 2017(https://doi.org/10.1186/s12862-017-1091-z) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre05.g248400 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Mitochondria | Non chloroplastic | 0 | 1 |
| Cre05.g248450 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Mitochondria | Non chloroplastic | 0 | 1 |
| Cre06.g258850 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre06.g263550 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre06.g271800 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Cre06.g275000 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre06.g284150 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre06.g295450 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Mitochondria | Non chloroplastic | 0 | 1 |
| Cre07.g313800 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre07.g321800 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre07.g336350 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Nuclear envelope | Non chloroplastic | 0 | 1 |
| Cre07.g339900 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Golgi | Non chloroplastic | 0 | 1 |
| Cre07.g344260 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Golgi | Non chloroplastic | 0 | 1 |
| Cre08.g360200 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Golgi | Non chloroplastic | 0 | 1 |
| Cre09.g386650 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Mitochondria | Non chloroplastic | 0 | 1 |
| Cre09.g396400 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre09.g402219 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre09.g403900 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre09.g415850 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Mitochondria | Non chloroplastic | 0 | 1 |
| Cre09.g417150 | CAT1 | Kato et al., 2021 (https://doi.org/10.3390/cells10081940) | Peroxisome | Non chloroplastic | 0 | 1 |
| Cre10.g427250 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre10.g435100 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre10.g439500 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre10.g457750 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre12.g483950 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Mitochondria | Non chloroplastic | 0 | 1 |
| Cre12.g485050 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Flagella | Non chloroplastic | 0 | 1 |
| Cre12.g486250 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre12.g487900 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre12.g507300 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Nuclear envelope | Non chloroplastic | 0 | 1 |
| Cre12.g516450 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Mitochondria | Non chloroplastic | 0 | 1 |
| Cre12.g519500 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre12.g528750 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre12.g531400 | NPHP4 | Jack et al., 2019 (https://doi.org/10.1016/j.celrep.2019.04.087) | Basal body | Non chloroplastic | 0 | 1 |
| Cre12.g552450 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre12.g554850 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Vacuole | Non chloroplastic | 0 | 1 |
| Cre13.g603700 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Cre14.g632350 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Flagella | Non chloroplastic | 0 | 1 |
| Cre16.g661700 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre16.g661750 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre16.g671350 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre16.g683600 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre16.g685000 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre16.g692228 | MARS1 | Perlaza et al., 2019 (https://doi.org/10.7554/eLife.49577) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre17.g698400 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Mitochondria | Non chloroplastic | 0 | 1 |
| Cre17.g724500 | | Mackinder et al., 2017 (DOI: 10.1016/j.cell.2017.08.044) | Centriole | Non chloroplastic | 0 | 1 |
| Cre03.g176961 | MAR1 | Pinello et al., 2021 (https://doi.org/10.1016/j.devcel.2021.10.023) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre09.g395950 | AOX1 | Kaye et al., 2021 (https://doi.org/10.1074/jbc.RA118.004667) | Mitochondria | Non chloroplastic | 0 | 1 |
| Cre03.g169550 | AOX2 | Kaye et al., 2021 (https://doi.org/10.1074/jbc.RA118.004667) | Mitochondria | Non chloroplastic | 0 | 1 |
| Cre11.g467753 | HMO2 | Duanmu et al., 2013 (https://doi.org/10.1073/pnas.1222375110) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre03.g169500 | FLS2 | Zhao et al., 2020 (https://doi.org/10.1371/journal.pgen.1008561) | Cytosolic | Non chloroplastic | 0 | 1 |
| Cre07.g329300 | MSC1 | Nakayama et al., 2007 (https://doi.org/10.1073/pnas.0609996104) | Chloroplast envelope | Stroma | 0 | 1 |
| 28269772 | rbcL | Putative | Pyrenoid | Pyrenoid | 1 | 0 |
| Cre10.g430350 | - | Meyer et al., 2020 (DOI: 10.1126/sciadv.abd2408) | Pyrenoid | Pyrenoid | 1 | 0 |

**Supplemental Data Set S4.** TMT Mass spectrometry analysis of pyrenoid baits against stromal and wild type controls. Gene annotations and autodefinition obtained from Phytozome were used to annotate all genes detected. Localization prediction are marked under "PredAlgo" and was predicted by submitting each protein sequence to the PredAlgo localization prediction webserver (Tardiff et al., 2012), chloroplast translated proteins were manually anotated as Chloroplast localization. The symbols used denote the following C: Chloroplast; M: Mitochondria; SP: Signal Peptide and O: Others. Classification is based on the localization obtained from the literature according to Supplemental Data Set S3 which is then used to denote if a detected protein is a true or false positive. The $Log_2$ FC of detected proteins for each binary comparison was calculated and a PEAKSQ test was used to calculate the statistical signficance as a -LogP value. **This material is too large to include in the thesis and is instead available online at https://doi.org/10.1093/plcell/koad131.**

**Supplemental Data Set S5**. Pyrenoid proxiome. The Pyrenoid proxiome is generated by comparing the significantly enriched proteins in all comparison groups. The number of appearance for each protein in each comparison group is denoted under "Occurence", proteins are considered to be within the pyrenoid proxiome when they appear in more than 1 comparison group. The associated Log2 FC in each comparison group is also listed.

| Protein Group | Protein ID | Accession | Description | Gene name | Anotations | TP | FP | Pred Algo | $Log_2FC$ (RBCS2 vs Wildtype) | $Log_2FC$ (RBCS2 vs RPE1) | $Log_2FC$ (RBCS2 vs PRK1) | $Log_2FC$ (EPYC1 vs Wildtype) | $Log_2FC$ (EPYC1 vs RPE1) | $Log_2FC$ (EPYC1 vs PRK1) | Occurences | Present In |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 9 | 28269780 | TPA_inf: CF1 ATP synthase beta subunit (chloroplast) [Chlamydomonas reinhardtii] | #N/A | #N/A | 0 | 0 | C | 2.28509468 | -1.78945985 | -1.64868413 | 2.3714383 | -1.70311624 | -1.56234052 | 2 | RBCS WT, EPYC1 WT |
| 9 | 60 | 28269772 | TPA_inf: ribulose bisphosphate carboxylase large subunit (chloroplast) [Chlamydomonas reinhardtii] | rbcL | #N/A | 1 | 0 | C | 5.01448182 | 6.550856617 | 8.96991073 | 3.1681094 | 4.704484244 | 7.12353835 | 6 | RBCS WT, EPYC1 WT, RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 10 | 36 | 28269770 | TPA_inf: RNA polymerase beta' subunit (chloroplast) [Chlamydomonas reinhardtii] | #N/A | #N/A | 0 | 0 | C | 4.375305148 | -1.3518767 | -0.17828333 | 5.5809374 | -0.14624441 | 1.02734896 | 2 | RBCS WT, EPYC1 WT |
| 11 | 7 | Cre06.g259100.t1.1 | Cre06.g259100.t1.1 | SAGA3 | - | 0 | 0 | C | 3.733046272 | 2.674747371 | 2.45843266 | 3.2159012 | 2.157602274 | 1.94128757 | 6 | RBCS WT, EPYC1 WT, RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 14 | 23 | Cre13.g592200.t1.2 | Cre13.g592200.t1.2 | GSN1 | NADH-dependent glutamate synthase (EC 1.4.1.14); NADH-dependent glutamine-2-oxoglutarate aminotransferase (NADH-GOGAT); plastid targeting predicted by Target-P | 0 | 0 | C | 3.085868644 | -0.8123261 | -1.80853199 | 3.5265719 | -0.37162284 | -1.36782873 | 2 | RBCS WT, EPYC1 WT |
| 16 | 20 | Cre12.g514850.t1.2 | Cre12.g514850.t1.2 | HSP90C | HSP90C is one of 3 HSP90 genes in Chlamydomonas; HSP90C is localized to the chloroplast; it is inducible by heat shock and light, forms dimers and interacts with plastidic HSP70B; it exhibits weak ATPase activity, which is inhibited by radicicol; PMID: 16143837 and 15995001 | 0 | 0 | M | 2.187635033 | -0.77050258 | -0.60322954 | 2.2221337 | -0.73600389 | -0.56873085 | 2 | RBCS WT, EPYC1 WT |
| 17 | 10 | Cre09.g394621.t2.1 | Cre09.g394621.t2.1 | SAGA2 | - | 1 | 0 | C | 5.2021542 | 4.258121133 | 3.6437226 | 4.6131783 | 3.669145198 | 3.05474667 | 6 | RBCS WT, EPYC1 WT, RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 20 | 46 | 32880375 | TPA_inf: hypothetical protein (chloroplast) [Chlamydomonas reinhardtii] | #N/A | #N/A | 0 | 0 | C | 0.795163678 | -1.05880976 | -0.50539795 | 2.291506 | 0.437532592 | 0.9909444 | 1 | EPYC1 WT |
| 22 | 185 | Cre10.g436550.t1.2 | Cre10.g436550.t1.2 | EPYC1 | low-complexity repeat protein, probably serving as a linker between RuBisCO proteins in the pyrenoid matrix; regulated by CCM1 [PMID: 15235119]; phopshorylated under low CO2; essential for pyrenoid formation | 1 | 0 | C | 4.021402245 | 4.116820333 | 4.43276869 | 3.3348675 | 3.430285547 | 3.7462339 | 6 | RBCS WT, EPYC1 WT, RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 23 | 24 | Cre11.g467712.t1.1 | Cre11.g467712.t1.1 | SAGA1 | - | 1 | 0 | M | 3.869855779 | 2.630592599 | 2.62222098 | 3.2846242 | 2.045360991 | 2.03698937 | 6 | RBCS WT, EPYC1 WT, RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 24 | 39 | Cre04.g229300.t1.1 | Cre04.g229300.t1.1 | RCA1 | RubisCO activase, chloroplast precursor; required for optimal photosynthesis in a low CO2 atmosphere [PMID: 16667924; independent cDNA sequence: M62962]. | 1 | 0 | C | 2.969100247 | 2.059120347 | 2.76935841 | 2.5587587 | 1.648778809 | 2.35901687 | 5 | RBCS WT, EPYC1 WT, RBCS RPE1, RBCS PRK1, EPYC1 PRK1 |
| 29 | 29 | Cre10.g435800.t1.2 | Cre10.g435800.t1.2 | CSP41B | Cleaves RNA in vitro, but appears necessary for stabilizing Cp mRNA in Arabidopsis and spinach; Arabidopsis homolog binds multiple mRNA and rRNA and interacts with Csp41a in multimeric complexes; also found associated with ribosomes and RNA polymerase; cp ribosome-associated protein RAP38; similar to NAD-dependent nucleoside-diphosphate-sugar epimerase/dehydratase | 1 | 0 | C | 2.773297467 | 4.885544098 | 5.34780073 | 0.8800006 | 2.992247234 | 3.45450387 | 5 | RBCS WT, RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 32 | 30 | Cre15.g635650.t2.1 | Cre15.g635650.t2.1 | - | - | 0 | 0 | C | 2.089740918 | -1.57168169 | -0.50763148 | 3.2797829 | -0.3816397 | 0.68241052 | 2 | RBCS WT, EPYC1 WT |
| 32 | 31 | Cre15.g635650.t1.2 | Cre15.g635650.t1.2 | - | - | 0 | 0 | C | 2.089740918 | -1.57168169 | -0.50763148 | 3.2797829 | -0.3816397 | 0.68241052 | 2 | RBCS WT, EPYC1 WT |
| 35 | 17 | Cre09.g416850.t1.2 | Cre09.g416850.t1.2 | RBMP2 | - | 1 | 0 | M | 3.45374719 | 3.066857359 | 3.02620735 | 1.3973574 | 1.010467521 | 0.96981751 | 2 | RBCS WT, RBCS RPE1 |
| 38 | 38 | Cre02.g101400.t1.2 | Cre02.g101400.t1.2 | CHC1 | Clathrin Heavy Chain, vesicle coat protein; Expressed Protein. ts-lethal mutant was isolated (PMID 29743196) | 0 | 0 | O | 1.92840315 | 0.741817998 | 0.79875972 | 1.8404941 | 0.653908907 | 0.71085063 | 2 | RBCS WT, EPYC1 WT |
| 40 | 43 | Cre17.g718950.t1.2 | Cre17.g718950.t1.2 | RCA2 | Related to RuBisCO activase (RCA); catalyzes the activation of RuBisCO (ribulose-1,5-bisphosphate carboxylase/oxygenase; EC 4.1.1.39); via the ATP-dependent carboxylation of the epsilon-amino group of lysine. | 1 | 0 | O | 4.630701324 | 4.831344904 | 4.7068665 | 3.8796068 | 4.080250389 | 3.95577198 | 6 | RBCS WT, EPYC1 WT, RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 41 | 465 | Cre02.g093650.t1.2 | Cre02.g093650.t1.2 | - | - | 0 | 0 | C | 5.523821199 | -0.92500632 | -0.87828967 | 5.8957887 | -0.5530388 | -0.50632215 | 2 | RBCS WT, EPYC1 WT |
| 42 | 93 | Cre10.g451900.t1.1 | Cre10.g451900.t1.1 | THS1 | putative threonine synthase (EC:4.2.3.1); organelle (mitochondria) targeting predicted by Target-P; plastid location expected based on homology | 0 | 0 | C | 3.125837127 | -0.8848192 | -0.33663828 | 2.9736642 | -1.03699209 | -0.48881118 | 2 | RBCS WT, EPYC1 WT |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 44 | 100 | Cre06.g269050.t1.2 | Cre06.g269050.t1.2 | TIC62 | Short chain dehydrogenase belonging to the extended family [PMID: 1223552]. Similarity to the N-terminal domain of Tic62 [PMID: 17374152]., Similar to Ycf39-like proteins from Synechocystis sp. PCC 683 (BAA1812), Cyanophora paradoxa (AAA81188) and Guillardia theta | 0 | 0 | C | 3.92410909 | -0.51062014 | -0.74696054 | 4.7220193 | 0.28729003 | 0.05094963 | 2 | RBCS WT, EPYC1 WT |
| 45 | 28 | Cre12.g517900.t1.1 | Cre12.g517900.t1.1 | SECA1 | SecA protein associated with import of proteins into chloroplast. Strong homology to SecA | 0 | 0 | C | 3.017747875 | -0.72261657 | -0.28825177 | 3.1257369 | -0.61462752 | -0.18026273 | 2 | RBCS WT, EPYC1 WT |
| 47 | 16 | Cre06.g282000.t1.1 | Cre06.g282000.t1.1 | STA3 | - | 0 | 0 | C | 3.372246244 | 0.21400823 | 0.68057025 | 2.4555951 | -0.70264291 | -0.2360809 | 2 | RBCS WT, EPYC1 WT |
| 48 | 21 | Cre10.g444700.t1.1 | Cre10.g444700.t1.1 | SBE3 | catalyses the formation of alpha-1,6 linkages within the polymer by cleaving a preexisting linear chain and transferring the fragment at the non reducing end of the cleaved glucan to an alpha-1,6 position | 1 | 0 | C | 2.49464823 | 0.23641574 | 0.40470184 | 2.7352504 | 0.47701787 | 0.64530396 | 2 | RBCS WT, EPYC1 WT |
| 52 | 53 | Cre10.g440050.t1.2 | Cre10.g440050.t1.2 | CSP41A | Homologous to CSP41b; Arabidopsis homologs form multimeric complexes and bind multiple mRNA and rRNA; similar to NAD-dependent nucleoside-diphosphate-sugar epimerase/dehydratase | 1 | 0 | C | 5.629072594 | 7.73112714 | 7.26108708 | 3.8837087 | 5.85058551 | 5.43689008 | 6 | RBCS WT, EPYC1 WT, RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 54 | 138 | 28269765 | TPA_inf: RNA polymerase beta subunit I (chloroplast) [Chlamydomonas reinhardtii] | #N/A | #N/A | 0 | 0 | C | 3.325501462 | -1.51065228 | -0.72423582 | 4.7769251 | -0.05922866 | 0.72718781 | 2 | RBCS WT, EPYC1 WT |
| 56 | 76 | Cre09.g386735.t1.1 | Cre09.g386735.t1.1 | DLA1 | Dihydrolipoamide acetyltransferase, probably mitochondrial; EC 2.3.1.12); E2 component of pyruvate dehydrogenase complex; null-allele mutant was isolated (PMID 29743196) | 0 | 0 | M | 3.530378936 | 0.95724366 | 1.67287511 | 5.494328 | 2.92119268 | 3.63682414 | 5 | RBCS WT, EPYC1 WT, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 57 | 96 | Cre02.g097800.t1.1 | Cre02.g097800.t1.1 | HLA3 | associated with bicarbonate uptake for CO2-cocentrating mechanism [PMID:19321421, 25660294, 26015566]; high light-induced, requiring both high light and low CO2 (ambient) levels for activation [PMID: 12000678]; MRP subfamily of ABC transporters [PMID: 15710683]; HLA3 is regulated by CCM1 [PMID: 15235119]; localized at plasma membrane [PMID: 15710683] | 0 | 1 | O | 2.044347485 | 2.23646443 | 1.88549887 | 0.5669569 | 0.75907181 | 0.40810824 | 3 | RBCS WT, RBCS RPE1, RBCS PRK1 |
| 61 | 56 | Cre12.g519180.t1.1 | Cre12.g519180.t1.1 | EFT1a | Polyprotein of EF-Ts; Chloroplast small ribosomal subunit protein PSRP-7 and elongation factor Ts are encoded in this single transcript | 0 | 0 | C | 2.02961635 | -1.42141167 | -0.68997867 | 1.6177186 | -1.83330946 | -1.10187646 | 1 | RBCS WT |
| 61 | 55 | Cre12.g519180.t2.1 | Cre12.g519180.t2.1 | EFT1a | Polyprotein of EF-Ts; Chloroplast small ribosomal subunit protein PSRP-7 and elongation factor Ts are encoded in this single transcript | 0 | 0 | C | 2.02961635 | -1.42141167 | -0.68997867 | 1.6177186 | -1.83330946 | -1.10187646 | 1 | RBCS WT |
| 64 | 95 | Cre01.g054500.t1.2 | Cre01.g054500.t1.2 | - | - | 0 | 0 | M | 2.800297088 | -0.52278102 | -0.38915074 | 2.905314 | -0.41776409 | -0.28413381 | 2 | RBCS WT, EPYC1 WT |
| 67 | 498 | Cre02.g120150.t1.2 | Cre02.g120150.t1.2 | RBCS2 | RuBisCO small subunit 2, chloroplast precursor [PMID: 3820291]; almost identical to SwissProt P08475 and to nearby RBCS1; RBCS2 is the only isoform present in the crystallized complex [PMID: 11866526] | 1 | 0 | C | 3.140641542 | 4.84885271 | 4.37881266 | 0 | 0 | 0 | 3 | RBCS WT, RBCS RPE1, RBCS PRK1 |
| 68 | 68 | Cre07.g356350.t1.1 | Cre07.g356350.t1.1 | DXS1 | 1-deoxy-D-xylulose 5-phosphate synthase (1-deoxyxylulose-5-phosphate synthase) (DXP synthase) (DXPS) (CLA1) (DEF), chloroplast precursor | 0 | 0 | C | 1.852116788 | -1.11483243 | -1.15301854 | 2.4364866 | -0.53046206 | -0.56864871 | 2 | RBCS WT, EPYC1 WT |
| 70 | 126 | Cre06.g259900.t1.2 | Cre06.g259900.t1.2 | ATPC | ATP synthase CF1 gamma chain, chloroplast precursor, gi:114638, PMID: 2904436, PMID: 8349573, PMID: 8543042 | 0 | 1 | C | 3.376042534 | -1.68376079 | -1.54202887 | 4.0285883 | -1.03121504 | -0.88948313 | 2 | RBCS WT, EPYC1 WT |
| 71 | 61 | Cre07.g352350.t1.2 | Cre07.g352350.t1.2 | FHL3 | Similar to FtsH protease, but lacks peptidase HEXXH motif, like At FtsHi; may function with Tic214 (Ycf2) in the TIC translocon | 0 | 0 | M | 1.359370161 | -1.64895078 | -0.53486801 | 3.0331903 | 0.02486932 | 1.1389521 | 1 | EPYC1 WT |
| 74 | 129 | Cre09.g397697.t1.1 | Cre09.g397697.t1.1 | RPL4 | Component of cytosolic 80S ribosome and 60S large subunit | 0 | 0 | O | 1.181004136 | -0.70433829 | -1.44199843 | 2.160611 | 0.27526858 | -0.46239156 | 1 | EPYC1 WT |
| 76 | 54 | Cre17.g746997.t1.1 | Cre17.g746997.t1.1 | ADH1 | Dual function alcohol dehydrogenase / acetaldehyde dehydrogenase [EC:1.1.1.1 /1.2.1.10], probably mitochondrial [PMID: 14756315]. Consolidates 3 previous v3 models, probably correctly | 0 | 0 | C | 3.768835939 | 0.22072239 | 0.30772322 | 2.9568717 | -0.59124118 | -0.50424097 | 2 | RBCS WT, EPYC1 WT |
| 78 | 177 | Cre06.g250100.t1.2 | Cre06.g250100.t1.2 | HSP70B | HSP70B is one of 9 genes encoding HSP70-like proteins in Chlamydomonas; HSP70B is localized to the chloroplast, there mainly to the stroma, but also to thylakoid and low density membranes; HSP70B is inducible by heat shock and light; it interacts with the GrpE-type nucleotide exchange factor CGE1 in an ATP-sensitive manner; HSP70B also interacts with HSP90C; HSP70B by an unknown mechanism protects PSII from photodamage and improves the repair of photodamaged PSII; with its cochaperone CDJ2, HSP70B interacts with VIPP1 and catalyzes assembly and disassembly of VIPP1 rings, as well as the disassembly of VIPP1 rods; HSP70B constitutes about 0.2% of cellular protein [PMID: 2779571, 8914534, 10368186, 11497999, 11752390, 15635096, 15995001, 16143837] | 0 | 0 | C | 1.46906482 | -1.00872051 | -0.775134 | 1.6963301 | -0.78145523 | -0.54786872 | 1 | EPYC1 WT |
| 81 | 66 | Cre12.g483700.t1.2 | Cre12.g483700.t1.2 | RB38 | RNA-binding protein of 38 kDa, OSB/PTAC9 homolog; targeted to chloroplast, binds psbA 5' UTR; homology to DUF731 proteins of unknown function found in plants | 0 | 0 | C | 2.1617684 | -0.45746281 | -0.00777529 | 1.5479869 | -1.07124435 | -0.62155683 | 1 | RBCS WT |
| 81 | 65 | Cre12.g483700.t2.1 | Cre12.g483700.t2.1 | RB38 | RNA-binding protein of 38 kDa, OSB/PTAC9 homolog; targeted to chloroplast, binds psbA 5' UTR; homology to DUF731 proteins of unknown function found in plants | 0 | 0 | C | 2.1617684 | -0.45746281 | -0.00777529 | 1.5479869 | -1.07124435 | -0.62155683 | 1 | RBCS WT |
| 82 | 105 | 28269758 | TPA_inf: ribosomal protein S2 (chloroplast) [Chlamydomonas reinhardtii] | #N/A | #N/A | 0 | 0 | C | 1.913381014 | -1.9402854 | -1.22833587 | 3.7430252 | -0.11064126 | 0.60130827 | 2 | RBCS WT, EPYC1 WT |
| 84 | 18 | Cre12.g497350.t1.1 | Cre12.g497350.t1.1 | PPR13 | - | 0 | 0 | O | 0 | -0.75179676 | 0.76512104 | 0.4842812 | 0.17035781 | 1.68727562 | 1 | EPYC1 PRK1 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 88 | 47 | Cre05.g240850.t1.2 | Cre05.g240850.t1.2 | THIC | involved in the regulation of thiamine biosynthesis. Precise catalytic function of this ThiC homolog is not known, though it participates in the formation of 4-amino-5-hydroxymethyl-2-methylpyrimidine | 0 | 0 | C | 2.90387472 | -0.25848151 | -2.36046879 | 3.9116881 | 0.749331896 | -1.35265538 | 2 | RBCS WT, EPYC1 WT |
| 93 | 135 | Cre02.g085450.t1.2 | Cre02.g085450.t1.2 | CPX1 | Coproporphyrinogen III oxidase, chloroplast precursor; GI:837299; CPO1; HEM13; HEM6 [PMID: 15849308]; Predicted chloroplast transit peptide 1-31 | 0 | 0 | C | 2.191723273 | -1.5879431 | -1.85404729 | 2.2106494 | -1.56901693 | -1.83512112 | 2 | RBCS WT, EPYC1 WT |
| 94 | 205 | Cre04.g223300.t1.2 | Cre04.g223300.t1.2 | CCP1 | low-CO2-inducible 36kDa chloroplast envelope protein (LIP36-1), CO2-responsive gene; related to mitochondrial substrate carrier proteins; CCP1 and LCIE are located in head-to-head orientation, like nearby CCP2 and LCID; regulated by CCM1 [PMID: 15235119]; involved in acclimation to changing CO2 concentrations | 0 | 1 | O | 0.843723297 | -0.27401358 | -0.92279022 | 2.0639345 | 0.94619764 | 0.297421 | 1 | EPYC1 WT |
| 96 | 195 | Cre06.g261750.t1.2 | Cre06.g261750.t1.2 | BST4 | Similar to Bestrophin, RFP-TM, chloride channel; similar to Arabidopsis AT3G61320 and AT2G45870; part of a small multigene family that includes Low-CO2-inducible membrane protein LCI11 | 1 | 0 | C | 2.631181124 | 4.473028046 | 3.02179269 | 1.4297798 | 3.271626683 | 1.82039133 | 5 | RBCS WT, RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 98 | 88 | Cre17.g720050.t1.2 | Cre17.g720050.t1.2 | FTSH2 | FtsH protease 2, type B; AAA metalloprotease protease, chloroplast; ortholog of Arabidopsis FtsH2=Var2 and FtsH8 | 0 | 0 | C | 1.853064952 | -1.52453539 | -1.06789676 | 2.6480707 | -0.7295296 | -0.27289097 | 2 | RBCS WT, EPYC1 WT |
| 100 | 168 | Cre01.g015350.t1.1 | Cre01.g015350.t1.1 | POR1 | Light-dependent protochlorophyllide reductase, chloroplast precursor; Converts protochlorophyllide to chlorophyllide using NADPH and light as the reductant; Chlamydomonas mutant known as pc-1 has a two-nucleotide deletion within the fourth and fifth codon | 0 | 0 | C | 2.11624253 | -0.21550326 | 0.0530128 | 1.652819 | -0.67892679 | -0.41041073 | 2 | RBCS WT, EPYC1 WT |
| 102 | 124 | 28269764 | TPA_inf: RNA polymerase beta subunit II (chloroplast) [Chlamydomonas reinhardtii] | #N/A | #N/A | 0 | 0 | C | 0.81542978 | -1.59073098 | -0.34497868 | 2.2982688 | -0.107892 | 1.1378603 | 1 | EPYC1 WT |
| 103 | 64 | Cre10.g452800.t1.2 | Cre10.g452800.t1.2 | LCIB | low-CO2 inducible protein; homologous to LCIC with which it forms a complex; regulated by CCM1; LCIB is indispensable for the CO2-concentrating mechanism especially under low-CO2 conditions; localization is diffuse in chloroplast stroma under high-CO2 and low-CO2 conditions, but aggregated around pyrenoid as ring-like structure under very low-CO2 conditions; LCIB expression is also detected under HC conditions; Mutations in this gene include pmp1 (Spalding et al., 1983 [PMID: 16663206]) and ad1 (Wang et al., [PMID:16777959]). Crystal structure of LCIB and LCIC harbor motifs bearing close resemblance to the active site of canonical beta type carbonic anhydrase, suggesting that LCIB/LCIC complex is required for the uptake of external CO2 into the stromal bicarbonate pool and the prevention of CO2 leakage from the pyrenoid. | 1 | 0 | C | 2.841762031 | -0.0674289 | 0.35536313 | 1.8516115 | -1.05757946 | -0.63478742 | 2 | RBCS WT, EPYC1 WT |
| 105 | 155 | Cre06.g270100.t1.1 | Cre06.g270100.t1.1 | SBE2 | catalyses the formation of alpha-1,6 linkages within the polymer by cleaving a preexisting linear chain and transferring the fragment at the non reducing end of the cleaved glucan to an alpha-1,6 position | 0 | 0 | C | 0 | 0 | 0 | 1.4215092 | 2.358652845 | 2.51231336 | 1 | EPYC1 WT |
| 106 | 308 | Cre12.g550850.t1.2 | Cre12.g550850.t1.2 | PSBP1 | Chloroplast precursor (OEE2) (23 kDa subunit of oxygen evolving complex of photosystem II) (OEC 23 kDa subunit) (23 kDa thylakoid membrane protein) (OEC23);PMID: 3468511; targeted to thylakoid lumen by TAT pathway | 0 | 0 | C | 1.483628914 | 1.989277074 | 1.52363188 | 0.1408531 | 0.646501236 | 0.18085605 | 2 | RBCS RPE1, RBCS PRK1 |
| 108 | 40 | Cre12.g519100.t1.2 | Cre12.g519100.t1.2 | ACC1 | Alpha-carboxyltransferase subunit of multimeric plastid enzyme | 0 | 0 | C | 1.421981936 | -1.59918722 | -1.60542413 | 2.1138672 | -0.90730198 | -0.91353889 | 1 | EPYC1 WT |
| 109 | 73 | 28269792 | TPA_inf: chloroplast RNA polymerase beta' chain (chloroplast) [Chlamydomonas reinhardtii] | #N/A | #N/A | 0 | 0 | C | 1.228864102 | -2.16070036 | -0.71418954 | 3.3787991 | -0.01076534 | 1.43574548 | 2 | EPYC1 WT, EPYC1 PRK1 |
| 110 | 145 | Cre05.g245158.t1.1 | Cre05.g245158.t1.1 | - | - | 0 | 0 | C | 2.600656871 | -0.61468556 | -0.22447678 | 2.3540334 | -0.86130903 | -0.47110025 | 2 | RBCS WT, EPYC1 WT |
| 112 | 172 | 28269756 | TPA_inf: RNA polymerase alpha subunit (chloroplast) [Chlamydomonas reinhardtii] | #N/A | #N/A | 0 | 0 | C | 2.288985094 | -1.62839178 | -0.2493822 | 3.5199559 | -0.39742093 | 0.98158865 | 2 | RBCS WT, EPYC1 WT |
| 116 | 83 | Cre03.g201100.t1.2 | Cre03.g201100.t1.2 | FHL1 | Similar to FtsH protease, but lacks peptidase HEXXH motif, like At FtsHi; may function with Tic214 (Ycf2) in the TIC translocon | 0 | 0 | C | 1.510469529 | -1.09766675 | -0.74881319 | 3.2627735 | 0.654637185 | 1.00349003 | 1 | EPYC1 WT |
| 119 | 110 | Cre09.g400550.t1.2 | Cre09.g400550.t1.2 | NOA1 | NO-associated; carries a GTPase domain, homologue in Arabidopsis appears necessary for synthesis of nitric oxide, but does not carry out the reaction (PMID: 17030145) | 0 | 0 | C | 1.786481025 | -0.68824178 | -0.15719806 | 2.5100089 | 0.035286118 | 0.56632984 | 1 | EPYC1 WT |
| 120 | 121 | Cre09.g386758.t1.1 | Cre09.g386758.t1.1 | ALS1 | Acetolactate synthase, large subunit (EC 2.2.1.6); (acetohydroxy acid synthase, large subunit); precursor, expected plastidic by homology, and predicted plastidic by Target-P. Previously annotated as ALSL1. Can be mutated to herbicide resistance. Requi | 0 | 0 | C | 1.990910068 | -0.78926526 | -0.64013098 | 2.4953758 | -0.28479949 | -0.13566522 | 2 | RBCS WT, EPYC1 WT |
| 121 | 119 | Cre07.g335400.t1.1 | Cre07.g335400.t1.1 | - | - | 0 | 0 | C | 1.532105027 | -0.82197115 | -0.39792348 | 1.7576697 | -0.59640647 | -0.1723588 | 1 | EPYC1 WT |
| 125 | 174 | Cre06.g278255.t1.1 | Cre06.g278255.t1.1 | - | - | 0 | 0 | C | 3.012903121 | -0.10379172 | 0.10508414 | 2.7304721 | -0.38622269 | -0.17734683 | 2 | RBCS WT, EPYC1 WT |
| 126 | 136 | Cre17.g734100.t1.2 | Cre17.g734100.t1.2 | - | - | 0 | 0 | C | 3.026535764 | -1.69153309 | -0.77080426 | 2.6625926 | -2.05547622 | -1.13474739 | 2 | RBCS WT, EPYC1 WT |
| 128 | 151 | Cre18.g749497.t1.1 | Cre18.g749497.t1.1 | - | - | 0 | 0 | C | 2.395060178 | -0.31850115 | -0.11511628 | 1.7311257 | -0.98243565 | -0.77905077 | 2 | RBCS WT, EPYC1 WT |
| 130 | 156 | Cre12.g512300.t1.2 | Cre12.g512300.t1.2 | - | - | 0 | 0 | M | 1.10064198 | -0.42121902 | 2.11716842 | 0 | -1.28810509 | 0 | 1 | RBCS PRK1 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 138 | 128 | Cre03.g149100.t1.2 | Cre03.g149100.t1.2 | CIS2 | Citrate synthase (EC 2.3.3.1), glyoxysomal/microbody form; similarity to Arabidopsis citrate synthase glyoxysomal precursor (GenBank Q9LXS6); PMID: 1921471 | 0 | 0 | M | 1.020974129 | -1.72462954 | -0.95347176 | 2.4509818 | -0.29462185 | 0.47653593 | 1 | EPYC1 WT |
| 139 | 150 | Cre02.g099850.t1.1 | Cre02.g099850.t1.1 | PDC2 | pyruvate dehydrogenase (EC 1.2.4.1), E1 component alpha subunit (ODPA, PDC2_E1a). The pyruvate dehydrogenase complex catalyzes the overall conversion of pyruvate to acetyl-CoA and CO(2). It contains multiple copies of three enzymatic components: pyruvate dehydrogenase (E1), dihydrolipoamide acetyltransferase (E2) and lipoamide dehydrogenase (E3) (By similarity). [CATALYTIC ACTIVITY] Pyruvate + [dihydrolipoyllysine-residue acetyltransferase] lipoyllysine = [dihydrolipoyllysine-residue acetyltransferase] S-acetyldihydrolipoyllysine + CO(2). [COFACTOR] Thiamine pyrophosphate. [SUBUNIT] Heterodimer of an alpha and a beta chain (By similarity). [SUBCELLULAR LOCATION] Plastid; chloroplast; This is the second PDC, localized in the plastid stroma, where it provides acetyl-CoA for fatty acid biosynthesis; different from the mitochondrial PDC | 0 | 0 | O | 1.869675507 | -1.63843186 | -1.5312786 | 2.1897594 | -1.31834795 | -1.21119469 | 2 | RBCS WT, EPYC1 WT |
| 145 | 127 | Cre17.g727700.t1.1 | Cre17.g727700.t1.1 | HEL65 | - | 0 | 0 | C | 0.264230313 | 0.929868841 | 1.37733125 | 0.408554 | 1.074192557 | 1.52165496 | 1 | EPYC1 PRK1 |
| 152 | 133 | Cre08.g372950.t1.2 | Cre08.g372950.t1.2 | IDS1 | 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (isopentenyl diphosphate/dimethylallyl diphosphate synthase) (ispH) (lytB) [PMID:11004185]; chloroplast precursor | 0 | 0 | C | 2.762286124 | -0.08865869 | -0.45477404 | 2.8977048 | 0.046760032 | -0.31935532 | 2 | RBCS WT, EPYC1 WT |
| 155 | 113 | Cre03.g158050.t1.1 | Cre03.g158050.t1.1 | - | - | 0 | 0 | C | 2.953536314 | 1.817758503 | 1.40520961 | 2.4842322 | 1.348454389 | 0.9359055 | 4 | RBCS WT, EPYC1 WT, RBCS RPE1, RBCS PRK1 |
| 156 | 435 | Cre12.g484200.t1.2 | Cre12.g484200.t1.2 | GGPS | polyprenyl synthetase (geranylgeranyl diphosphate synthase) (GGPPS); Low-CO2 inducible gene LCI14 revealed by cDNA array analyses; regulated by CCM1 [PMID: 15235119] | 0 | 0 | C | 1.984977803 | -1.59806875 | -1.58354393 | 2.3004836 | -1.28256298 | -1.26803816 | 2 | RBCS WT, EPYC1 WT |
| 159 | 169 | Cre09.g415600.t2.1 | Cre09.g415600.t2.1 | - | - | 0 | 0 | C | 3.943735925 | 0.064800345 | -0.44822616 | 4.2819428 | 0.403007259 | -0.11001925 | 2 | RBCS WT, EPYC1 WT |
| 159 | 170 | Cre09.g415600.t1.1 | Cre09.g415600.t1.1 | - | - | 0 | 0 | C | 3.943735925 | 0.064800345 | -0.44822616 | 4.2819428 | 0.403007259 | -0.11001925 | 2 | RBCS WT, EPYC1 WT |
| 163 | 106 | Cre13.g607900.t1.1 | Cre13.g607900.t1.1 | CPLD68 | - | 0 | 0 | M | 2.453919773 | -0.43256171 | -0.68157988 | 2.7262238 | -0.16025768 | -0.40927585 | 2 | RBCS WT, EPYC1 WT |
| 165 | 199 | Cre05.g234550.t2.1 | Cre05.g234550.t2.1 | FBA3 | Fructose-1,6-bisphosphate aldolase 1 (EC:4.1.2.13); probably plastidic, but predicted mitochondrial by Target-P; 87% identical to cDNA (Q42690); glycolytic enzyme that catalyses the reversible aldol cleavage or condensation of fructose-1,6-bisphosphate into dihydroxyacetone-phosphate and glyceraldehyde 3-phosphate [PMID: 8053679; Genbank entry X69969] | 0 | 1 | C | 1.50935307 | -2.2320128 | -1.88463822 | 2.6148269 | -1.12653896 | -0.77916438 | 1 | EPYC1 WT |
| 165 | 200 | Cre05.g234550.t1.2 | Cre05.g234550.t1.2 | FBA3 | Fructose-1,6-bisphosphate aldolase 1 (EC:4.1.2.13); probably plastidic, but predicted mitochondrial by Target-P; 87% identical to cDNA (Q42690); glycolytic enzyme that catalyses the reversible aldol cleavage or condensation of fructose-1,6-bisphosphate into dihydroxyacetone-phosphate and glyceraldehyde 3-phosphate [PMID: 8053679; Genbank entry X69969] | 0 | 1 | C | 1.50935307 | -2.2320128 | -1.88463822 | 2.6148269 | -1.12653896 | -0.77916438 | 1 | EPYC1 WT |
| 166 | 144 | Cre13.g583550.t1.2 | Cre13.g583550.t1.2 | VIPP1 | Initially thought involved in vesicle formation/transport in the chloroplast, since VIPP1-underexpressing Arabidopsis mutants were abolished in vesicle budding from the inner membrane and had distorted thylakoids; VIPP1 in Chlamy is located to stroma, thylakoids, and low density membranes; it forms ring-like oligomers, which also assemble into rod-like structures; these are disassembled by the HSP70B-CDJ2-CGE1 chaperones; VIPP1 is overexpressed in Alb3.2 mutants [PMID: 15635096, 16679460] | 0 | 0 | C | 2.222262856 | -0.71414499 | -0.79241704 | 2.3937784 | -0.54262947 | -0.62090152 | 2 | RBCS WT, EPYC1 WT |
| 167 | 178 | Cre08.g360050.t1.1 | Cre08.g360050.t1.1 | DUR1 | urea carboxylase [EC:6.3.4.]; Allophanate hydrolase subunit 2; catalyses first reaction of ATP:urea amidolyase; homologous to C-terminus of yeast protein, whose N-terminus is homologous to DUR2, upstream divergent; in green algae, the two activities are separated at the protein level | 0 | 0 | O | 1.047252744 | 1.958847309 | 2.20659072 | -0.1377239 | 0.773870705 | 1.02161412 | 1 | RBCS PRK1 |
| 167 | 179 | Cre08.g360050.t2.1 | Cre08.g360050.t2.1 | DUR1 | urea carboxylase [EC:6.3.4.]; Allophanate hydrolase subunit 2; catalyses first reaction of ATP:urea amidolyase; homologous to C-terminus of yeast protein, whose N-terminus is homologous to DUR2, upstream divergent; in green algae, the two activities are separated at the protein level | 0 | 0 | O | 1.047252744 | 1.958847309 | 2.20659072 | -0.1377239 | 0.773870705 | 1.02161412 | 1 | RBCS WT |
| 170 | 116 | Cre08.g365692.t1.1 | Cre08.g365692.t1.1 | SIR4 | May contain additional coding sequence in intronic N-island PMID: 27175017 | 0 | 0 | C | 1.906597756 | -1.42252025 | -1.60525271 | 2.2939434 | -1.03517465 | -1.21790711 | 2 | RBCS WT, EPYC1 WT |
| 174 | 84 | Cre04.g214500.t1.1 | Cre04.g214500.t1.1 | IDH3 | NADP specific isocitrate dehydrogenase (EC 1.1.1.42), mitochondrial precursor; similar to mammalian mitochondrial ICDH (e.g., GenBank XP_536192), predicted by Target-P to have an organellar targeting sequence (mitochondrial) | 0 | 0 | M | 1.101132714 | 0.791997398 | 1.71687603 | 0 | 0.138879552 | 0 | 1 | RBCS PRK1 |
| 180 | 175 | Cre03.g206600.t1.2 | Cre03.g206600.t1.2 | AAD1 | acetohydroxyacid dehydratase (EC 4.2.1.9); ILVD; dihydroxy-acid dehydrase; probable plastid location, based on homology and on Target-P prediction | 0 | 0 | C | 2.247312309 | 0.244764552 | 0.86771455 | 3.5404174 | 1.537869605 | 2.1608196 | 3 | RBCS WT, EPYC1 WT, EPYC1 PRK1 |
| 184 | 245 | Cre06.g252550.t1.1 | Cre06.g252550.t1.1 | DLA3 | Dihydrolipoamide acetyltransferase; E2 component of pyruvate dehydrogenase complex. Independent sequence at AY596299. | 0 | 0 | M | 1.261702259 | 2.566559812 | 1.41392626 | 2.9132897 | 4.218147261 | 3.06551371 | 5 | EPYC1 WT, RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 188 | 153 | Cre06.g278221.t1.1 | Cre06.g278221.t1.1 | - | Trehalose-6-phosphate synthase/phosphatase, a glycosyl transferase | 0 | 0 | M | -0.409462997 | -0.47328186 | 0.92124925 | 0.1155633 | 0.051744446 | 1.44627555 | 1 | EPYC1 PRK1 |
| 190 | 114 | Cre02.g080200.t1.2 | Cre02.g080200.t1.2 | TRK1 | - | 0 | 0 | C | 2.462634104 | -0.17472168 | 0.43235851 | 1.6856725 | -0.95168333 | -0.34460314 | 1 | RBCS WT |
| 191 | 85 | Cre13.g578650.t1.1 | Cre13.g578650.t1.1 | CPLD10 | Arabidopsis homolog is involved in psbA translation initiation; similar to Mitochondrial complex I intermediate-associated protein 30 and NmrA negative transcriptional regulator protein family. Conserved in the plant lineage and Diatoms | 0 | 0 | C | 2.156234015 | -0.65325844 | -0.5651503 | 2.4075552 | -0.40193728 | -0.31382914 | 2 | RBCS WT, EPYC1 WT |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 195 | 224 | Cre07.g318750.t1.2 | Cre07.g318750.t1.2 | PURM | PurM (Aminoimidazole Ribonucleotide [AIR] synthetase), one of eleven enzymes required for purine biosynthesis; catalyzes the conversion of formylglycinamide ribonucleotide (FGAM) and ATP to AIR, ADP, and Pi, the fifth step in de novo purine biosynthesis; | 0 | 0 | C | 3.137588428 | -1.5453283 | -1.27166405 | 3.3781052 | -1.30481151 | -1.03114727 | 2 | RBCS WT, EPYC1 WT |
| 196 | 534 | 28269786 | TPA_inf: photosystem II reaction center protein D2 (chloroplast) [Chlamydomonas reinhardtii] | #N/A | #N/A | 0 | 0 | C | 0.687778238 | -0.14860516 | 1.92594935 | 0 | -1.62120013 | 0 | 1 | RBCS PRK1 |
| 199 | 98 | Cre03.g199000.t1.2 | Cre03.g199000.t1.2 | PHOT1 | Blue-light receptor; found in flagella; regulates many biological processes including gametogenesis, chemotaxis, zygote maturation, light-stress response; contains two PAS and two protein kinase domains | 0 | 0 | C | 2.033323827 | 0.953842438 | 1.43898092 | 2.4666354 | 1.387153987 | 1.87229247 | 4 | RBCS WT, EPYC1 WT, RBCS PRK1, EPYC1 PRK1 |
| 200 | 166 | Cre01.g032650.t2.1 | Cre01.g032650.t2.1 | TAL1 | EC 2.2.1.2; possible cytosolic isoform, although weak predicted organelle targeting by Target-P; catalyzes the reversible transfer of a three-carbon ketol unit from sedoheptulose 7-phosphate to glyceraldehyde 3-phosphate to form erythrose 4-phosphate and fructose 6-phosphate; this enzyme, together with transketolase, provides a link between the glycolytic and pentose-phosphate pathways | 0 | 0 | C | 0.712143436 | -1.29149537 | 1.95031455 | 1.1095877 | -0.89405111 | 2.26892567 | 1 | EPYC1 PRK1 |
| 200 | 165 | Cre01.g032650.t1.2 | Cre01.g032650.t1.2 | TAL1 | EC 2.2.1.2; possible cytosolic isoform, although weak predicted organelle targeting by Target-P; catalyzes the reversible transfer of a three-carbon ketol unit from sedoheptulose 7-phosphate to glyceraldehyde 3-phosphate to form erythrose 4-phosphate and fructose 6-phosphate; this enzyme, together with transketolase, provides a link between the glycolytic and pentose-phosphate pathways | 0 | 0 | C | 0.712143436 | -1.29149537 | 1.95031455 | 1.1095877 | -0.89405111 | 2.26892567 | 1 | RBCS PRK1 |
| 203 | 120 | Cre15.g643600.t1.2 | Cre15.g643600.t1.2 | SUFB | ABC-type system involved in Fe-S cluster assembly, likely localized to plastid | 0 | 0 | C | 2.078705092 | -0.96372497 | -0.0668624 | 1.9049637 | -1.13746636 | -0.24060379 | 2 | RBCS WT, EPYC1 WT |
| 205 | 420 | Cre09.g393200.t1.2 | Cre09.g393200.t1.2 | HSP70C | HSP70C is one of 9 genes encoding HSP70-like proteins in Chlamydomonas; HSP70C is localized to mitochondria; it is inducible by heat shock and light; PMID: 2779571, 16143837 | 0 | 0 | M | 1.193392011 | 2.402131127 | 1.20175732 | 0 | 0 | 0 | 2 | RBCS RPE1, RBCS PRK1 |
| 211 | 246 | 28269728 | TPA_inf: light-independent protochlorophyllide reductase subunit B (chloroplast) [Chlamydomonas reinhardtii] | #N/A | #N/A | 0 | 0 | C | 1.437260523 | -0.65104125 | -0.8428826 | 1.8918831 | -0.19641864 | -0.38826 | 1 | EPYC1 WT |
| 215 | 320 | Cre12.g520200.t1.1 | Cre12.g520200.t1.1 | CPL8 | Pyridoxamine 5'-phosphate oxidase-like protein, but enzymatic activity is unproven | 0 | 0 | C | 1.669608551 | -1.55528485 | -1.0685289 | 2.1492694 | -1.07562397 | -0.58886802 | 1 | EPYC1 WT |
| 218 | 158 | Cre07.g325736.t1.1 | Cre07.g325736.t1.1 | - | - | 0 | 0 | C | 1.683205993 | -0.12414893 | -0.6125908 | 2.15998 | 0.352625065 | -0.13581687 | 1 | EPYC1 WT |
| 230 | 236 | Cre11.g467400.t1.2 | Cre11.g467400.t1.2 | - | - | 0 | 0 | C | 2.856933742 | -1.41462995 | -2.6504815 | 3.7006347 | -0.57092898 | -1.80678055 | 2 | RBCS WT, EPYC1 WT |
| 234 | 252 | Cre12.g485800.t1.2 | Cre12.g485800.t1.2 | FTSH1 | FtsH protease 1, type A; chloroplast-targeted; membrane-bound metalloprotease (M41) with AAA domain; ortholog of Arabidopsis FtsH1=Var1 and FtsH5 | 0 | 0 | C | 2.302657076 | 0.043784796 | 0.3792431 | 1.9419158 | -0.3169565 | 0.01850185 | 2 | RBCS WT, EPYC1 WT |
| 235 | 209 | Cre12.g553250.t1.2 | Cre12.g553250.t1.2 | PFK2 | Phosphofructokinase family protein (EC 2.7.1.11 or 2.7.1.90), probably plastid targeted | 0 | 0 | C | 0.82556559 | -1.586072 6 | -0.9247129 7 | 1.9065978 | -0.5050443 | 0.15631919 | 1 | EPYC1 WT |
| 248 | 276 | Cre17.g702200.t1.2 | Cre17.g702200.t1.2 | ANK29 | protein of unknown function with ankyrin repeats | 0 | 0 | C | 1.85385164 | -0.26658932 | -0.4793729 9 | 1.9785 | -0.14194097 | -0.35472464 | 2 | RBCS WT, EPYC1 WT |
| 249 | 341 | Cre06.g278750.t1.2 | Cre06.g278750.t1.2 | UCP1 | Belong to mitochondrial substrate carrier family | 0 | 0 | M | 1.979983769 | 0.303451969 | -0.5603262 | 1.7980812 | 0.121549384 | -0.74222879 | 2 | RBCS WT, EPYC1 WT |
| 253 | 251 | Cre01.g031100.t1.2 | Cre01.g031100.t1.2 | TEF30 | - | 0 | 0 | C | 3.199171182 | 0.715780218 | 0.1595307 8 | 3.0517429 | 0.568351964 | 0.01210253 | 2 | RBCS WT, EPYC1 WT |
| 265 | 244 | Cre08.g384750.t1.1 | Cre08.g384750.t1.1 | AMA3 | Belongs to glycoside hydrolase family 13; catalyses hydrolysis of (1-4)-alpha-D-glucosidic linkages in polysaccharides so as to remove successive alpha-maltose residues from the non-reducing ends of the chains in the conversion of starch to maltose | 0 | 0 | C | 1.786671183 | 0.650872946 | 3.0248423 | 1.7716991 | 0.635900817 | 2.93103703 | 3 | EPYC1 WT, RBCS PRK1, EPYC1 PRK1 |
| 277 | 344 | Cre04.g223850.t1.2 | Cre04.g223850.t1.2 | HEL23 | Cytoplasmic DExD/H-box helicase, stimulates mRNA 5'-decapping. Homologue of S. cerevisiae Dhh1. Might be involved in the translational control of some specific mRNAs. | 0 | 0 | O | 0.865098478 | 1.406786045 | 0.9601234 6 | 0.6573054 | 1.198992962 | 0.75233038 | 1 | RBCS RPE1 |
| 278 | 442 | Cre10.g446500.t1.1 | Cre10.g446500.t1.1 | CPL24 | Conserved in photosynthetic organisms; contains pfam07534: TLD domain; null-allele mutant was isolated (PMID 29743196) | 0 | 0 | C | 3.347280855 | 0.643068104 | 0.0940623 | 3.4360892 | 0.731876409 | 0.18287061 | 2 | RBCS WT, EPYC1 WT |
| 286 | 529 | Cre12.g534700.t1.2 | Cre12.g534700.t1.2 | FKB19 | FKBP-type peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8) (PPIase) (Rotamase) (Immunophilin); probably directed to thylakoid lumen (homology to AtFKBP19 and bipartite cTP with RR motif) | 0 | 0 | SP | 2.09112731 | 2.289943957 | 2.6648671 6 | 0 | 0 | 0.70100701 | 3 | RBCS WT, RBCS RPE1, RBCS PRK1 |
| 287 | 396 | Cre12.g497300.t2.1 | Cre12.g497300.t2.1 | CAS1 | Present in thylakoid-enriched fraction based on mass spectrometric peptide identification [PMID: 17078018]. CAS is required for the expression of HLA3 and LCIA which are cooperatively associated with bicarbonate uptake for CO2-concentraing mechanism [PMID:24549931, 27791081]; CAS localization displays a mesh-like structure in the chloroplast under high-CO2 conditions, but CAS gathers inside the pyrenoid as a distinct wheel-like structure at 2 h after transfer to low-CO2 condition, and then localized to the center of the pyrenoid at 12 h [PMID: 29372336]. | 1 | 0 | C | -0.823519551 | 2.751918285 | 0.5989546 5 | -1.144491 | 2.430946884 | 0.27798325 | 1 | RBCS RPE1 |
| 287 | 397 | Cre12.g497300.t1.2 | Cre12.g497300.t1.2 | CAS1 | Present in thylakoid-enriched fraction based on mass spectrometric peptide identification [PMID: 17078018]. CAS is required for the expression of HLA3 and LCIA which are cooperatively associated with bicarbonate uptake for CO2-concentraing mechanism [PMID:24549931, 27791081]; CAS localization displays a mesh-like structure in the chloroplast under high-CO2 conditions, but CAS gathers inside the pyrenoid as a distinct wheel-like structure at 2 h after transfer to low-CO2 condition, and then localized to the center of the pyrenoid at 12 h [PMID: 29372336]. | 0 | 0 | C | -0.823519551 | 2.751918285 | 0.5989546 5 | -1.144491 | 2.430946884 | 0.27798325 | 1 | RBCS RPE1 |
| 290 | 338 | Cre12.g523300.t1.2 | Cre12.g523300.t1.2 | GTR22 | similar to ribophorin I, an essential subunit of oligosaccharyltransferase (OST), a glycosyl transferase which is also known as dolichyl-diphosphooligosaccharide--protein glycosyltransferase [EC 2.4.1.119]; | 0 | 0 | SP | 0.507093619 | -0.35977348 | -0.9411936 2 | 1.9290881 | 1.062221 03 | 0.48080089 | 1 | EPYC1 WT |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | catalyses the transfer of an oligosaccharide from dolichol pyrophosphate to selected asparagine residues of nascent polypeptides | | | | | | | | | | | |
| 292 | 379 | Cre01.g054150.t1.2 | Cre01.g054150.t1.2 | NTRC1 | NADPH-dependent thioredoxin reductase C type (NTR/thioredoxin fusion); NTRC is a two domain protein containing an NADPH dependent thioredoxin reductase (N-terminal part) and a thioredoxin (C-terminal part). The protein was shown to be localized in chloroplasts in rice and appears to protect from abiotic stress in Arabidopsis [PMID: 15292215]. | 0 | 0 | C | 2.205917422 | 0.367514688 | 0.74948539 | 1.5192695 | -0.31913322 | 0.06283748 | 1 | RBCS WT |
| 293 | 241 | Cre06.g298350.t1.2 | Cre06.g298350.t1.2 | FAP224 | Conserved uncharacterized Flagellar Associated Protein; found in the flagellar proteome; may the target of a miRNA directing cleavage in the extended 3' UTR, after the main polyadenylation site, in a region that appears to be deleted in strain S1D2 | 0 | 0 | O | 1.191436046 | -0.32859615 | 0.27744983 | 2.216825 | 0.696792788 | 1.30283877 | 1 | EPYC1 WT |
| 294 | 345 | Cre08.g365632.t1.1 | Cre08.g365632.t1.1 | - | - | 0 | 0 | C | 1.770579542 | 0.340525662 | -0.7932252 | 1.9430709 | 0.513017032 | -0.62073383 | 1 | EPYC1 WT |
| 295 | 458 | Cre43.g760497.t1.1 | Cre43.g760497.t1.1 | - | - | 0 | 0 | C | -0.179074974 | -2.39048009 | -0.01328706 | 1.7820369 | -0.42936824 | 1.94782479 | 1 | EPYC1 WT |
| 296 | 778 | 28269757 | TPA_inf: ribosomal protein S2 (chloroplast) [Chlamydomonas reinhardtii] | #N/A | #N/A | 0 | 0 | C | 0.093135318 | -1.93462197 | -1.56446279 | 2.4285536 | 0.400796295 | 0.77095548 | 1 | EPYC1 WT |
| 299 | 424 | Cre06.g260350.t1.2 | Cre06.g260350.t1.2 | DNJ6 | DnaJ-like protein; similar to E. coli Caj1p and to presumably cytosolic Arabidopsis Q9T24 and Q9SJS8; contains long N-terminal extension, which is absent from Arabidopsis homologs; TargetP gives no clear prediction for DNJ6; contains N-terminal J-domain | 0 | 0 | O | -0.183721324 | 1.389312089 | 0.85738714 | 0.0250079 | 1.598041267 | 1.06611632 | 1 | RBCS RPE1 |
| 304 | 436 | Cre16.g691000.t1.2 | Cre16.g691000.t1.2 | EFP1 | Putative organellar translation elongation factor P (EF-P) | 0 | 0 | C | 2.486711053 | -0.51335149 | -0.05522787 | 1.0360991 | -1.96396348 | -1.50583986 | 1 | RBCS WT |
| 318 | 483 | Cre05.g234637.t1.1 | Cre05.g234637.t1.1 | - | - | 0 | 0 | O | 1.895779151 | -0.19636724 | -0.09169859 | 2.1578791 | 0.06573267 | 0.17040132 | 2 | RBCS WT, EPYC1 WT |
| 322 | 449 | Cre12.g498550.t1.1 | Cre12.g498550.t1.1 | CHLM | Magnesium protoporphyrin IX S-adenosyl methionine O-methyl transferase, chloroplast precursor (Magnesium-protoporphyrin IX methyltransferase) (PPMT); Magnesium-protoporphyrin O-methyltransferase (chlM) [PMID: 12828371; PMID: 12489983; PMID: 4436384]; ChloroP 1.1 predicts chloroplast location | 0 | 0 | C | 1.070971114 | 1.404559994 | 2.30914223 | 0 | 0 | 0 | 1 | RBCS PRK1 |
| 323 | 486 | Cre09.g401515.t1.1 | Cre09.g401515.t1.1 | - | - | 0 | 0 | C | 0 | 0 | 0 | 0.1454779 | 1.718511347 | -0.45975033 | 1 | EPYC1 RPE1 |
| 326 | 512 | Cre13.g603500.t1.2 | Cre13.g603500.t1.2 | - | - | 0 | 0 | C | 4.341942238 | -0.58794279 | -0.34318633 | 4.8136244 | -0.11626059 | 0.1284959 | 2 | RBCS WT, EPYC1 WT |
| 338 | 353 | Cre12.g528000.t1.2 | Cre12.g528000.t1.2 | FAP303 | Similar 5'-AMP-activated protein kinase gamma subunit | 0 | 0 | M | 0 | 0 | 0 | 1.1106008 | 1.375375355 | 2.26993873 | 1 | EPYC1 PRK1 |
| 340 | 373 | Cre14.g621351.t1.1 | Cre14.g621351.t1.1 | - | - | 0 | 0 | SP | 1.251910691 | 0.174603985 | -0.36377346 | 1.8351582 | 0.757851471 | 0.21947403 | 1 | EPYC1 WT |
| 352 | 233 | Cre13.g562150.t1.2 | Cre13.g562150.t1.2 | - | - | 0 | 0 | C | 1.858502479 | -0.89099586 | -0.43811393 | 1.303499 | -1.44599936 | -0.99311743 | 1 | RBCS WT |
| 362 | 280 | Cre12.g483650.t1.2 | Cre12.g483650.t1.2 | - | - | 0 | 0 | O | 0.297456156 | 1.700295375 | 1.17587786 | 0 | 0 | 0 | 2 | RBCS RPE1, RBCS PRK1 |
| 367 | 337 | Cre12.g498500.t1.2 | Cre12.g498500.t1.2 | DEG11 | DegP-type protease, thylakoid lumen; one PDZ domain; similar to At DegP1; v5.5 model is truncated at N-term | 0 | 0 | M | 0.045691537 | 1.753902711 | 1.11998634 | 0 | 0 | 0 | 2 | RBCS RPE1, RBCS PRK1 |
| 375 | 506 | Cre11.g467749.t1.1 | Cre11.g467749.t1.1 | - | - | 0 | 0 | C | 2.284772843 | 0.709759224 | 0.38091902 | 2.9655404 | 1.390526819 | 1.06168661 | 2 | RBCS WT, EPYC1 WT |
| 383 | 318 | Cre09.g391650.t1.2 | Cre09.g391650.t1.2 | HCP4 | Prismane/CO dehydrogenase family | 0 | 0 | C | 0.684244218 | -0.96682364 | 1.76101432 | 1.6345413 | -0.01652654 | 2.71131143 | 2 | EPYC1 WT, RBCS PRK1 |
| 433 | 439 | Cre06.g269650.t1.2 | Cre06.g269650.t1.2 | - | - | 0 | 0 | C | 0.764435462 | 2.05977061 | 1.31829928 | 0.8079074 | 2.10324251 | 1.36177118 | 2 | RBCS RPE1, EPYC1 RPE1 |
| 442 | 598 | Cre04.g221250.t1.2 | Cre04.g221250.t1.2 | - | - | 0 | 0 | C | 1.897646155 | -0.67771709 | -0.60412383 | 1.3170786 | -1.25833861 | -1.18469135 | 1 | RBCS WT |
| 445 | 968 | Cre02.g078300.t1.1 | Cre02.g078300.t1.1 | GPX1 | Gene model is truncated at N-terminus due to UGA Selenocystein codon; add sequence MSLLGNLFGGGSKPTSSTSNFHQLSALDIDKKNVDFKSLNNRVVLVVNVASKUGLTAANYKEFATLLGKY PATDLTIVAFPCNQFGGQEPGTNAEIKAFASARGFSGAGALL with a Selenocysteine in position 53; most similar to plant homologues, that have a Cys at the corresponding position; contains a SECIS element in 3'-UTR | 0 | 0 | O | 1.532407448 | -0.69804188 | -1.57788946 | 1.9995978 | -0.23085153 | -1.11069911 | 1 | EPYC1 WT |
| 460 | 482 | Cre03.g199647.t1.1 | Cre03.g199647.t1.1 | - | Forms the core of the Exon Junction Complex which associates to mRNAs upstream of the splicing sites and accompanies it until the first round of translation | 0 | 0 | O | 1.101763233 | 1.767166947 | 1.00319256 | 0.0311437 | 0.696547375 | -0.06742701 | 1 | RBCS RPE1 |
| 484 | 472 | Cre16.g663150.t1.2 | Cre16.g663150.t1.2 | STR18 | - | 0 | 0 | M | 2.06810578 | 3.776316953 | 3.30627689 | 0 | 0 | 0 | 3 | RBCS WT, RBCS RPE1, RBCS PRK1 |
| 512 | 456 | Cre06.g271850.t1.2 | Cre06.g271850.t1.2 | ABCF1 | - | 0 | 0 | C | 0.490449923 | 0.733648336 | 1.72862104 | 0 | 1.469915022 | 0 | 1 | RBCS PRK1 |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 542 | 820 | 28269742 | TPA_inf: ribosomal protein S8 (chloroplast) [Chlamydomonas reinhardtii] | #N/A | #N/A | 0 | 0 | C | 0.478261311 | -0.24796249 | 1.71643243 | 0 | -1.90156913 | 0 | 1 | RBCS PRK1 |
| 546 | 921 | Cre10.g430350.t1.2 | Cre10.g430350.t1.2 | - | - | 1 | 0 | C | 0.901046112 | 2.241862983 | 2.13921723 | 1.5408496 | 2.881666446 | 2.70018755 | 4 | RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 554 | 540 | Cre14.g624201.t1.1 | Cre14.g624201.t1.1 | - | - | 0 | 0 | M | 1.042589312 | 1.679639487 | 2.28076043 | 0 | 0 | 0 | 1 | RBCS PRK1 |
| 556 | 610 | Cre10.g459200.t1.2 | Cre10.g459200.t1.2 | PMA2 | Putative plasma membrane-type proton ATPase GI:16751320; Low-CO2 inducible gene revealed by cDNA array analyses; regulated by CCM1 [PMID: 15235119] | 0 | 0 | O | 0.059938709 | 1.801877989 | 0.09574647 | 0 | 0 | 0 | 1 | RBCS RPE1 |
| 565 | 683 | Cre06.g278236.t1.1 | Cre06.g278236.t1.1 | - | null-allele passenger mutation was isolated (PMID 29743196) | 0 | 0 | M | 2.976278665 | 3.291620519 | 2.11079791 | 1.8986435 | 2.213985379 | 1.03316277 | 5 | RBCS WT, EPYC1 WT, RBCS RPE1, RBCS PRK1, EPYC1 RPE1 |
| 604 | 913 | Cre09.g393358.t1.1 | Cre09.g393358.t1.1 | - | - | 0 | 0 | O | 2.632356326 | 2.506238849 | 1.48348132 | 2.7540761 | 2.627958638 | 1.60520111 | 4 | RBCS WT, EPYC1 WT, EPYC1 PRK1, EPYC1 RPE1 |
| 625 | 709 | Cre05.g237000.t1.2 | Cre05.g237000.t1.2 | MAP1D | Involved in N-terminal maturation of proteins; Metallopeptidase family M24; probably targeted to both chloroplast and mitochondrial [PMID: 11060042, 14532271] | 0 | 0 | C | 1.853246533 | -0.01547125 | 0.25869536 | 0.4542658 | -1.41445193 | -1.14028533 | 1 | RBCS WT |
| 658 | 484 | Cre11.g481500.t1.2 | Cre11.g481500.t1.2 | HIS7 | glutamine amidotransferase/cyclase involved in histidine biosynthesis; imidazole glycerol phosphate synthase (IGPS).May contain additional coding sequence in intronic N-island PMID: 27175017 | 0 | 0 | C | 1.012554427 | 0.309919622 | 1.85000602 | 0 | -1.87798013 | 0 | 1 | RBCS PRK1 |
| 681 | 720 | Cre03.g190281.t1.1 | Cre03.g190281.t1.1 | - | - | 0 | 0 | O | 0.011317586 | 1.819606851 | 1.71662816 | 0 | 0 | 0 | 2 | RBCS RPE1, RBCS PRK1 |
| 717 | 602 | Cre12.g541750.t1.1 | Cre12.g541750.t1.1 | ZPD1 | - | 0 | 0 | M | -0.025834099 | 1.512102544 | 0.66417295 | 0 | 0 | 0 | 1 | RBCS RPE1 |
| 770 | 894 | Cre08.g362900.t1.1 | Cre08.g362900.t1.1 | PSBP4 | shows similarity to PSBP (OEE2, oxygen-evolving enhancer protein 2 of photosystem II, OEC23, 23 kDa subunit of oxygen evolving complex : PMID: 3468511); orthologous to At4g15510; targeted to thylakoid lumen by TAT pathway | 1 | 0 | C | 1.354571875 | 3.316741918 | 2.84670186 | 0 | 0 | 0 | 1 | RBCS PRK1 |
| 779 | 775 | Cre04.g216950.t1.2 | Cre04.g216950.t1.2 | - | - | 0 | 0 | C | 0.421267676 | 1.50967374 | 1.65943879 | 0 | 0 | 0 | 1 | RBCS PRK1 |
| 788 | 974 | Cre14.g615050.t1.2 | Cre14.g615050.t1.2 | - | - | 0 | 0 | O | 0.406651516 | 1.365591698 | -0.27963262 | 0 | 0 | 0 | 1 | RBCS RPE1 |
| 844 | 1138 | Cre01.g054700.t1.2 | Cre01.g054700.t1.2 | - | - | 0 | 0 | M | 0.607526432 | 2.315737606 | 1.84569755 | 0 | 0 | 0 | 1 | RBCS PRK1 |
| 868 | 1116 | Cre13.g573250.t1.2 | Cre13.g573250.t1.2 | STR16 | - | 0 | 0 | C | 0.376600842 | 2.084812016 | 1.61477196 | 0.9050221 | 2.478055505 | 2.06436007 | 4 | RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 888 | 1143 | Cre03.g169900.t1.2 | Cre03.g169900.t1.2 | - | - | 0 | 0 | C | 0 | 0 | 0 | 0.9760008 | 2.549034171 | 2.13533874 | 2 | EPYC1 PRK1, EPYC1 RPE1 |
| 965 | 1196 | Cre24.g755497.t1.1 | Cre24.g755497.t1.1 | - | - | 0 | 0 | O | 0.010050573 | 1.064055853 | 1.24822169 | 0 | 0 | 0 | 1 | RBCS PRK1 |
| 1037 | 1112 | Cre12.g556250.t1.2 | Cre12.g556250.t1.2 | - | Homolog of septin molecules used in cytokinesis | 0 | 0 | O | -0.010117172 | 2.195018668 | 1.15654159 | 0 | 0 | 0 | 1 | RBCS RPE1 |

**Supplemental Data Set S6**. High-confidence pyrenoid proxiome. The HC-pyrenoid proxiome is generated by comparing the significantly enriched proteins in pyrenoid vs stromal control comparison groups. Only proteins which are significantly enriched in at least two comparisons are included. The number of appearances for each protein across comparison groups is denoted under "occurence". The associated Log2FC in each comparison group is also listed.

| Protein Group | Protein ID | Accession | Description | Gene name | Anotations | TP | FP | PredAlgo | $Log_2FC$ (RBCS2 vs Wildtype) | $Log_2FC$ (RBCS2 vs RPE1) | $Log_2FC$ (RBCS2 vs PRK1) | $Log_2FC$ (EPYC1 vs Wildtype) | $Log_2FC$ (EPYC1 vs RPE1) | $Log_2FC$ (EPYC1 vs PRK1) | Occurences | Present In |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 17 | 10 | Cre09.g394621.t2.1 | Cre09.g394621.t2.1 | SAGA2 | - | 1 | 0 | C | 5.2021542 | 4.25812113 | 3.6437226 | 4.613178263 | 3.6691452 | 3.05474667 | 4 | RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 868 | 1116 | Cre13.g573250.t1.2 | Cre13.g573250.t1.2 | STR16 | - | 0 | 0 | C | 0.37660084 | 2.084812012 | 1.614771916 | 0.90502209 | 2.4780555 | 2.06436007 | 4 | RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 22 | 185 | Cre10.g436550.t1.2 | Cre10.g436550.t1.2 | EPYC1 | low-complexity repeat protein, probably serving as a linker between RuBisCO proteins in the pyrenoid matrix; regulated by CCM1 [PMID: 15235119]; phopshorylated under low CO2; essential for pyrenoid formation | 1 | 0 | C | 4.02140225 | 4.116820343 | 4.432768669 | 3.33486746 | 3.43028555 | 3.7462339 | 4 | RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 96 | 195 | Cre06.g261750.t1.2 | Cre06.g261750.t1.2 | BST4 | Similar to Bestrophin, RFP-TM, chloride channel; similar to Arabidopsis AT3G61320 and AT2G45870; part of a small multigene family that includes Low-CO2-inducible membrane protein LCI11 | 1 | 0 | C | 2.63118112 | 4.473028055 | 3.021792669 | 1.42977976 | 3.27162668 | 1.82039133 | 4 | RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 23 | 24 | Cre11.g467712.t1.1 | Cre11.g467712.t1.1 | SAGA1 | - | 1 | 0 | M | 3.86985578 | 2.630592626 | 2.622209838 | 3.28462417 | 2.04536099 | 2.03698937 | 4 | RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 184 | 245 | Cre06.g252550.t1.1 | Cre06.g252550.t1.1 | DLA3 | Dihydrolipoamide acetyltransferase; E2 component of pyruvate dehydrogenase complex. Independent sequence at AY596299. | 0 | 0 | M | 1.26170226 | 2.566559861 | 1.413926226 | 2.91328971 | 4.218147216 | 3.06551371 | 4 | RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 29 | 29 | Cre10.g435800.t1.2 | Cre10.g435800.t1.2 | CSP41B | Cleaves RNA in vitro, but appears necessary for stabilizing Cp mRNA in Arabidopsis and spinach; Arabidopsis homolog binds multiple mRNA and rRNA and interacts with Csp41a in multimeric complexes; also found associated with ribosomes and RNA polymerase; cp ribosome-associated protein RAP38; similar to NAD-dependent nucleoside-diphosphate-sugar epimerase/dehydratase | 1 | 0 | C | 2.77329747 | 4.885544153 | 5.347800733 | 0.8800006 | 2.99224723 | 3.45450387 | 4 | RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 40 | 43 | Cre17.g718950.t1.2 | Cre17.g718950.t1.2 | RCA2 | Related to RuBisCO activase (RCA); catalyzes the activation of RuBisCO (ribulose-1,5-bisphosphate carboxylase/oxygenase; EC 4.1.1.39); via the ATP-dependent carboxylation of the epsilon-amino group of lysine. | 1 | 0 | O | 4.63070132 | 4.831344949 | 4.706866538 | 3.87960681 | 4.080250339 | 3.95577198 | 4 | RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 52 | 53 | Cre10.g440050.t1.2 | Cre10.g440050.t1.2 | CSP41A | Homologous to CSP41b; Arabidopsis homologs form multimeric complexes and bind multiple mRNA and rRNA; similar to NAD-dependent nucleoside-diphosphate-sugar epimerase/dehydratase | 1 | 0 | C | 5.62907259 | 7.731127174 | 7.261087008 | 3.88370872 | 5.850585551 | 5.43689008 | 4 | RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 9 | 60 | 28269772 | TPA_inf: ribulose bisphosphate carboxylase large subunit (chloroplast) [Chlamydomonas reinhardtii] | rbcL | #N/A | 1 | 0 | C | 5.01448182 | 6.550856662 | 8.969910733 | 3.16810945 | 4.704484244 | 7.12353835 | 4 | RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 11 | 7 | Cre06.g259100.t1.1 | Cre06.g259100.t1.1 | SAGA3 | - | 0 | 0 | C | 3.73304627 | 2.674747377 | 2.458432666 | 3.21590117 | 2.157602217 | 1.94128757 | 4 | RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 546 | 921 | Cre10.g430350.t1.2 | Cre10.g430350.t1.2 | - | - | 1 | 0 | C | 0.90104611 | 2.241862928 | 1.392172323 | 1.54084957 | 2.881666455 | 2.70018755 | 4 | RBCS RPE1, RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 24 | 39 | Cre04.g229300.t1.1 | Cre04.g229300.t1.1 | RCA1 | RubisCO activase, chloroplast precursor; required for optimal photosynthesis in a low CO2 atmosphere [PMID: 16667924; independent cDNA sequence: M62962]. | 1 | 0 | C | 2.96910025 | 2.059120355 | 2.769358441 | 2.55875871 | 1.648778881 | 2.35901687 | 3 | RBCS RPE1, RBCS PRK1, EPYC1 PRK1 |
| 565 | 683 | Cre06.g278236.t1.1 | Cre06.g278236.t1.1 | - | null-allele passenger mutation was isolated (PMID 29743196) | 0 | 0 | M | 2.97627866 | 3.291620522 | 2.110797911 | 1.89864352 | 2.213985318 | 1.03316277 | 3 | RBCS RPE1, RBCS PRK1, EPYC1 RPE1 |
| 56 | 76 | Cre09.g386735.t1.1 | Cre09.g386735.t1.1 | DLA1 | Dihydrolipoamide acetyltransferase, probably mitochondrial; EC 2.3.1.12); E2 component of pyruvate dehydrogenase complex; null-allele mutant was isolated (PMID 29743196) | 0 | 0 | M | 3.53037894 | 0.957243606 | 1.672875111 | 5.49432796 | 2.921192639 | 3.636824144 | 3 | RBCS PRK1, EPYC1 PRK1, EPYC1 RPE1 |
| 155 | 113 | Cre03.g158050.t1.1 | Cre03.g158050.t1.1 | - | - | 0 | 0 | C | 2.95353631 | 1.81775851 | 1.405209611 | 2.4842322 | 1.348454349 | 0.9359055 | 2 | RBCS RPE1, RBCS PRK1 |
| 888 | 1143 | Cre03.g169900.t1.2 | Cre03.g169900.t1.2 | - | - | 0 | 0 | C | 0 | 0 | 0 | 0.97600076 | 2.549034127 | 2.13533874 | 2 | EPYC1 PRK1, EPYC1 RPE1 |
| 265 | 244 | Cre08.g384750.t1.1 | Cre08.g384750.t1.1 | AMA3 | Belongs to glycoside hydrolase family 13; catalyses hydrolysis of (1-4)-alpha-D-glucosidic linkages in polysaccharides so as to remove successive alpha-maltose residues from the non-reducing ends of the chains in the conversion of starch to maltose | 0 | 0 | C | 1.78667118 | 0.650872955 | 3.024842331 | 1.77169905 | 0.635900882 | 2.93103703 | 2 | RBCS PRK1, EPYC1 PRK1 |
| 362 | 280 | Cre12.g483650.t1.2 | Cre12.g483650.t1.2 | - | - | 0 | 0 | O | 0.29745616 | 1.700295318 | 1.175877866 | 0 | 0 | 0 | 2 | RBCS RPE1, RBCS PRK1 |
| 106 | 308 | Cre12.g550850.t1.2 | Cre12.g550850.t1.2 | PSBP1 | Chloroplast precursor (OEE2) (23 kDa subunit of oxygen evolving complex of photosystem II) (OEC 23 kDa subunit) (23 kDa thylakoid membrane protein) (OEC23);PMID: 3468511; targeted to thylakoid lumen by TAT pathway | 0 | 0 | C | 1.48362891 | 1.989277007 | 1.523631888 | 0.14085308 | 0.646501204 | 0.18085605 | 2 | RBCS RPE1, RBCS PRK1 |
| 367 | 337 | Cre12.g498500.t1.2 | Cre12.g498500.t1.2 | DEG11 | DegP-type protease, thylakoid lumen; one PDZ domain; similar to At DegP1; v5.5 model is truncated at N-term | 0 | 0 | M | 0.04569154 | 1.753902711 | 1.119986344 | 0 | 0 | 0 | 2 | RBCS RPE1, RBCS PRK1 |
| 205 | 420 | Cre09.g393200.t1.2 | Cre09.g393200.t1.2 | HSP70C | HSP70C is one of 9 genes encoding HSP70-like proteins in Chlamydomonas; HSP70C is localized to mitochondria; it is inducible by heat shock and light; PMID: 2779571, 16143837 | 0 | 0 | M | 1.19339201 | 2.402131113 | 1.201757332 | 0 | 0 | 0 | 2 | RBCS RPE1, RBCS PRK1 |
| 433 | 439 | Cre06.g269650.t1.2 | Cre06.g269650.t1.2 | - | - | 0 | 0 | C | 0.76443546 | 2.059770611 | 1.318299288 | 0.80790736 | 2.103242511 | 1.36177118 | 2 | RBCS RPE1, EPYC1 RPE1 |
| 484 | 472 | Cre16.g663150.t1.2 | Cre16.g663150.t1.2 | STR18 | - | 0 | 0 | M | 2.06810578 | 3.776316955 | 3.306276899 | 0 | 0 | 0 | 2 | RBCS RPE1, RBCS PRK1 |
| 67 | 498 | Cre02.g120150.t1.2 | Cre02.g120150.t1.2 | RBCS2 | RuBisCO small subunit 2, chloroplast precursor [PMID: 3820291]; almost identical to SwissProt P08475 and to nearby RBCS1; RBCS2 is the only isoform present in the crystallized complex [PMID: 11866526] | 1 | 0 | C | 3.14064154 | 4.848852742 | 4.378812666 | 0 | 0 | 0 | 2 | RBCS RPE1, RBCS PRK1 |
| 286 | 529 | Cre12.g534700.t1.2 | Cre12.g534700.t1.2 | FKB19 | FKBP-type peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8) (PPIase) (Rotamase) (Immunophilin); probably directed to thylakoid lumen (homology to AtFKBP19 and bipartite cTP with RR motif) | 0 | 0 | SP | 2.09112731 | 2.289943926 | 2.664867166 | 0 | 0 | 0.70100701 | 2 | RBCS RPE1, RBCS PRK1 |
| 681 | 720 | Cre03.g190281.t1.1 | Cre03.g190281.t1.1 | - | - | 0 | 0 | O | 0.01131759 | 1.819606815 | 1.716628166 | 0 | 0 | 0 | 2 | RBCS RPE1, RBCS PRK1 |
| 604 | 913 | Cre09.g393358.t1.1 | Cre09.g393358.t1.1 | - | - | 0 | 0 | O | 2.63235633 | 2.506238855 | 1.483481332 | 2.75407612 | 2.627958614 | 1.60520111 | 2 | EPYC1 PRK1, EPYC1 RPE1 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 57 | 96 | Cre02.g097800.t1.1 | Cre02.g097800.t1.1 | HLA3 | associated with bicarbonate uptake for CO2-cocentrating mechanism [PMID:19321421, 25660294, 26015566]; high light-induced, requiring both high light and low CO2 (ambient) levels for activation [PMID: 12000678]; MRP subfamily of ABC transporters [PMID: 15710683]; HLA3 is regulated by CCM1 [PMID: 15235119]; localized at plasma membrane [PMID: 15710683] | 0 | 1 | O | 2.04434748 | 2.23646244 | 1.88549887 | 0.56695685 | 0.759071811 | 0.40810824 | 2 | RBCS RPE1, RBCS PRK1 |
| 199 | 98 | Cre03.g199000.t1.2 | Cre03.g199000.t1.2 | PHOT1 | Blue-light receptor; found in flagella; regulates many biological processes including gametogenesis, chemotaxis, zygote maturation, light-stress response; contains two PAS and two protein kinase domains | 0 | 0 | C | 2.03332383 | 0.953842444 | 1.43898092 | 2.46663538 | 1.387153991 | 1.87229247 | 2 | RBCS PRK1, EPYC1 PRK1 |

**Supplemental Data Set S7.** TMT Mass spectrometry analysis of RBCS2-TurboID at different $CO_2$ levels and in the *epyc1* mutant. Gene annotations and autodefinition obtained from Phytozome were used to annotate all genes detected. The $Log_2$ Fold Change of detected proteins for each binary comparison was calculated and a PEAKSQ test was used to calculate the statistical signficance as a -$Log_P$ value. **This material is too large to include in the thesis and is instead available online at https://doi.org/10.1093/plcell/koad131.**

**Supplemental Data Set S8**. The statistical summary of results shown in Figure 2E. Two-Way repeated measures ANOVA was carried out on proteins with known localization. Tukey's multiple comparisons tests and the multiple-comparison adjusted P-value presented below.

| Test details | Mean 1 | Mean 2 | Mean Diff. | SE of diff. | N1 | N2 | q | DF | 95.00% CI of diff. | Significant? | Summary | Adjusted P Value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pyrenoid | | | | | | | | | | | | |
| 1 Hour vs. 2 hour | 0.6909 | 0.7594 | -0.0685 | 0.03272 | 18 | 18 | 2.96 | 17 | -0.1615 to 0.02453 | No | ns | 0.195 |
| 1 Hour vs. 4 Hour | 0.6909 | 0.9878 | -0.2969 | 0.04393 | 18 | 18 | 9.559 | 17 | -0.4218 to -0.1721 | Yes | **** | <0.0001 |
| 1 Hour vs. 8 Hour | 0.6909 | 0.7998 | -0.1089 | 0.06465 | 18 | 18 | 2.382 | 17 | -0.2927 to 0.07489 | No | ns | 0.3617 |
| 2 Hour vs. 4 Hour | 0.7594 | 0.9878 | -0.2284 | 0.03646 | 18 | 18 | 8.86 | 17 | -0.3321 to -0.1248 | Yes | **** | <0.0001 |
| 2 Hour vs. 8 Hour | 0.7594 | 0.7998 | -0.0404 | 0.04796 | 18 | 18 | 1.191 | 17 | -0.1767 to 0.09593 | No | ns | 0.8336 |
| 4 Hour vs. 8 Hour | 0.9878 | 0.7998 | 0.188 | 0.02992 | 18 | 18 | 8.888 | 17 | 0.1030 to 0.2731 | Yes | **** | <0.0001 |
| Stroma | | | | | | | | | | | | |
| 1 Hour vs. 2 hour | 0.09441 | 0.116 | -0.02154 | 0.02341 | 15 | 15 | 1.302 | 14 | -0.08959 to 0.04650 | No | ns | 0.7946 |
| 1 Hour vs. 4 Hour | 0.09441 | 0.2829 | -0.1885 | 0.04813 | 15 | 15 | 5.537 | 14 | -0.3284 to -0.04856 | Yes | ** | 0.0075 |
| 1 Hour vs. 8 Hour | 0.09441 | 0.06294 | 0.03147 | 0.07518 | 15 | 15 | 0.592 | 14 | -0.1870 to 0.2500 | No | ns | 0.9744 |
| 2 Hour vs. 4 Hour | 0.116 | 0.2829 | -0.1669 | 0.03221 | 15 | 15 | 7.329 | 14 | -0.2605 to -0.07330 | Yes | *** | 0.0007 |
| 2 Hour vs. 8 Hour | 0.116 | 0.06294 | 0.05301 | 0.06309 | 15 | 15 | 1.188 | 14 | -0.1304 to 0.2364 | No | ns | 0.8345 |
| 4 Hour vs. 8 Hour | 0.2829 | 0.06294 | 0.2199 | 0.0383 | 15 | 15 | 8.121 | 14 | 0.1086 to 0.3313 | Yes | *** | 0.0003 |
| Pyrenoid + Stroma | | | | | | | | | | | | |
| 1 Hour vs. 2 hour | 0.1045 | 0.07833 | 0.02615 | 0.03161 | 10 | 10 | 1.17 | 9 | -0.07253 to 0.1248 | No | ns | 0.8403 |
| 1 Hour vs. 4 Hour | 0.1045 | 0.1932 | -0.08875 | 0.05208 | 10 | 10 | 2.41 | 9 | -0.2513 to 0.07383 | No | ns | 0.3754 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 Hour vs. 8 Hour | 0.1045 | -0.09057 | 0.1951 | 0.06967 | 10 | 10 | 3.96 | 9 | -0.02243 to 0.4125 | No | ns | 0.0811 |
| 2 Hour vs. 4 Hour | 0.07833 | 0.1932 | -0.1149 | 0.04931 | 10 | 10 | 3.295 | 9 | -0.2689 to 0.03905 | No | ns | 0.162 |
| 2 Hour vs. 8 Hour | 0.07833 | -0.09057 | 0.1689 | 0.06861 | 10 | 10 | 3.481 | 9 | -0.04529 to 0.3831 | No | ns | 0.1339 |
| 4 Hour vs. 8 Hour | 0.1932 | -0.09057 | 0.2838 | 0.03003 | 10 | 10 | 13.36 | 9 | 0.1900 to 0.3776 | Yes | **** | <0.0001 |
| Non-Chloroplastic | | | | | | | | | | | | |
| 1 Hour vs. 2 hour | -0.1325 | -0.223 | 0.09042 | 0.01652 | 13 | 13 | 7.74 | 12 | 0.04137 to 0.1395 | Yes | *** | 0.0007 |
| 1 Hour vs. 4 Hour | -0.1325 | -0.1327 | 0.0001447 | 0.06054 | 13 | 13 | 0.00338 | 12 | -0.1796 to 0.1799 | No | ns | >0.9999 |
| 1 Hour vs. 8 Hour | -0.1325 | -0.4771 | 0.3446 | 0.08702 | 13 | 13 | 5.6 | 12 | 0.08623 to 0.6030 | Yes | ** | 0.0089 |
| 2 Hour vs. 4 Hour | -0.223 | -0.1327 | -0.09028 | 0.0562 | 13 | 13 | 2.272 | 12 | -0.2571 to 0.07659 | No | ns | 0.4113 |
| 2 Hour vs. 8 Hour | -0.223 | -0.4771 | 0.2542 | 0.08206 | 13 | 13 | 4.38 | 12 | 0.01054 to 0.4978 | Yes | * | 0.04 |
| 4 Hour vs. 8 Hour | -0.1327 | -0.4771 | 0.3444 | 0.03738 | 13 | 13 | 13.03 | 12 | 0.2335 to 0.4554 | Yes | **** | <0.0001 |

| Two-way RM ANOVA | Matching: Stacked |
|---|---|
| Assume sphericity? | No |
| Alpha | 0.05 |

| Source of Variation | % of total variation | P value | P value summary | Significant? | Geisser-Greenhouse's epsilon |
|---|---|---|---|---|---|
| Incubation time x Localization | 0.9829 | <0.0001 | **** | Yes | |
| Incubation time | 1.889 | <0.0001 | **** | Yes | 0.4571 |
| Localization | 38.67 | <0.0001 | **** | Yes | |
| Protein identity | 55.37 | <0.0001 | **** | Yes | |

| ANOVA table | SS | DF | MS | F (DFn, DFd) | P value |
|---|---|---|---|---|---|
| Incubation time x Localization | 0.943 | 9 | 0.1048 | F (9, 156) = 5.548 | P<0.0001 |
| Incubation time | 1.812 | 3 | 0.6039 | F (1.371, 71.30) = 31.98 | P<0.0001 |
| Localization | 37.1 | 3 | 12.37 | F (3, 52) = 12.11 | P<0.0001 |
| Protein identity | 53.12 | 52 | 1.022 | F (52, 156) = 54.10 | P<0.0001 |
| Residual | 2.946 | 156 | 0.01888 | | |

| Data summary | |
|---|---|
| Number of columns (Localization) | 4 |
| Number of rows (Incubation time) | 4 |
| Number of subjects (Protein identity) | 56 |
| Number of missing values | 0 |

**Supplemental Data Set S9.** Oligo sequences used.

| Oligo name | Sequence | Description |
|---|---|---|
| oLM2919 | TAGAAGACAACTCAAATGCAGGCCCTCCAGATGC | Forward primer to amplify RPE1 (Cre12.g511900) genomic region for L0 assembly into L0-RPE1-TurboID plasmid |
| oLM2920 | TAGAAGACTTTGCCCGCCATCACGTTGGCGGGG | Reverse primer to amplify RPE1 (Cre12.g511900) genomic region for L0 assembly into L0-RPE1-TurboID plasmid |
| oLM2921 | TAGAAGACAACTCAAATGGCTTTCACTATGCGCGC | Forward primer to amplify RPE1 (Cre12.g554800) genomic region for L0 assembly into L0-RPE1-TurboID plasmid |
| oLM2922 | TAGAAGACTTTGCCCACGGGCACAACGTCCTTC | Reverse primer to amplify RPE1 (Cre12.g554800) genomic region for L0 assembly into L0-RPE1-TurboID plasmid |
| oLM2207 | ATAGTTAACATGGGAAGACAAGGCAGCGGCAGCACCAGCGGCAGCGGCAGCAAGTCTTCGAAGTTAACAATC | Flesible linker sequence anneals with oLM2208 as a Level -1 part used to assembling with the TurboID expression plasmid used in MS experiment |
| oLM2208 | GATTGTTAACTTCGAAGACTTGCTGCCGCTGCCGCTGGTGCTGCCGCTGCCTTGTCTTCCCATGTTAACTAT | Flesible linker sequence anneals with oLM2207 as a Level -1 part used to assembling with TurboID expression plasmid used in the MS experiment |
| oLM3219 | ATAGTTAACATGGGAAGACAACTCAAGGTGGCAGCGGCAGCACCAGCGGCAGCGGCAGCAAGTCTTCGAAGTTAACAATC | Flesible linker sequence anneals with oLM3219 to assemble the TurboID sequence into L0-TurboID at B4 position in pUAP1 |
| oLM3220 | GATTGTTAACTTCGAAGACTTGCTGCCGCTGCCGCTGGTGCTGCCGCTGCCACCTTGAGTTGTCTTCCCATGTTAACTAT | Flesible linker sequence anneals with oLM3220 to assemble the TurboID sequence into L0-TurboID at B4 position in pUAP1 |
| oLM2591 | TTCTAGAAATGCAAACGCCATGTAGTTTACATCACTCTTCAGGAGGAACTGAAGATCCTTTGATCTTTTCTACGGG | Reverse primer to clone Cre13.g573250 genomic region and its native promoter into recombineering plasmid pLM162 |
| oLM2592 | CCGGCTTCGACAAGGTGGACAAGGCCGACCGCTGGTCGCTGCCCGTCTACGGAGATCTGGGTGGCTCCG | Forward primer to clone Cre13.g573250 genomic region and its native promoter into recombineering plasmid pLM162 |
| oLM665 | GTGCTCGTCGCAGTGCGGCACCCACGACCGGGCACGCGGGGTTTGGCTGGGAAGATCCTTTGATCTTTTCTACGGG | Reverse primer to clone Cre16.g663150 genomic region and its native promoter into recombineering plasmid pLM099 |
| oLM664 | ATGGTGGGCAGGCGGAGGCTGCGGCCCCCAAGAAGTACAGCGCGTGGCGGGGAGATCTGGGTGGCTCCG | Forward primer to clone Cre16.g663150 genomic region and its native promoter into recombineering plasmid pLM099 |

| | | |
|---|---|---|
| oLM813 | GGTGGTAGTCCTTGCAAAGCTTAACGTGCGGTAAGCCCTGACCCAGACTCGAAGATCCTTTGATCTTTTCTACGGG | Reverse primer to clone Cre06.g271850 genomic region and its native promoter into recombineering plasmid pLM099 |
| oLM498 | AGAGCAAGGAGGTGCGCTGGCTTGAGCTAGCGGAAATCGCCGGCGACATCGGAGATCTGGGTGGCTCCG | Forward primer to clone Cre06.g271850 genomic region and its native promoter into recombineering plasmid pLM099 |
| oLM2885 | AGTGAGAGTGAGAGTGAGAGCACAAGGGTTGTGTGTGTTAAAAAACGAAGGAAGATCCTTTGATCTTTTCTACGGG | Reverse primer to clone Cre03.g172700 genomic region and its native promoter into recombineering plasmid pLM099 |
| oLM2886 | CAGAAGGCGCGAAGCGCGCCGGTGGCAACTGGCGAGAGACGGCCGACGTGGGAGATCTGGGTGGCTCCG | Forward primer to clone Cre03.g172700 genomic region and its native promoter into recombineering plasmid pLM099 |
| oLM2871 | ACAACTGTGAAGACTTATTCCATAATGAATTAACAACCTGCAATATTGCAGAAGATCCTTTGATCTTTTCTACGGG | Reverse primer to clone Cre17.g720450 genomic region and its native promoter into recombineering plasmid pLM099 |
| oLM2872 | CGGGAGCTGGGCGGAAACCCGGCGGCAACAGCCCAGGCCGCAACGGCCTCGGAGATCTGGGTGGCTCCG | Forward primer to clone Cre17.g720450 genomic region and its native promoter into recombineering plasmid pLM099 |
| poLM001-E01 | ATGGCAGCTGCTCAGACCTCGACGTGATGTATCCTACCGGTGCTTGACTAGAAGATCCTTTGATCTTTTCTACGGG | Reverse primer to clone Cre09.g394510 genomic region and its native promoter into recombineering plasmid pLM099 |
| poLM003-E01 | CTGGCGCCACCGCTGCCGCGCCGGTTACGGCCAACGCCACCAATGGGCGGGGAGATCTGGGTGGCTCCG | Forward primer to clone Cre09.g394510 genomic region and its native promoter into recombineering plasmid pLM099 |
| oLM3223 | TAAAAAACAGCCATGTAGTATCAAACCGGTACATAGAATAGGTATGGCGAGAAGATCCTTTGATCTTTTCTACGGG | Reverse primer to clone Cre02.g093650 genomic region and its native promoter into recombineering plasmid pLM099 |
| oLM3224 | CGCCTCCGCAGGTGCGCATTAGCGACAACGGCGACATTGAGGTCAATGTGGGAGATCTGGGTGGCTCCG | Forward primer to clone Cre02.g093650 genomic region and its native promoter into recombineering plasmid pLM099 |

## Funding

C.S.L. was supported by the Bill and Melinda Gates Agricultural Innovations (Investment ID 53197). The project was supported by a United Kingdom Research and Innovation Future Leaders Fellowship to L.C.M.M. (MR/T020679/1) and Biotechnology and Biological Sciences Research Council Grants (BB/T017589/1, BB/S015337/1, BB/R001014/1, and BB/X003035/1) and Engineering and Physical Research Council Grant (EP/W024063/1). P.G. was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project number 456013262. The York Centre of Excellence in Mass Spectrometry was created thanks to a major capital investment through Science City York, supported by Yorkshire Forward with funds from the Northern Way Initiative, and subsequent support from EPSRC (EP/K039660/1 and EP/M028127/1).

References

**Adler L, Díaz-Ramos A, Mao Y, Pukacz KR, Fei C, and McCormick AJ**. New horizons for building pyrenoid-based CO2-concentrating mechanisms in plants to improve yields. Plant Physiol. 2022:**190**(3):1609–1627.

**Arias C, Obudulu O, Zhao X, Ansolia P, Zhang X, Paul S, Bygdell J, Pirmoradian M, Zubarev RA, Samuelsson G, et al.** Nuclear proteome analysis of Chlamydomonas with response to CO2 limitation. Algal Research. 2020:**46**:101765.

**Atkinson N, Mao Y, Chan KX, and McCormick AJ**. Condensation of Rubisco into a proto-pyrenoid in higher plant chloroplasts. Nat Commun. 2020:**11**(1):6303.

**Baier T, Wichmann J, Kruse O, and Lauersen KJ**. Intron-containing algal transgenes mediate efficient recombinant gene expression in the green microalga Chlamydomonas reinhardtii. Nucleic Acids Res. 2018:**46**(13):6909–6919.

**Banani SF, Lee HO, Hyman AA, and Rosen MK**. Biomolecular condensates: organizers of cellular biochemistry. Nat Rev Mol Cell Biol. 2017:**18**(5):285–298.

**Barrett J, Girr P, and Mackinder LCM**. Pyrenoids: CO2-fixing phase separated liquid organelles. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2021:**1868**(5):118949.

**Boël G, Smith PC, Ning W, Englander MT, Chen B, Hashem Y, Testa AJ, Fischer JJ, Wieden H-J, Frank J, et al.** The ABC-F protein EttA gates ribosome entry into the translation elongation cycle. Nat Struct Mol Biol. 2014:**21**(2):143–151.

**Bonomi F, Pagani S, and Cerletti P**. Insertion of sulfide into ferredoxins catalyzed by rhodanese. FEBS Lett. 1977:**84**(1):149–152.

**Borkhsenious ON, Mason CB, and Moroney JV**. The intracellular localization of ribulose-1,5-bisphosphate Carboxylase/Oxygenase in chlamydomonas reinhardtii. Plant Physiol. 1998:**116**(4):1585–1591.

**Boyd ES, Thomas KM, Dai Y, Boyd JM, and Outten FW**. Interplay between oxygen and Fe-S cluster biogenesis: insights from the Suf pathway. Biochemistry. 2014:**53**(37):5834–5847.

**Bracha D, Walls MT, and Brangwynne CP**. Probing and engineering liquid-phase organelles. Nat Biotechnol. 2019:**37**(12):1435–1445.

**Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, Feldman JL, Perrimon N, and Ting AY**. Efficient proximity labeling in living cells and organisms with TurboID. Nat Biotechnol. 2018:**36**(9):880–887.

**Brueggeman AJ, Gangadharaiah DS, Cserhati MF, Casero D, Weeks DP, and Ladunga I**. Activation of the carbon concentrating mechanism by CO2 deprivation coincides with massive transcriptional restructuring in Chlamydomonas reinhardtii. Plant Cell. 2012:**24**(5):1860–1875.

**Chng S-S, Dutton RJ, Denoncin K, Vertommen D, Collet J-F, Kadokura H, and Beckwith J**. Overexpression of the rhodanese PspE, a single cysteine-containing protein, restores disulphide bond formation to an Escherichia coli strain lacking DsbA. Mol Microbiol. 2012:**85**(5):996–1006.

**Choi J-M, Holehouse AS, and Pappu RV**. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. Annu Rev Biophys. 2020:**49**:107–133.

**Christopher JA, Stadler C, Martin CE, Morgenstern M, Pan Y, Betsinger CN, Rattray DG, Mahdessian D, Gingras A-C, Warscheid B, et al.** Subcellular proteomics. Nat Rev Methods Primers. 2021:**1**. https://doi.org/10.1038/s43586-021-00029-y

**Cox J and Mann M**. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008:**26**(12):1367–1372.

**Crozet P, Navarro FJ, Willmund F, Mehrshahi P, Bakowski K, Lauersen KJ, Pérez-Pérez M-E, Auroy P, Gorchs Rovira A, Sauret-Gueto S, et al.** Birth of a photosynthetic chassis: A MoClo toolkit enabling synthetic biology in the microalga *Chlamydomonas reinhardtii*. ACS Synth Biol. 2018:**7**(9):2074–2086.

**Dahlgren KK, Gates C, Lee T, and Cameron JC**. Proximity-based proteomics reveals the thylakoid lumen proteome in the cyanobacterium Synechococcus sp. PCC 7002. Photosynth Res. 2021:**147**(2):177–195.

**Emrich-Mills TZ, Yates G, Barrett J, Girr P, Grouneva I, Lau CS, Walker CE, Kwok TK, Davey JW, Johnson MP, et al.** A recombineering pipeline to clone large and complex genes in Chlamydomonas. Plant Cell. 2021:**33**(4):1161–1181.

**Engel BD, Schaffer M, Kuhn Cuellar L, Villa E, Plitzko JM, and Baumeister W**. Native architecture of the Chlamydomonas chloroplast revealed by in situ cryo-electron tomography. Elife. 2015:**4**. https://doi.org/10.7554/elife.04889

**Fang W, Si Y, Douglass S, Casero D, Merchant SS, Pellegrini M, Ladunga I, Liu P, and Spalding MH**. Transcriptome-wide changes in Chlamydomonas reinhardtii gene expression regulated by carbon dioxide and the CO2-concentrating mechanism regulator CIA5/CCM1. Plant Cell. 2012:**24**(5):1876–1893.

**Fei C, Wilson AT, Mangan NM, Wingreen NS, and Jonikas MC**. Modelling the pyrenoid-based CO2-concentrating mechanism provides insights into its operating principles and a roadmap for its engineering into crops. Nat Plants. 2022:**8**(5):583–595.

**Freeman Rosenzweig ES, Xu B, Kuhn Cuellar L, Martinez-Sanchez A, Schaffer M, Strauss M, Cartwright HN, Ronceray P, Plitzko JM, Förster F, et al.** The

Eukaryotic CO2-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization. Cell. 2017:**171**(1):148-162.e19.

**Ganapathy US, Bai L, Wei L, Eckartt KA, Lett CM, Previti ML, Carrico IS, and Seeliger JC**. Compartment-Specific Labeling of Bacterial Periplasmic Proteins by Peroxidase-Mediated Biotinylation. ACS Infect Dis. 2018:**4**(6):918–925.

**Gargouri M, Park J-J, Holguin FO, Kim M-J, Wang H, Deshpande RR, Shachar-Hill Y, Hicks LM, and Gang DR**. Identification of regulatory network hubs that control lipid metabolism in Chlamydomonas reinhardtii. J Exp Bot. 2015:**66**(15):4551–4566.

**Han J, Pluhackova K, and Böckmann RA**. The Multifaceted Role of SNARE Proteins in Membrane Fusion. Front Physiol. 2017:**8**:5.

**Han S, Li J, and Ting AY**. Proximity labeling: spatially resolved proteomic mapping for neurobiology. Curr Opin Neurobiol. 2018:**50**:17–23.

**Harvey SH, Krien MJE, and O'Connell MJ**. Structural maintenance of chromosomes (SMC) proteins, a family of conserved ATPases. Genome Biol. 2002:**3**(2):REVIEWS3003.

**He S, Chou H-T, Matthies D, Wunder T, Meyer MT, Atkinson N, Martinez-Sanchez A, Jeffrey PD, Port SA, Patena W, et al.** The structural basis of Rubisco phase separation in the pyrenoid. Nat Plants. 2020:**6**(12):1480–1490.

**Hubstenberger A, Courel M, Bénard M, Souquere S, Ernoult-Lange M, Chouaib R, Yi Z, Morlot J-B, Munier A, Fradet M, et al.** P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. Mol Cell. 2017:**68**(1):144-157.e5.

**Hyman AA, Weber CA, and Jülicher F**. Liquid-liquid phase separation in biology. Annu Rev Cell Dev Biol. 2014:**30**:39–58.

**Itakura AK, Chan KX, Atkinson N, Pallesen L, Wang L, Reeves G, Patena W, Caspari O, Roth R, Goodenough U, et al.** A Rubisco-binding protein is required for normal pyrenoid number and starch sheath morphology in *Chlamydomonas reinhardtii*. Proc Natl Acad Sci U S A. 2019:**116**(37):18445–18454.

**Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al.** Highly accurate protein structure prediction with AlphaFold. Nature. 2021:**596**(7873):583–589.

**Kim DI, Birendra KC, Zhu W, Motamedchaboki K, Doye V, and Roux KJ**. Probing nuclear pore complex architecture with proximity-dependent biotinylation. Proc Natl Acad Sci U S A. 2014:**111**(24):E2453-61.

**Kim DI, Jensen SC, Noble KA, Kc B, Roux KH, Motamedchaboki K, and Roux KJ**. An improved smaller biotin ligase for BioID proximity labeling. Mol Biol Cell. 2016:**27**(8):1188–1196.

**Kreis E, König K, Misir M, Niemeyer J, Sommer F, and Schroda M**. TurboID reveals the proxiomes of *Chlamydomonas* proteins involved in thylakoid biogenesis and stress response. Plant Physiol. 2023:**193**(3):1772–1796.

**Kropat J, Hong-Hermesdorf A, Casero D, Ent P, Castruita M, Pellegrini M, Merchant SS, and Malasarn D**. A revised mineral nutrient supplement increases biomass and growth rate in Chlamydomonas reinhardtii. Plant J. 2011:**66**(5):770–780.

**Kuchitsu K, Tsuzuki M, and Miyachi S**. Changes of Starch Localization within the Chloroplast Induced by Changes in CO2 Concentration during Growth of Chlamydomonas reinhardtii: Independent Regulation of Pyrenoid Starch and Stroma Starch. Plant Cell Physiol. 1988:**29**(8):1269–1278.

**Küken A, Sommer F, Yaneva-Roder L, Mackinder LC, Höhne M, Geimer S, Jonikas MC, Schroda M, Stitt M, Nikoloski Z, et al.** Effects of microcompartmentation on flux distribution and metabolic pools in Chlamydomonas reinhardtii chloroplasts. Elife. 2018:**7**. https://doi.org/10.7554/eLife.37960

**Lam SS, Martell JD, Kamer KJ, Deerinck TJ, Ellisman MH, Mootha VK, and Ting AY**. Directed evolution of APEX2 for electron microscopy and proximity labeling. Nat Methods. 2015:**12**(1):51–54.

**Li X, Patena W, Fauser F, Jinkerson RE, Saroussi S, Meyer MT, Ivanova N, Robertson JM, Yue R, Zhang R, et al.** A genome-wide algal mutant library and functional screen identifies genes required for eukaryotic photosynthesis. Nat Genet. 2019:**51**(4):627–635.

**Li X, Zhang R, Patena W, Gang SS, Blum SR, Ivanova N, Yue R, Robertson JM, Lefebvre PA, Fitz-Gibbon ST, et al.** An Indexed, Mapped Mutant Library Enables Reverse Genetics Studies of Biological Processes in Chlamydomonas reinhardtii. Plant Cell. 2016:**28**(2):367–387.

**Li-Beisson Y, Beisson F, and Riekhof W**. Metabolism of acyl-lipids in Chlamydomonas reinhardtii. Plant J. 2015:**82**(3):504–522.

**Long SP, Burgess S, and Causton I**. Redesigning crop photosynthesis. Sustaining global food security: the nexus of science and policy.

**Mackinder LCM**. The Chlamydomonas CO2 -concentrating mechanism and its potential for engineering photosynthesis in plants. New Phytol. 2018:**217**(1):54–61.

**Mackinder LCM, Chen C, Leib RD, Patena W, Blum SR, Rodman M, Ramundo S, Adams CM, and Jonikas MC**. A Spatial Interactome Reveals the Protein Organization of the Algal CO2-Concentrating Mechanism. Cell. 2017:**171**(1):133-147.e14.

**Mackinder LCM, Meyer MT, Mettler-Altmann T, Chen VK, Mitchell MC, Caspari O, Freeman Rosenzweig ES, Pallesen L, Reeves G, Itakura A, et al.** A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. Proc Natl Acad Sci U S A. 2016:**113**(21):5958–5963.

**Mair A and Bergmann DC**. Advances in enzyme-mediated proximity labeling and its potential for plant research. Plant Physiol. 2022:**188**(2):756–768.

**Mair A, Xu S-L, Branon TC, Ting AY, and Bergmann DC**. Proximity labeling of protein complexes and cell-type-specific organellar proteomes in Arabidopsis enabled by TurboID. Elife. 2019:**8**. https://doi.org/10.7554/eLife.47864

**Meyer MT, Genkov T, Skepper JN, Jouhet J, Mitchell MC, Spreitzer RJ, and Griffiths H**. Rubisco small-subunit α-helices control pyrenoid formation in *Chlamydomonas*. Proc Natl Acad Sci U S A. 2012:**109**(47):19474–19479.

**Meyer MT, Itakura AK, Patena W, Wang L, He S, Emrich-Mills T, Lau CS, Yates G, Mackinder LCM, and Jonikas MC**. Assembly of the algal $CO_2$-fixing organelle, the

pyrenoid, is guided by a Rubisco-binding motif. Sci Adv. 2020:**6**(46). https://doi.org/10.1126/sciadv.abd2408

**Neofotis P, Temple J, Tessmer OL, Bibik J, Norris N, Pollner E, Lucker B, Weraduwage SM, Withrow A, Sears B, et al.** The induction of pyrenoid synthesis by hyperoxia and its implications for the natural diversity of photosynthetic responses in Chlamydomonas. Elife. 2021:**10**. https://doi.org/10.7554/eLife.67565

**Ray DK, Mueller ND, West PC, and Foley JA**. Yield Trends Are Insufficient to Double Global Crop Production by 2050. PLoS One. 2013:**8**(6):e66428.

**Rhee H-W, Zou P, Udeshi ND, Martell JD, Mootha VK, Carr SA, and Ting AY**. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. Science. 2013:**339**(6125):1328–1331.

**Roux KJ, Kim DI, Raida M, and Burke B**. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J Cell Biol. 2012:**196**(6):801–810.

**Rydz L, Wróbel M, and Jurkowska H**. Sulfur Administration in Fe-S Cluster Homeostasis. Antioxidants (Basel). 2021:**10**(11). https://doi.org/10.3390/antiox10111738

**Sun Y, Valente-Paterno M, Bakhtiari S, Law C, Zhan Y, and Zerges W**. Photosystem Biogenesis Is Localized to the Translation Zone in the Chloroplast of Chlamydomonas. Plant Cell. 2019:**31**(12):3057–3072.

**Tardif M, Atteia A, Specht M, Cogne G, Rolland N, Brugière S, Hippler M, Ferro M, Bruley C, Peltier G, et al.** PredAlgo: a new subcellular localization prediction tool dedicated to green algae. Mol Biol Evol. 2012:**29**(12):3625–3639.

**Toyokawa C, Yamano T, and Fukuzawa H**. Pyrenoid Starch Sheath Is Required for LCIB Localization and the CO2-Concentrating Mechanism in Green Algae. Plant Physiol. 2020:**182**(4):1883–1893.

**Turnšek J, Brunson JK, Viedma MDPM, Deerinck TJ, Horák A, Oborník M, Bielinski VA, and Allen AE**. Proximity proteomics in a marine diatom reveals a putative cell surface-to-chloroplast iron trafficking pathway. Elife. 2021:**10**. https://doi.org/10.7554/eLife.52770

**Uniacke J, Colón-Ramos D, and Zerges W**. FISH and immunofluorescence staining in Chlamydomonas. Methods Mol Biol. 2011:**714**:15–29.

**Uniacke J and Zerges W**. Stress induces the assembly of RNA granules in the chloroplast of Chlamydomonas reinhardtii. J Cell Biol. 2008:**182**(4):641–646.

**Wang Q, Chen Z, Zhang X, Xin Y, Xia Y, Xun L, and Liu H**. Rhodanese Rdl2 produces reactive sulfur species to protect mitochondria from reactive oxygen species. Free Radic Biol Med. 2021:**177**:287–298.

**Wunder T, Cheng SLH, Lai S-K, Li H-Y, and Mueller-Cajar O**. The phase separation underlying the pyrenoid-based microalgal Rubisco supercharger. Nat Commun. 2018:**9**(1):5076.

**Wurzinger B, Stael S, Leonardelli M, Perolo C, Melzer M, Chaturvedi P, Afjehi-Sadat L, Weckwerth W, and Teige M**. Proximity labelling allows to study novel factors in chloroplast development[a]. bioRxiv. 2022. https://doi.org/10.1101/2022.12.08.519630

**Yamano T, Tsujikawa T, Hatano K, Ozawa S-I, Takahashi Y, and Fukuzawa H**. Light and low-CO2-dependent LCIB-LCIC complex localization in the chloroplast supports the carbon-concentrating mechanism in Chlamydomonas reinhardtii. Plant Cell Physiol. 2010:**51**(9):1453–1468.

**Youn J-Y, Dunham WH, Hong SJ, Knight JDR, Bashkurov M, Chen GI, Bagci H, Rathod B, MacLeod G, Eng SWM, et al.** High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies. Mol Cell. 2018:**69**(3):517-532.e11.

**Zhan Y, Dhaliwal JS, Adjibade P, Uniacke J, Mazroui R, and Zerges W**. Localized control of oxidized RNA. J Cell Sci. 2015:**128**(22):4210–4219.

**Zhan Y, Marchand CH, Maes A, Mauries A, Sun Y, Dhaliwal JS, Uniacke J, Arragain S, Jiang H, Gold ND, et al.** Pyrenoid functions revealed by proteomics in Chlamydomonas reinhardtii. PLoS One. 2018:**13**(2):e0185039.

**Zhang Y, Song G, Lal NK, Nagalakshmi U, Li Y, Zheng W, Huang P-J, Branon TC, Ting AY, Walley JW, et al.** TurboID-based proximity labeling reveals that UBR7 is a regulator of N NLR immune receptor-mediated immunity. Nat Commun. 2019:**10**(1):3252.

**Zhou Y and Zou P**. The evolving capabilities of enzyme-mediated proximity labeling. Curr Opin Chem Biol. 2021:**60**:30–38.

# Chapter 3 – Proteomic mapping of *Chlamydomonas* pyrenoid interfaces

## Introduction to Chapter 3

The present chapter builds on chapter 2 with the use of proximity labelling strategy, here we further expanded the proximity labelling toolkit by incorporating the plasmid into the recombineering pipeline (Emrich-Mills et al. 2021). Previously, the cloning of targeted genes into the constructed TurboID plasmids such as in (Lau et al. 2023) or (Kreis et al. 2023) requires gene domestication to remove Type IIS enzyme restriction sites from their sequence. This requirement largely restricted the cloning of *Chlamydomonas* genes which are commonly large and complex. In the present chapter, by replacing the gene encoding for the Venus fluorescent protein in the pLM099 plasmid with the TurboID coding sequence, we generated the rc-TurboID plasmid backbone which facilitated the cloning of large and complex genes into TurboID-tagging plasmids. This novel rc-TurboID backbone plasmid differs from the previous plasmid in two major facets: the inclusion of the native promoter and a much longer flexible linker between cloned gene and the TurboID tag. The latter difference is particularly crucial as it extends the biotinylation radius of the TurboID tag by approximately 4 nm which aids in the pursuit of identifying proximal proteins in addition to interactors. Taking the newly developed plasmid backbone, we expanded our labelling target beyond the pyrenoid matrix. To do so, we identified multiple proteins with well documented localisation such as FZL, PSAF, LCI9 and RBMP2. We then use these as targeting proteins for biotinylation at thylakoids, pyrenoid tubules and the starch sheath, the latter two being particularly relevant to the assembly of the pyrenoid. It is hypothesised that pyrenoid assembly is driven through the Rubisco interacting motif, however previous work identified 4099 proteins that contain at least 1 Rubisco interacting motif, making it an unrealistic task to search for a pyrenoid assembly protein simply by this sole criterion. The proximity labelling targeted towards pyrenoid organelle interfaces such as the pyrenoid tubules and starch sheath periphery helps identify proteins present at these critical locations, combined with previous pulldown work using pyrenoid components and bioinformatics analysis, the present chapter provides additionally information to inform our search for the elusive pyrenoid assembly proteins.

The present chapter is written as a manuscript in the Plant Cell journal format. However, the content of this chapter might be later split into two separate manuscripts: the proximity labelling results of FZL, RBMP2 and PSAF are planned to be included in a manuscript that focuses on characterisation of the *fzl* mutant; while that of LCI9 will be contributing to a manuscript characterising the *lci9* mutant.

### Declaration of Authorship

All experimental work presented below was carried out by Chun Sing Lau with the following exception:

Adam Dowle performed the LC-MS/MS experiment and the associated quantitation.
Liat Adler performed the LCI9-TurboID Streptavidin purification experiments together with Chun Sing Lau.

Philipp Girr generated the plasmids encoding FZL-mNeon for confocal imaging. Other confocal imaging experiments are performed by Chun Sing Lau

All figures were assembled by Chun Sing Lau.

## Proteomic mapping of *Chlamydomonas* pyrenoid interfaces

### Authors

Chun Sing Lau[1], Adam Dowle[1], Liat Adler[2], Philipp Girr[1], Luke CM Mackinder[1]

[1]The University of York, Centre of Novel Agricultural Products, Department of Biology, United Kingdom
[2]Carnegie Institute of Science, Biosphere Sciences and Engineering, United States

### Abstract

The *Chlamydomonas* pyrenoid is composed of three elements, a $CO_2$-fixing Rubisco dense matrix, a matrix-enclosing starch sheath, and a trans-pyrenoidal thylakoid membrane network. The cooperation from all three elements is critical to the efficient operation of the $CO_2$ concentrating mechanism (CCM) which enables the cells' survival in the $CO_2$ scarce aquatic environment. Despite their importance, very little is known on the molecular basis of starch sheath and trans-pyrenoid thylakoid network biogenesis in the pyrenoid. Previously, proximity labelling of the pyrenoid matrix via RBCS2-TurboID recovered a substantial list of pyrenoid proteins demonstrating the techniques strength in identifying spatially specific proteomes. The protein LCI9 lines the adjacent surface where two starch sheath granules meet, while RBMP2 localises to the interface between the pyrenoid matrix and pyrenoid tubules. These interface locations represent proposed critical pyrenoid biogenesis sites, therefore making the two proteins prime targets for TurboID tagging. To enable generation of TurboID fusion proteins with large and complex genes such as RBMP2 we first expanded the TurboID tagging pipeline using the recombination cloning method. Subsequent proximity labelling and analysis using LCI9 and RBMP2 highlights multiple candidate proteins which might participate in the biogenesis of the starch sheath and pyrenoid tubules, as well as generated novel insights in to starch metabolism. The modified vectors and application at

sub-organelle level further establishes proximity labelling as a powerful tool to explore *Chlamydomonas* physiology.

## Introduction

To survive the limitation of $CO_2$ diffusion in an aquatic environment, eukaryotic algae, such as *Chlamydomonas,* operate a $CO_2$ concentration mechanism (CCM) that condenses the carbon-fixing enzyme Rubisco into a phase-separated organelle called the pyrenoid. In many eukaryotic algae, the pyrenoid organelle is often found to harbour two additional components besides its Rubisco-dense matrix: a starch layer external to the pyrenoid; and a thylakoid membrane network which traverses the pyrenoid matrix. Across species, the starch layer is seen to take multiple different arrangements when encasing the Rubisco matrix. For example, in *Chlorella sorokiniana* the pyrenoid matrix is encased by two starch plates, whereas in *Orbulina universa* and *Aureodinium pigmentosum* there is almost complete encasement by a single starch plate (Dodge 1968; LeKieffre et al. 2018). The trans-pyrenoid thylakoid network is also structurally diverse. In *Phaeodactylum triconutum*, the trans-pyrenoid thylakoids (or referred to as pyrenoid-penetrating thylakoid) takes on a simpler configuration where a few thylakoid membranes were joined together to bisect the pyrenoid matrix (Flori et al. 2016; Shimakawa et al. 2023). More extreme cases are found in some red algal species like *Porphyridium pupureum* where the thylakoid membranes is interspersed throughout the pyrenoid matrix (Markina et al. 2021). *Chlamydomonas reinhardtii* represents the best characterised pyrenoid system to date and its pyrenoid harbours both a starch sheath and traversing thylakoids, termed pyrenoid tubules (Engel et al. 2015). The *Chlamydomonas* pyrenoid starch sheath builds from multiple cup-shaped granules which collectively encases the pyrenoid. Thylakoid membranes which run in proximity to the pyrenoid bend and merge into tubule-shaped membranes entering the pyrenoid matrix through the opening between the edge of starch sheath granules. Once within the pyrenoid matrix, the pyrenoid tubules propagate into a disorganised network forming a region called the “reticulated knot”.

The engineering of a pyrenoid-based CCM into crop plants is an increasingly attractive strategy to address the growing food security issue due to its potential to increase crop yield by 60% (Fei et al. 2022). Recent modelling results suggest that the key to achieving an efficient pyrenoid-based CCM *in planta* is the co-assembly of the pyrenoid components which reduces $CO_2$ leakage (starch sheath) and provide a bicarbonate delivery route (pyrenoid tubules). The engineering of these critical components in crop plants must be guided by a list of fundamental molecular parts involved in their assembly in *Chlamydomonas reinhardtii*. However, our understanding of their assembly process is still at its infancy. Many outstanding questions such as how the thylakoid membranes contort from

a sheet to a tubular ultrastructure as it traverses the pyrenoid matrix; how does starch sheath assembly enclose the pyrenoid matrix in coordination with the pyrenoid tubule traversion, remain unanswered. This prohibits us from maximising the proposed yield gain from implementing a pyrenoid-based CCM in crop plants. A previous report identified a Rubisco-binding motif [D/N]W[R/K]XX[L/I/V/A] (RBM) found across many pyrenoid proteins binds to the α-helix of the Rubisco small subunit (Meyer et al. 2012, 2020; He et al. 2020). This Rubisco interacting motif is not only critical for the intrinsically disordered linker protein EPYC1 (ESSENTIAL PYRENOID COMPONENT 1) to drive Rubisco phase separation but was also proposed to direct the assembly of the other structural components of the pyrenoid. Bioinformatic analysis identified ~4099 proteins containing at least one RBM (Meyer et al. 2020). Nevertheless, not all proteins containing a predicted RBM are pyrenoid proteins let alone function to guide pyrenoid assembly. Combining proteomics data of the pyrenoid such as those obtained from pyrenoid purification (Zhan et al. 2018) or proximity labelling (Lau et al. 2023) will likely accelerate our efforts in searching for essential proteins that function in pyrenoid biogenesis.

Here we report the development of the recombination based cloning method for TurboID tagging. This method deviates from the previous cloning procedure via Golden Gate that requires gene domestication to remove type IIs enzyme recognition sites and the amplification of genes via PCR or gene synthesis. Recombineering allows the transfer of a target gene in a BAC or fosmid directly into a destination vector, therefor removing the need to amplify the open reading frames, which is often challenging in *Chlamydomonas* due to their length and complexity. The newly generated TurboID recombineering vectors were used to tag LCI9, RBMP2, PSAF and RBMP2 which drove targeted biotin labelling at sites proposed to be particularly important for pyrenoid assembly such as the pyrenoid tubules and the starch sheath boundaries. The analysis of the LCI9 proxiome identified multiple starch-related proteins therefore implicating them in the starch sheath synthesis process. Additionally, several new potential pyrenoid proteins are found in the RBMP2 proxiome, which might function in the pyrenoid tubule biogenesis process and highlights the pyrenoid's potential role in RNA metabolism. We envisioned that further characterisation of proteins found within the obtained proxiomes will lead to more insights on the pyrenoid assembly process.

## Results

### Construction of a recombineering-compatible TurboID plasmid

Our goal to expand proximity labelling to sites particularly important for the assembly of the pyrenoid organelle (i.e. pyrenoid tubules and starch sheath edge) was first met with

difficulties since *Chlamydomonas* genes are often long and complex which complicates attempts to clone them through the more traditional PCR (polymerase chain reaction)-cloning technique and can make gene synthesis either prohibitively expensive or prone to failure. To circumvent these difficulties, we turned to the previously developed recombineering cloning method (Emrich-Mills et al. 2021), which uses recombination with bacterial artificial chromosomes (BAC) to avoid direct PCR of long gene fragments alongside the domestication of restriction enzyme sites in the gene sequence. To transfer the TurboID enzyme tag into the *Chlamydomonas* recombineering-cloning tool kit, we started by removing a BsaI recognition site on the *ccdB* gene on the plasmid pLM099 through a silent mutation introduced by PCR site-directed mutagenesis (Fig. 1A and 1B). The domesticated pLM099 plasmid was then used as template DNA for later PCR amplifications. The modified pLM099 was amplified as two fragments using oligos which contain internal BsaI recognitions sites for Golden Gate cloning. The TurboID-tag alongside with its flexible linker were amplified from the L1-RBCS2-TurboID. The purified DNA fragments were then assembled to make the recombineering-compatible backbone plasmid via the Golden Gate cloning procedure and was subsequently named as rc-TurboID (Fig. 1A and 1B). Multiple colonies containing the rc-TurboID backbone were then verified by digestion with MlyI (Fig.1 C) and confirmed via sequencing

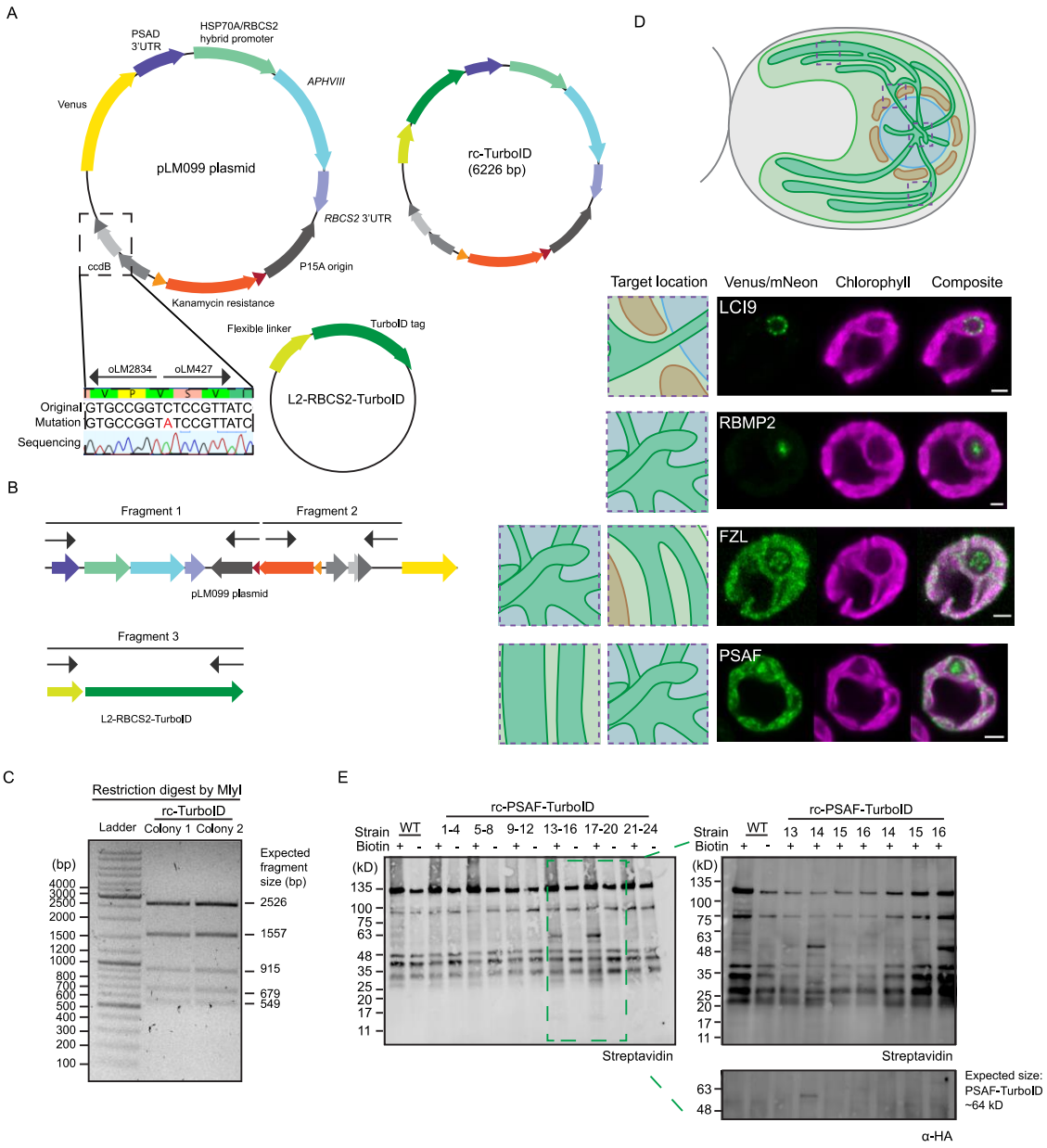



**Figure 1.** Construction of the rc-TurboID plasmid and screening of TurboID expression lines. **A)** The annotated plasmid map schematics of pLM099, L2-RBCS-TurboID and rc-TurboID. BsaI recognition sites on the ccdB gene of plasmid pLM099 is removed by a synonymous mutation introduced through amplification using primers oLM2834 and oLM427. The amplified products were joined through blunt-end ligation and sequenced to confirm mutation which is shown in the zoomed region. The mutation introduced is coloured red. **B)** Fragment 1 and Fragment 2 were then amplified using the modified pLM099 plasmid, alongside Fragment 3 amplified on the L2-RBCS2-TurboID, the three DNA fragments are joined by Golden Gate cloning method to make the resultant rc-TurboID plasmid. **C**) Restriction digest pattern of the rc-TurboID plasmid purified from two separate colonies. Sizes of DNA fragments expected to result from restriction digest are shown on the side. **D)** A schematic of the

*Chlamydomonas reinhardtii* cells and the TurboID target locations (dashed rectangles). The corresponding target protein used later for TurboID tagging and their localisation as demonstrated by confocal imaging of their fluorescent protein fusion is shown on the right. Scale bar 2 µm. **E)** Biotin labelling screen of *Chlamydomonas* cells transformed with rc-PSAF-TurboID. 4 transformed colonies were inoculated into the same well and grown until mid-log phase. Cells are sampled before and after addition of 2.5 mM biotin substrate to the cell culture for 2 hours. Biotin labelling activity was then visualized with immunoblotting protein extracts against streptavidin. Lanes of mixed culture which showed increased biotinylation due to biotin substrate addition are further isolated. The isolated cultures are then probed once more against Streptavidin and anti-HA antibody to detect for expression.

To direct proximity labelling at sites likely to be important for pyrenoid biogenesis, we identified multiple proteins either previously localised to the pyrenoid starch sheath gap (LCI9), pyrenoid tubules (RBMP2) or continuous through-out the thylakoid membrane (FZL and PSAF; Fig. 1D). LCI9 was previously found to localise to the pyrenoid starch sheath gap and was suggested to interact with multiple starch synthesis related proteins such as SBE3 (Starch branching enzyme 3), PFK1 and PFK 2 (Phosphofructokinase 1 and 2; (Mackinder et al. 2017). Their association alongside its localisation has been proposed to be important to regulate the formation of the pyrenoid starch sheath. Together this suggests LCI9 as a prime candidate to target for proximity labelling to expand our understanding on the pyrenoid starch sheath assembly process. As mentioned earlier, pyrenoid tubule traversion is a common occurrence in many pyrenoid-containing algal systems. In *Chlamydomonas*, the pyrenoid tubules exhibits a unique ultrastructure and multiple proteins have been found to exclusively localise to this specialized membrane. RBMP2 presents one of such proteins, which carries a rhodanese domain in the centre of its amino acid sequence along with 5 Rubisco binding motifs on its C-terminal section. This protein was previously hypothesised to function as a tether protein which binds or directs the Rubisco matrix to the pyrenoid tubules through their Rubisco binding motifs. RBMP2 also localises to a unique region on the pyrenoid tubules, called the reticulated knot, where the pyrenoid tubules are shown to merge and display additional complexity (Fig. 1D). RBMP2's unique characteristics, particularly their proposed tether function, make them an excellent candidate for expanding our understanding on the possible contributors of pyrenoid tubule formation as well as the pyrenoid formation mechanism.

FZL is a recently characterized membrane associated protein that shows a homogenous thylakoid localisation in *Chlamydomonas* (Findinier et al. 2019a). When we performed fluorescent protein tagging to confirm protein localisation, we found that FZL-mNeon signal does not uniformly coat the thylakoid membrane, but instead forms discrete signal enriched

on the thylakoid surface as seen by their partial overlap with the chlorophyll signal (Fig. 1D, Supplemental Fig. 2). Particularly of interest is that FZL-mNeon signal can be found at the pyrenoid position, which indicates FZL additionally lines the pyrenoid tubules due to their membrane association. As FZL is found to be critical for thylakoid fusion and is a structural determinant of thylakoid morphology (Gao et al. 2006), we hypothesised that it could participate in the conversion process of the sheet-like thylakoid into pyrenoid tubules or the fusion of pyrenoid tubule membranes at the centre of pyrenoid matrix.

To facilitate better isolation of the proteins that specifically localise to the pyrenoid tubule, a spatial control which labels thylakoid membrane associated proteins indiscriminately is necessary (Fig. 1D). PSAF was previously demonstrated to localise into the thylakoid membrane homogenously (Mackinder et al. 2017) and was chosen as a control. The PSAF protein constitutes a part of the reaction centre within the PSI (Photosystem 1) supercomplex (Farah et al. 1995; Su et al. 2019). While it is often annotated as a lumenal protein as a large part of the PSAF's structure is located at the thylakoid lumen, the recently solved crystal structure of *Chlamydomonas* PSI supercomplex (PDB: 6IJO) found PSAF's C-terminal extends towards the chloroplast stroma (Su et al. 2019). The C-terminal addition of TurboID tag to PSAF hence likely drives biotin labelling of proteins along the thylakoid membrane. Combined with its homogenous thylakoid localisation, PSAF is particularly well-placed as a thylakoid membrane spatial control to improve identification of pyrenoid tubule specific proteins.

LCI9 has previously been successfully cloned into the pLM005 plasmid for fluorescent protein tagging via PCR. Its coding sequence also contains no recognition site for BsaI and BpiI used in the Golden Gate cloning method. We therefore constructed the LCI9-TurboID plasmid using the Golden Gate cloning method as previously described (Lau et al. 2023) instead of using the newly developed rc-TurboID backbone. For FZL, RBMP2 and PSAF, these genes either have long coding regions (RBMP2 – 7669 bp) or contain multiple Type IIS enzyme recognition sites (PSAF – 4 recognition sites). This prompted us to make use of the recombineering method (Emrich-Mills et al. 2021). Briefly, primers carrying 50 bp homology regions that flanked the native promoter region (~2000 bp) and directly before its stop codon were used to amplify the TurboID-recombineering back bone. The modified backbone is transformed into *E. coli* containing a BAC corresponding to the gene of interest and a pRed plasmid necessary to drive recombination. Transformed colonies were selected on kanamycin-containing plates. The resultant plasmids were checked for correct assembly first by restriction digestion using MlyI and later sequenced (Fig. 1C). To identify colonies which express the full-length protein fused with the TurboID tag, we decided to pool transformed colonies together and incubate the mixed culture with biotin substrates. We then

extracted proteins from the mixed culture and performed immunoblotting against streptavidin to visualise biotinylated proteins. This workflow allows us to quickly detect biotinylation activity and expression of the tagged TurboID (Fig. 1E). Indeed, using the screening for PSAF-TurboID as an example, 24 transformants were pooled as 6 cultures and grown until mid-log phase. Each culture was then further split into two. Biotin substrate was added to only one which allows comparison against a biotin-absent control. Probing of labelled cells against streptavidin highlights that labelling activity is observed in two groups: strains 13 – 16 and strains 17 – 20. We therefore repeated biotin labelling individually per strains and identified that strains 14 and 16 both showed labelling activity, with strain 14 also showing the full-length expression of the PSAF-TurboID when probed against the HA antibody. Interestingly, despite the strong biotin labelling observed, full-length expression of PSAF-TurboID is not detected in strain 16. As we previously suggested that the strong biotinylated band at ~50 kD represents the Rubisco large subunit (Lau et al. 2023), the biotinylation signal observed in Strain 16 points to at least partial expression of PSAF-TurboID where the PSAF-N-terminal chloroplast transit peptide is necessary to target the mature protein into the chloroplast stroma. A likely explanation is that C-terminal HA tag is lost or degraded in the chloroplast stroma which results in the failure to detect the full-length protein by immunoblotting against the HA epitope.

**RC-TurboID drives biotinylation proportional to their expression level**

After identification of the TurboID-expressing cells, the biotinylation pattern in each expressing line was assessed individually. In agreement to previous work, native biotinylation bands are present in wildtype even prior to the biotin incubation treatment and did not show any increase with the addition of biotin. Several of the biotinylation bands found in wildtype, thus presumably natively biotinylated proteins, were absent in the TurboID expressing cells. Their absence likely reflected the depletion of internal free biotin due to the expression of TurboID proteins as previously postulated (Kreis et al. 2023). Interestingly, in LCI9-TurboID and RPE1-TurboID, both which are expressed form the PSAD promoter, strong biotinylation signals can be observed at 102 kD (LCI9-TurboID) and 67 kD (RPE1-TurboID) respectively, which correspond well to their full-length protein (Fig. 2). However, for the recombineering TurboID expression constructs, there is minimal biotinylation signals observed on position which corresponds to their full-length size (Fig. 2). One possibility is that the expression in recombineering TurboID being driven by their native promoter is much lower than that of the PSAD promoter driven LCI9-TurboID and RPE1-TurboID. We found that biotinylation signal as full-length TurboID proteins is difficult to observe in RBMP2-TurboID and FZL-TurboID. We reasoned that the expression level of these protein driven by their native promoter is relatively low. Indeed, QconCat determination of protein level in CC-

1690 finds that the endogenous protein level of RBMP2 and FZL to be 4 times smaller than that of RPE1 (Arend et al. 2023). We further determined the expression of full-length protein by immunoblotting the protein lysate against HA tag. Here we found corresponding full-length protein representing LCI9-TurboID, RPE1-TurboID, FZL-TurboID as well as PSAF-TurboID. RBMP2-TurboID is predicted to be a high molecular weight protein (205 kD) which we could not determine their expression with immunoblotting against the HA-tag. This difficulty in demonstrating their expression using epitope-tag has previously been demonstrated, where despite clear observation of fluorescence signal in confocal microscopy, resultant immunoblotting results show only weak signals (Meyer et al. 2020). We argue that since RBMP2-TurboID labelling has still been able to biotinylate the Rubisco large subunit, the expression of full-length RBMP2 is likely weak but present. We additionally saw multiple bands when immunoblotting FZL-TurboID against the HA-tag possibly representing truncated FZL-TurboID proteins. However, the most dominant band observed still corresponds well to the full-length product of FZL-TurboID (148 kD).

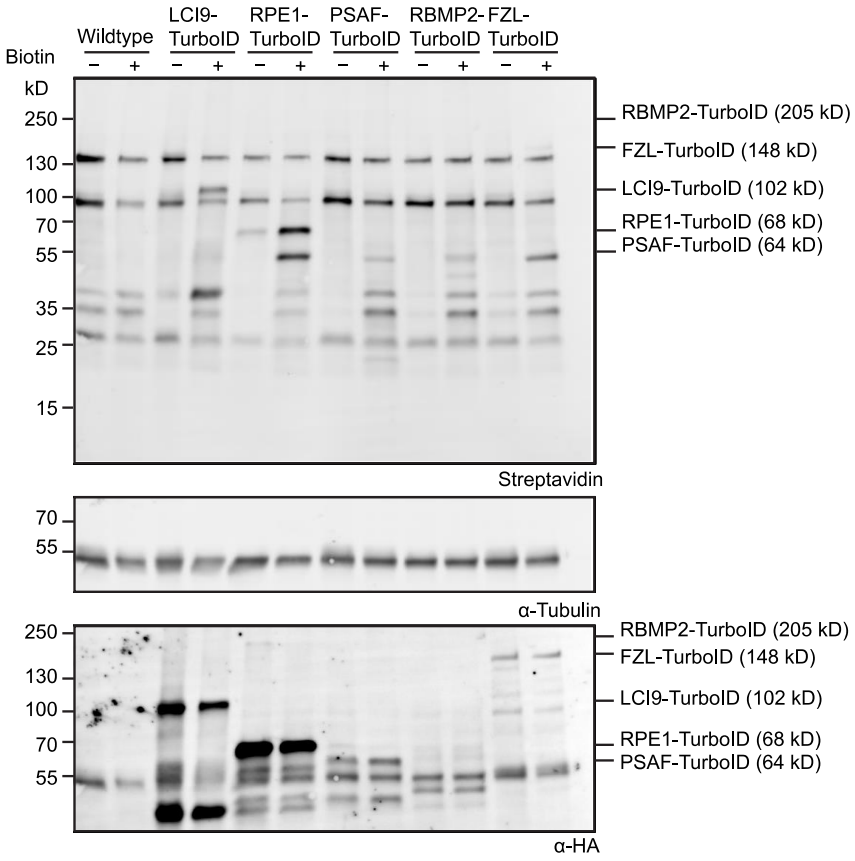



**Figure 2.** Assessment of biotin labelling activity of TurboID expression strains submitted for mass spectrometry analysis. The labelling activity of the TurboID expression strains, as determined in the absence (−) or presence (+) of 2.5 mM biotin for 2 h. Cultures are separately harvested both before and after the biotin labelling treatment. Biotinylation was visualized via immunoblotting whole-cell lysate with a streptavidin conjugate. The abundance of LCI9-TurboID (102 kD), RPE1-TurboID (67 kD), PSAF-TurboID (64 kD), RBMP2-TurboID (205 kD), and FZL-TurboID (148 kD) was probed by α-HA. α-tubulin was used as a loading control.

### LCI9-TurboID highlights a suite of starch-related genes and further implicating its role in starch sheath assembly process

In contrary to previous work where we make use of data-dependent acquisition mode incorporated with TMT-labelling for quantification of protein abundance. Data-independent acquisition method was chosen for the acquisition of MS data due to their improved replicability and sensitivity (Doerr 2015). The five TurboID tagged strains as well as wildtype were submitted in triplicates for mass spectrometry analysis. In total we identified 4610 proteins groups where proteins contain at least 2 unique peptides. Assessment of the total

protein level between submitted samples show that all replicas do not vary drastically and also generally correlate well within replicates (Supplemental Fig. S1A and S1B).

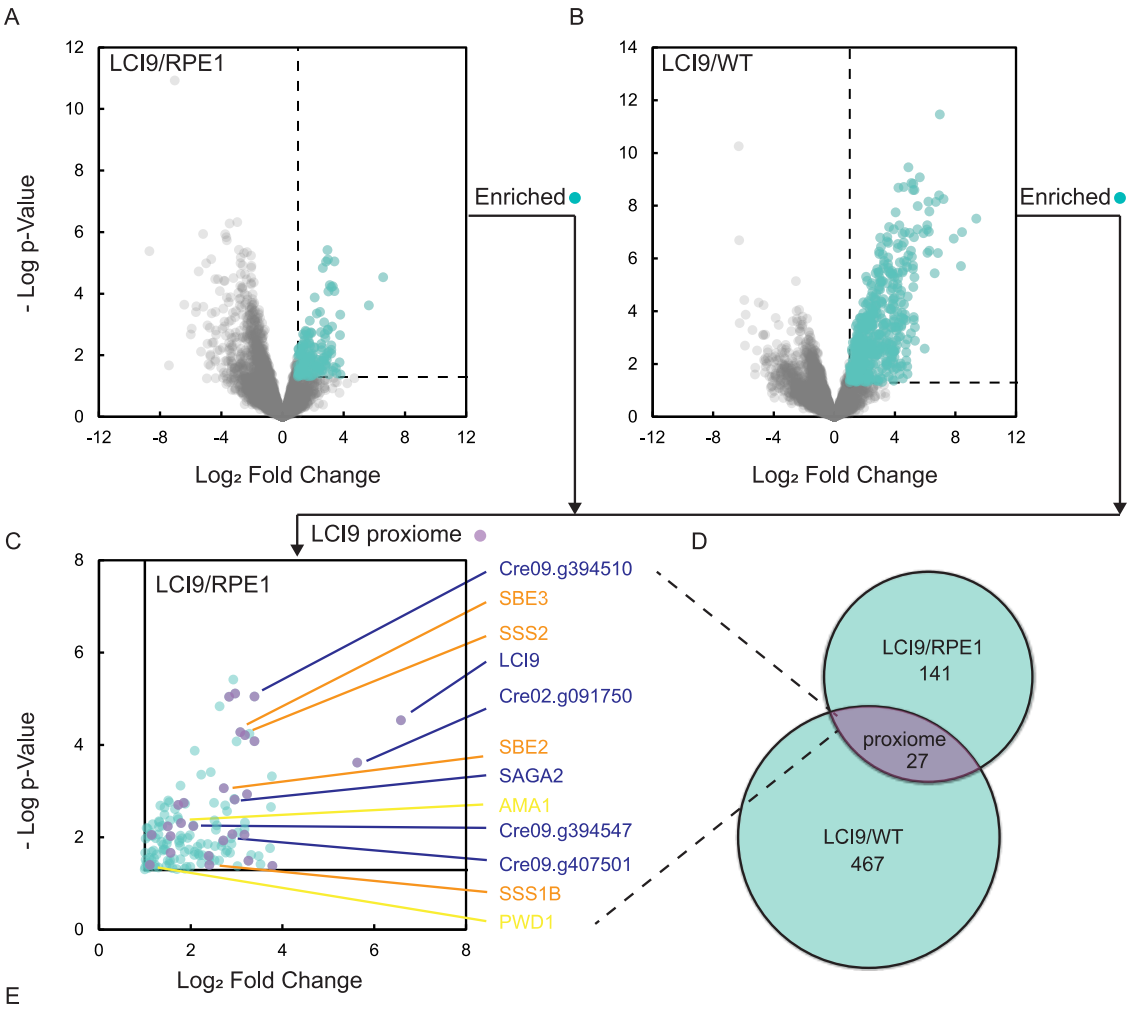


**Figure 3.** LCI9 TurboID labelling affirms its role in starch metabolism. Protein enrichment from LCI9-TurboID is compared to RPE1-TurboID **A)** or wildtype **B)**. $Log_2$ Fold Change of protein abundance was calculated for detected proteins and visualised as a scatterplot against the Log p-value. The enrichment threshold of $Log_2$ Fold Change >1 and p <0.05 is shown as the dotted line. Proteins considered significantly enriched are coloured in green. This yields 139 proteins in LCI9/RPE1 and 414 proteins in LCI9/WT respectively. **C)** To obtain a more stringent LCI9 proxiome, enriched proteins in LCI9/RPE1 were compared to LCI9/WT. Proteins which are found to be present in both groups are placed in the LCI9 proxiome (purple). These proteins are then shown in the inset of LCI9/RPE1 where starch-related proteins are highlighted according to function (synthesis: orange; degradation: yellow; CBM-containing: purple). **D)** Venn diagram demonstrating the comparison between significantly enriched proteins in each comparison group. **E)** Alphafold modelling of the uncharacterised starch-related proteins identified from LCI9-proxiome. Black arrow highlights the annotated CBM20 domain (orange). α-helical regions are coloured in turquoise. The annotated domains of each protein is displayed on the schematic of amino acid sequence on the right.

To highlight proteins which contribute to the starch sheath formation, LCI9-TurboID protein enrichment was compared against wildtype and RPE1-TurboID (which we referred to as LCI9/WT and LCI9/RPE1 respectively). An enrichment threshold of $log_2$FC above 1 was applied to the both datasets to obtain a total of 467 proteins in LCI9/WT and 141 proteins in LCI9/RPE1 where the p-value is below 0.05. To further filter the dataset (Figure  and Figure ), we adopted a similar strategy as previous work (Lau et al. 2023) and compared the obtained proxiome between LCI9/WT and LCI9/RPE1 which finds a total of 27 proteins to be enriched in both comparison groups (Fig. 3C and 3D; Supplemental Data Set 2). Additionally, highlighting the proteins which possibly contribute to the starch sheath synthesis, we highlighted proteins which contain CBM (Carbohydrate-binding family member) domains or are annotated as starch synthesis and degradation proteins (Figure 3C; Synthesis – Orange; Degradation – Yellow; CBM – Purple). The LCI9 protein itself is found to be the most enriched in both proxiomes as is expected due to self-biotinylation by the LCI9-TurboID fusion protein. We found that SBE3 (STARCH BRANCHING ENZYME 3 – Cre10.g 444700) is among the 27 high confidence proximal proteins, which shows agreement to previous affinity purification study (Mackinder et al. 2017). Interestingly, the same study identified PFK1 and PFK2 as high-confidence interacting protein, which are enzymes responsible for a commitment step of glycolysis by diverting Fructose-6-phosphate to Fructose-1,6-biphosphate for further break down. However, in the present study these proteins are only preferentially enriched in LCI9/WT and depleted in LCI9/RPE1 suggesting their association in the chloroplast is possibly either weak or transient. The high-confidence proximal protein of LCI9 include many canonical starch-related proteins besides SBE3,

proteins responsible in starch synthesis process such as: SBE2 (STARCH BRANCHING ENZYME 2 – Cre06.g270100), SSS2 (SOLUBLE STARCH SYNTHASE 2 – Cre03.g185250), SSS1B (SOLUBLE STARCH SYNTHASE IB – Cre04.g215150); or starch degradation: AMA1 (ALPHA-AMYLASE 1 – Cre08.g385500) and PWD1 (PHOSPHOGLUCAN WATER DIKINASE 1 – Cre17.g719900); or starch morphology: SAGA2 (STARCH GRANULE ABNORMAL 2 – Cre09.g394621) are amongst those identified. Further exploring starch associated proteins, we found multiple uncharacterised proteins that also carry a CBM20 domain: Cre02.g091750, Cre09.g407501, Cre09.g394510 and Cre09.g394547. Notably, the identification of Cre09.g394510 and Cre09.g394547 is particularly interesting as both genes are at the same genomic locus as *LCI9*. Alphafold modelling of these proteins shows that they share a similar structure where the CBM20 domain is followed by a long α-helix domain (Fig. 3E, black arrow). Additionally, Cre09.g394510 was previously localised to the starch sheath gaps of the pyrenoid (Lau et al. 2023) further supporting their similarity. Together, their close genetic and structural similarity hints that these proteins are likely homologues, Cre09.g394510 and Cre09.g374547 are explored later in Chapter 4. On the other hand, while Cre02.g091750 and Cre09.g407501 also shares the comparable domain arrangement, the α-helix domains in these two proteins are fragmented and separated by small, disordered regions which deviates from the LCI9 homologues (Fig. 3E). A recent report characterised a mutant of Cre02.g091750, which is designated as BSG1 (BIMODAL STARCH GRANULE 1), finds the protein to likely be responsible for pyrenoid starch degradation during nutrient starvation condition (Findinier et al. 2019b). Cre02.g091750 also carries a predicted SAND domain (Fig. 3E, lime green) which has been shown to regulate transcriptions through DNA binding (Bottomley et al. 2001), suggesting that it might display additional DNA-related function. Compared to Cre02.g091750 and LCI9, Cre09.g407501 has multiple short helical regions separated by regions with predicted disorder (Fig. 3E). Fluorescent protein tagging of this protein show that it localises into the chloroplast homogenously with no signal within the pyrenoid (Wang et al. 2023), which suggests Cre09.g407501 might not participate in the pyrenoid starch sheath synthesis exclusively. The *Arabidopsis* PTST2 protein presents a similar domain architecture, CBM48-and-helical motif, and has been found to be important in starch synthesis (Seung et al. 2018). Interestingly, PTST2 also localises into chloroplast stroma patches that do not always correlate to starch granules suggesting that starch-related proteins might not always associate with starch granules despite their related activity. Cre09.g407501 function in starch synthesis should therefore not be precluded due to its localisation and warrants further characterisation. Besides the starch-related proteins, multiple proteins containing Rubisco binding motifs (RBMs) are found in the LCI9 proxiome. SAGA2 exhibits 5 RBMs and is hypothesised to remodel starch sheath morphology by

mediating the interactions between the pyrenoid matrix and starch sheath via the RBM. Its presence in the LCI9 proxiome highlights a possible interaction between the two proteins to regulate the pyrenoid starch sheath morphology. Cre08.g358563 and Cre03.g179961 contains 2 and 3 RBMs respectively. Cre08.g358563 is annotated as an ATP-dependent RNA helicase, while Cre03.g179961 is annotated as a DNA polymerase epsilon subunit 1. Both proteins contain RNA and DNA-binding domains which echoes the previous discovery of multiple RNA-binding proteins in the pyrenoid proxiome and further supporting the pyrenoid's hypothesised role in RNA related processes. Collectively, the LCI9 proxiome highlighted a suite of starch-related proteins thus providing exciting candidates for characterisation of the starch sheath assembly process.

**RBMP2 proxiome identifies many known pyrenoid proteins and some new pyrenoid candidates**

Pyrenoid penetrating thylakoids are a common feature in many pyrenoid-containing eukaryotic algae. However, the precise molecular function of this specialised membrane structure is still largely unclear. Elucidating the pyrenoid tubule proteome likely provides additional clues to their function and could possibly inform their assembly process. To obtain the pyrenoid tubule proxiome, the calculated protein enrichment from RBMP2-TurboID was compared to that of wildtype, RPE1 and PSAF (Fig. 4A). The filtering of enriched proteins by RBMP2/wildtype and RBMP2/RPE1 comparison groups has effectively removed most proteins enriched by streptavidin beads and non-chloroplast targeted proteins. The additional inclusion of the RBMP2/PSAF filter further set apart enriched proteins that does not specifically associate with the pyrenoid tubules. Together these stringent filtering yields a total of 13 proteins in the RBMP2 proxiome (Fig. 4B and 4C; Supplemental Data Set 3). Like the previous pyrenoid proxiome obtained from RBCS2-TurboID, a high proportion of proteins within the RBMP2 proxiome contains RBMs (10 out of 16 proteins). Among these 16 proteins, 10 have also been previously identified as known pyrenoid proteins, they include rbcL (RUBISCO LARGE SUBUNIT), Cre03.g172700, BST4/RBMP1 (BESTROPHIN-LIKE PROTEIN 4/RUBISCO BINDING MEMBRANE PROTEIN 1 – Cre06.g261750), SAGA3 (STARCH GRANULE ABNORMAL 3 – Cre06.g259100), RBMP2, , EPYC1 (Cre10.g436550), CSP41A (CHLOROPLAST STEM-LOOP BINDING PROTEIN 41A – Cre10.g440050), CSP41B (CHLOROPLAST STEM-LOOP BINDING PROTEIN 41B – Cre10.g435800), Cre10.g430350 and RCA2 (RUBISCO ACTIVASE-LIKE PROTEIN 2 – Cre17.g718950), and thus provides high accuracy in the identification of pyrenoid proteins (Figure ; purple). Many of the RBMP2 proxiome proteins have previously been characterised and would be expected to be enriched within the RBMP2 proxiome, such as Rubisco which is known to interact with RBMP2 via its RBMs (Meyer et al. 2020). We therefore looked in

detail at some of the less explored proteins (Cre03.g172700, Cre03.g169300, Cre02.g097250, Cre12.g530950 and Cre16.g688302) (Fig. 4B; black). Cre02.g097250 encodes a protein which shows homology to the importin-β protein family. Importins are a class of carrier protein primarily responsible for nucleocytoplasmic protein transport through the recognition of nuclear targeting sequence with its interacting partners importin-α. After binding the cargo protein, the importin protein complex further interacts with the nuclear pore complex to translocate the bound cargo protein (Oka and Yoneda 2018). The importin-β annotation is further supported by the Alphafold modelling results, which found the modelled structure of Cre02.g097250 carries multiple armadillo-fold like helical hairpins, which is highly similar to the Human importin-β structure (PDB: 2q5d; (Mitrousis et al. 2008); Fig. 4E). Cre02.g097250 presence in the RBMP2 proxiome suggests that this protein might be chloroplast targeted instead. The pyrenoid tubules contain small membrane structure termed minitubules formed from the thylakoid invagination and are hypothesised to mediate metabolite transport across the pyrenoid matrix and external chloroplast stroma (Engel et al. 2015). With its activity in protein translocation, one hypothesis is that Cre02.g097250 aid in shuttling proteins to and from the pyrenoid matrix possibly through the small pyrenoid minitubule opening (~3.5 nm) which allows efficient protein turnover. Cre03.g169300 might similarly share a transport function as it carries an ABC-transporter domain. Particularly, it bares similarity towards the ABCG2 transporters which exhibit transport capability to large array of substrates: including multiple drugs and lipid substrates such as cholesterol (Kerr et al. 2021). Their similarity is also showcased in the alphafold modelling of Cre03.g169300, where Cre03.g169300's ATP-binding domain (Fig. 4E, black arrows) and transmembrane domain (Fig. 4E, blow arrows) echoes that of the human ABCG2 structure (PDB: 5nj3; (Taylor et al. 2017) . Cre03.g169300 presence at the pyrenoid tubules might promote substrate shuttling between the thylakoid lumen and pyrenoid matrix to facilitate better efficiency of the Calvin cycle. Additionally, exposure to limiting $C_i$ conditions triggers a spike in calcium levels at the pyrenoid matrix that is important for CCM-related gene induction, this process likely require a dedicated pyrenoid-localised calcium transporter (Wang et al. 2016). The ABC-type transporter identified here could potentially play a role in this regulatory response. Cre16.g688302 is annotated as a bifunctional nuclease. The modelled structure of Cre16.g688302 also aligns well to the *Thermotoga maritima* TM0160 protein (PDB: 1sj5), which shares the bifunctional nuclease domain annotation (Spraggon et al. 2004). Bifunctional nucleases in plants have been implicated in the nucleic acid degradation during leaf senescence (Pérez-Amador et al. 2000). Previous work found oxidised RNA is present in the pyrenoid (Zhan et al. 2015); multiple RNA-associated proteins are also identified in the pyrenoid proxiome, which supports the pyrenoids possible role in RNA metabolism (Lau et al. 2023). If localised to the pyrenoid, Cre16.g688302 might help regulate the accumulated

RNA within the pyrenoid matrix during stress response. Cre12.g530950 is also identified in the RBMP2 proxiome. The encoded proteins correspond to a triacylglycerol lipase which breaks down TAG into diacylglycerol molecules. The alphafold modelling of Cre12.g530950 also resembles the solved structure of *Solanum cardiophyllum* lipase Patatin-17 (PDB: 4pk9; (Wijeyesakere et al. 2014). Triacylglycerol (TAG) production is shown to increase during $CO_2$ limiting conditions. Their associated lipid remodelling is implicated in the transient oxidative stress protection (Abreu et al. 2020). We hypothesised that Cre12.g530950 is sequestered within the pyrenoid which avoids the undesirable catalytic breakdown of TAG within the lipid bodies in the chloroplast stroma. Alternatively, Cre12.g530950 might alter the pyrenoid tubules' lipid composition, which creates a specific membrane property which is important for recognition by the pyrenoid tubule proteins.

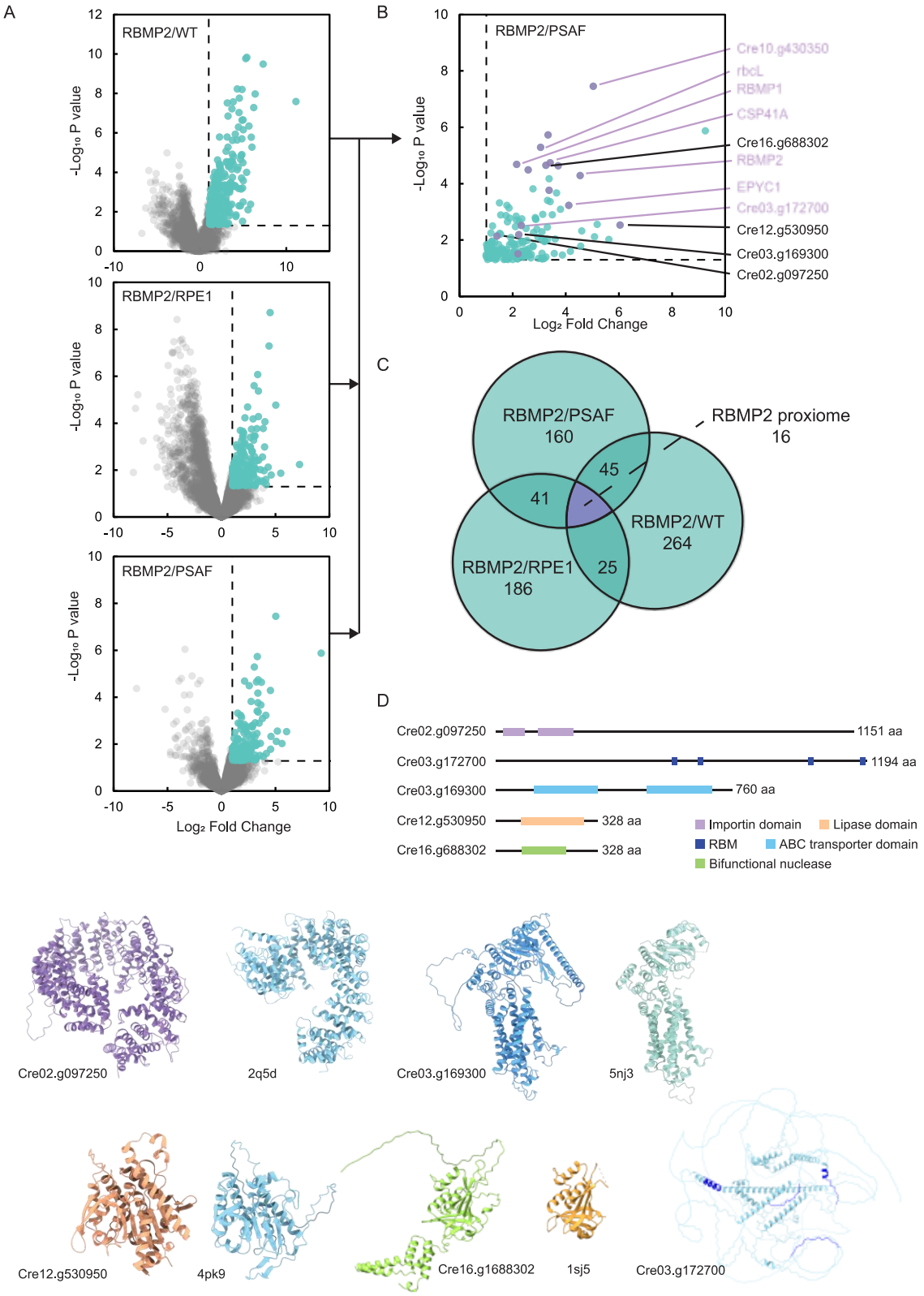



**Figure 4.** Analysis of the RBMP2 proxiome. **A)** Volcano plot of the protein enrichment in RBMP2 compared to wildtype (top), RPE1 (middle) and PSAF-TurboID (bottom). $Log_2$ Fold Change of proteins are plotted against the -log10 transformed p-value. The empirical thresholds of $Log_2$ FC > 1 and p-

value <0.05 (vertical and horizontal dashed line respectively) was applied to obtain the significantly enriched proteins (green). **B)** Proteins which lie above the enrichment threshold in all three comparisons are classified as within the RBMP2 proxiome (purple). The notable proteins (black), along with those previously identified pyrenoid proteins (purple) are labelled on the right. **C)** Venn diagram demonstrating the comparison between significantly enriched proteins in each comparison group. E) Alphafold 3 modelling of the uncharacterised proteins identified from RBMP2-proxiome. Annotated domains are highlighted on the schematic of amino acid sequence on the right. A reference structure carrying the same domain annotations are shown on the right of the respective modelled structure. Cre02.g097250 and 2q5d share the importin-β domain, chain A of the 2q5d is displayed here (Mitrousis et al. 2008). Cre03.g169300 and 5nj3 share the ABCG transporter annotation, chain A of the 5nj3 is displayed here (Taylor et al. 2017). Cre12.g530950 and 4pk9 share the Patatin/lipase domain (Wijeyesakere et al. 2014). Cre16.g688302 and 1sj5 both carries the Bifunctional nuclease domain, chain A of the 1sj5 is displayed here (Spraggon et al. 2004).

Cre03.g172700 is predominantly disordered protein that was previously localised to the pyrenoid tubules due to its punctate localisation pattern (Lau et al. 2023). Combined with 4 RBMs spaced along its amino acid sequence, Cre03.g172700 showed promise as a possible pyrenoid matrix tether protein. CRISPR-Cas9 mediated mutation was performed on the Cre03.g172700 locus and 4 independent mutant lines were obtained (Fig. 5A – B). However, none of the mutant lines present a CCM-dependent growth phenotype (Fig. 5C) suggesting Cre03.g172700 is unlikely an essential pyrenoid component and was therefore not further characterised.

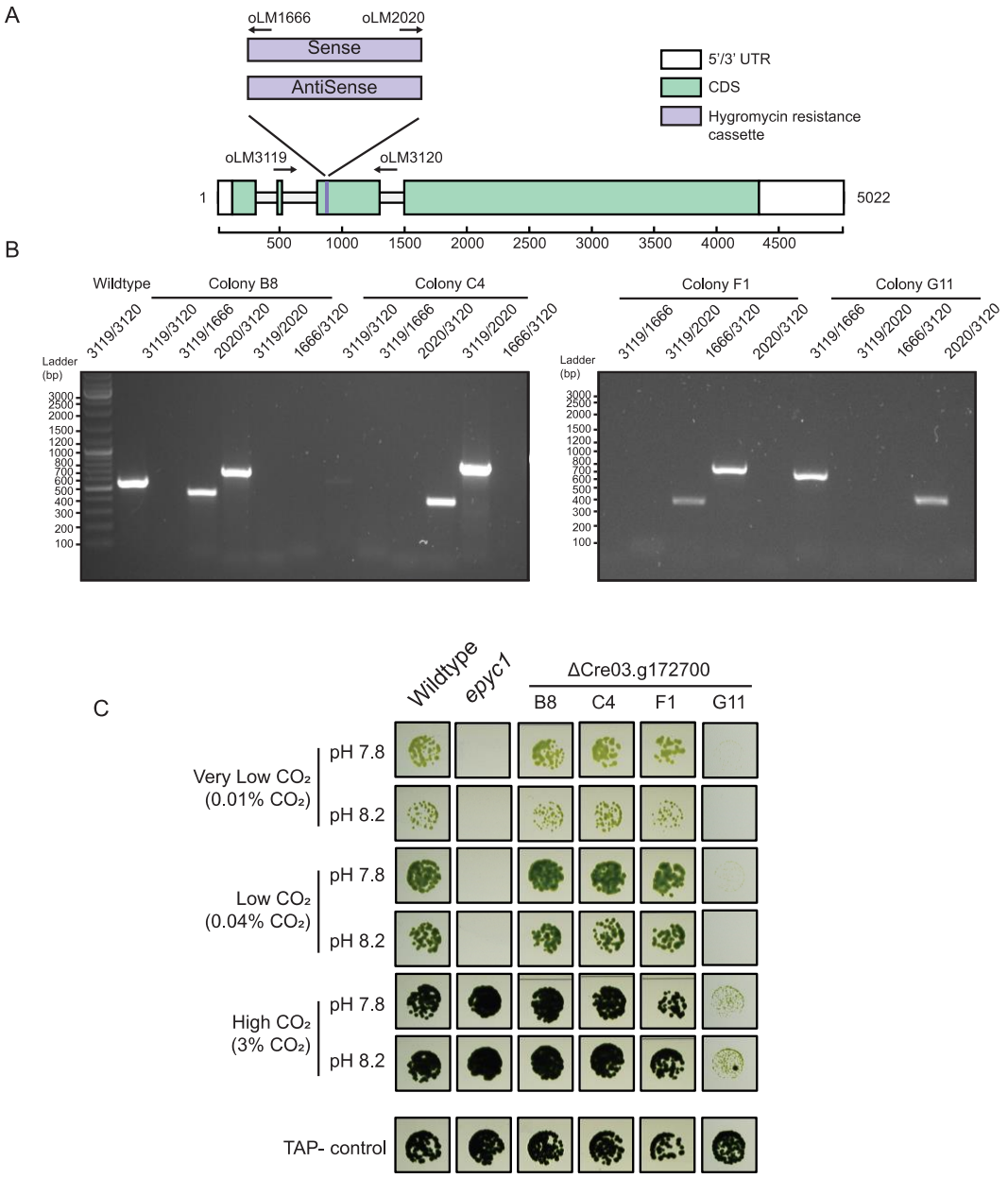



**Figure 5.** CRISPR mutant generation of Cre03.g172700 and its phenotyping. **A)** A schematic of the Cre03.g172700 gene model is shown to scale. The CRISPR-Cas9 targeting region is coloured in Yellow, the 5'/3' UTR is shown in white, CDS regions in green and the hygromycin resistance cassette in purple. The primers binding region are labelled on the schematic. **B)** The co-transformed hygromycin resistance cassette can be inserted either in sense or anti-sense orientation. To validate the insertion orientation at the targeted location, amplification around the targeted region is performed using different primer pairs as shown on top of the DNA gel. Results shows that lines B8 and G11 contain an insertion cassette in the sense orientation while the lines C4 and F1 contains the insertion cassette in the anti-sense orientation. **C)** $CO_2$ requiring phenotype screen of the ΔCre03.g172700 mutants. wildtype, *epyc1* and the ΔCre03.g172700 mutants are grown on TP-agar plates adjusted to either 7.8 or pH 8.2. Plates are placed in air-tight chambers with $CO_2$ supplied to a concentration of

3%, 0.04% and 0.01% $CO_2$ separately. Light intensity is maintained under 400 µmol photons $m^{-2}$ $s^{-1}$. None of the mutants are found to exhibit a growth defect at low $CO_2$ in comparison to the wildtype strain. However, the strain G11 exhibits a photosynthetic defect which is likely caused by insertion of hygromycin resistance cassette in another gene location.

**FZL proxiome contains many protein related to stress response**

The FZL proxiome was obtained as mentioned above (Fig. 6A and 6B). In total 12 proteins including FZL are highlighted after filtering all enriched proteins using the FZL/WT, FZL/RPE1, and FZL/PSAF comparison groups (Fig. 6C; Supplemental Data Set 4). Interestingly, multiple proteins within this dataset exhibit stress response element, such as HSP22C (HEAT SHOCK PROTEIN 22C – Cre03.g145787); stress-enhanced protein Cre03.g199535; senescence-related protein Cre08.g374100; AKC4 (ABC-1 LIKE KINASE – Cre13.g570350); a DnaJ domain containing protein Cre12.g801337. As *Chlamydomonas* FZL was suggested to function in thylakoid membrane fusion, they are likely found in proximity to newly merged thylakoid membranes where the insertion of nascent photosynthetic proteins is needed. The occurrence of chaperone proteins such as HSP22C are likely similarly positioned at the newly merged membrane to ensure coordinated folding and insertion. Surprisingly, while we expected the FZL proxiome to highlight membrane-associated proteins that could facilitate its function in thylakoid fusion, none of the highlighted proteins except for AKC4 possess a transmembrane domain. AKC4 is annotated as an ABC1 atypical kinase-like protein and was found in the eyespot proteome in *Chlamydomonas* (Eitzinger et al. 2015). As the majority of the ABC1-like kinase homologs in *Arabidopsis* are plastoglobule-associated, AKC4 was also hypothesised to share this association (Eitzinger et al. 2015; Rottet et al. 2015). We surmise that FZL might leverage their association with AKC4 and promote the lipid delivery from plastoglobules towards thylakoid membranes, which will be particularly important for these newly synthesised membranes structures. While not exhibiting any transmembrane domains, Cre09.g416350 is annotated to contain a RAD50-interacting protein (RINT-1) domain. The mammalian RINT-1 is important for the ER – Golgi vesicular transport system, where RINT-1 is shown to regulate SNARE complex formation by RINT-1's interaction with the Syntaxin 16 protein (Arasaki et al. 2013). It is attractive to hypothesize that the association between FZL and Cre09.g416350 pointed to a similar regulatory mechanism. However, further validation is necessary to show their interaction *in vivo* and additional work is required to characterise their precise function.

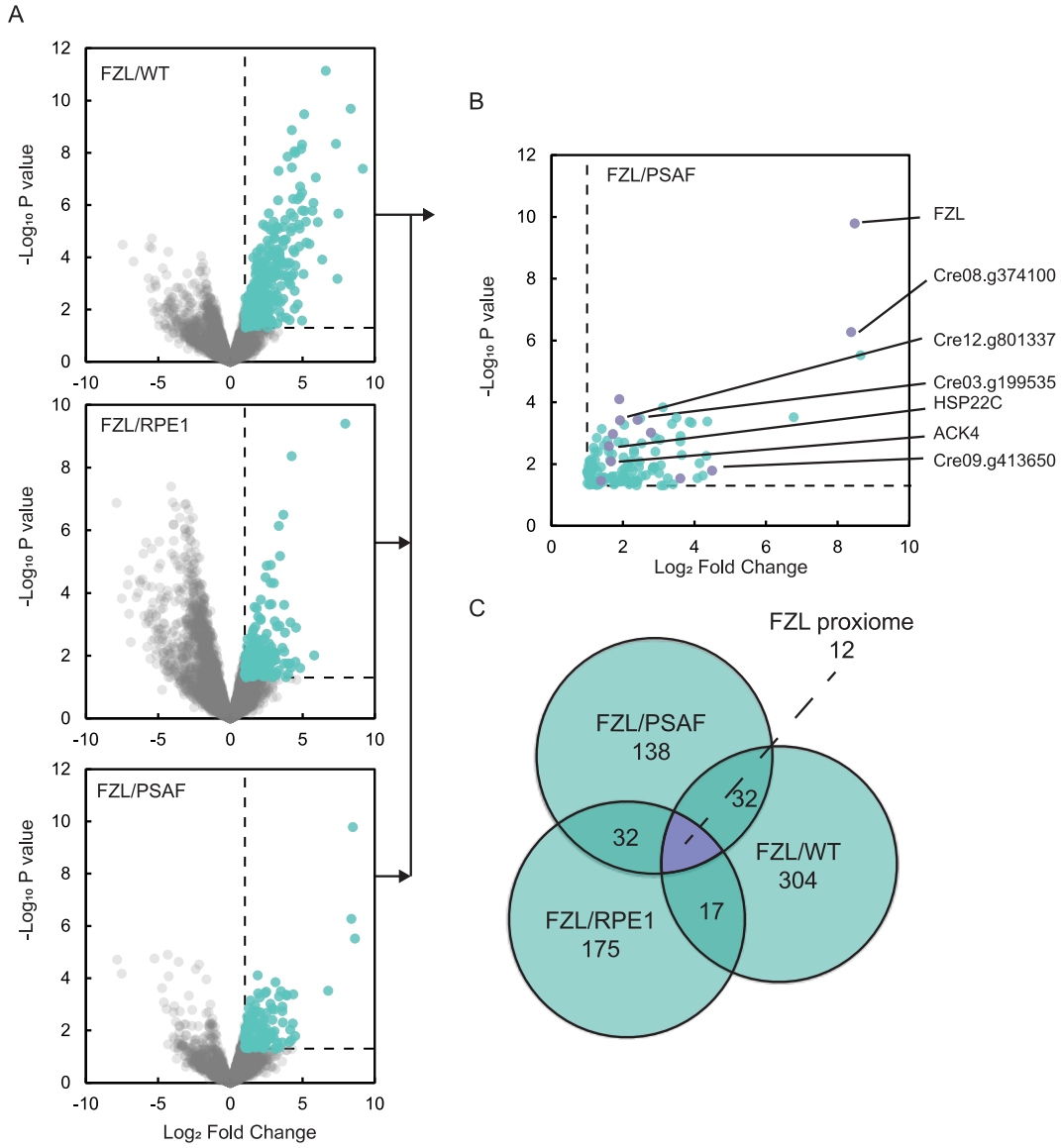



**Figure 6.** Analysis of the FZL proxiome. **A)** Volcano plot of the protein enrichment in RBMP2 compared to wildtype (top), RPE1 (middle) and PSAF-TurboID (bottom). $Log_2$Fold Change of proteins are plotted against the -$log_{10}$ transformed p-value. The empirical thresholds of $Log_2$FC > 1 and p-value <0.05 (vertical and horizontal dashed line respectively) was applied to obtain the significantly enriched proteins (green). **B)** Proteins which lie above the enrichment threshold in all three comparisons are classified as within the FZL proxiome (purple). The notable proteins, such as those mentioned in text below are labelled on the right. **C)** Venn diagram demonstrating the comparison between significantly enriched proteins in each comparison groups

## Discussion

To explore the different assembly factors required for building the two distinct layers of the *Chlamydomonas* pyrenoid, we first targeted LCI9 due to its unique localisation at the starch sheath, which likely represents the synthesis point of starch plates. Analysis of the LCI9

proxiome indeed reveals a suite of canonical starch associated proteins. Puzzlingly, instead of preferentially enriching only either starch synthesis or degradation proteins, both classes of proteins are found to be enriched by LCI9-TurboID which complicates the direct assignment of LCI9 molecular function towards either process. One might argue that many of the starch synthesis proteins such as SSS2 or SBE3 display a higher level of enrichment to that of starch degradation proteins, and thus inferring LCI9's role in starch synthesis. However, the Qconcat-derived protein level measurements indicate that the identified starch synthesis proteins are ~ 2 – 4 times higher than their counterparts, which increases the probability of their enrichment by LCI9 (Arend et al. 2023). This highlights that direct comparison on enrichment levels might unfairly discriminate against proteins of low abundance. Additionally, the presence of these starch modifying enzymes in the LCI9 proxiome could also stem from their shared substrates rather than a direct association with LCI9. In this work we identified four CBM containing proteins which share a similar domain arrangement to LCI9: a CBM20 domain followed by an α-helix region. We found that two of the CBM20 domain containing genes are situated next to LCI9 on chromosome 9, making them likely homologues born from a gene duplication event. In total 5 of these LCI9-like proteins can be found in the LCI9 proxiome, which deepens their association with starch sheath metabolism. Further characterisation of these CBM20 domain containing genes in conjunction with LCI9 might help elucidate the starch sheath assembly process.

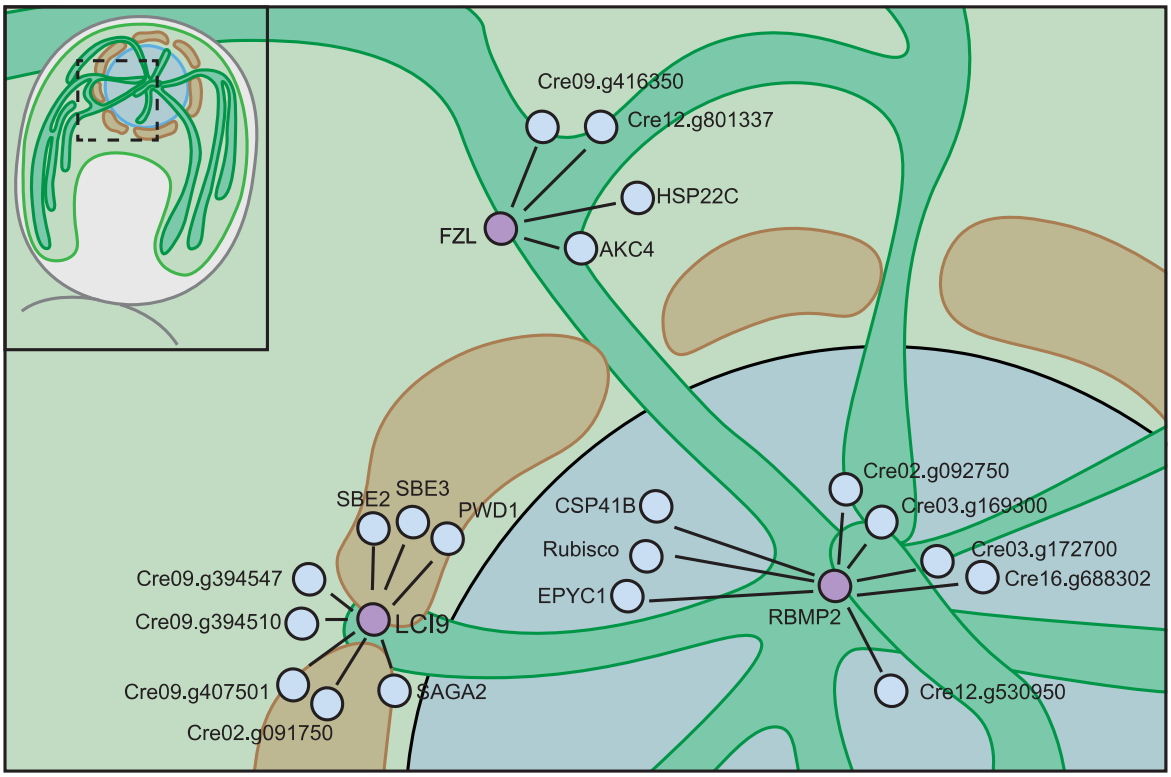



**Figure 7.** Summary of the obtained proxiome from the proximity labelling of LCI9, RBMP2 and FZL. The schematic representation of the Chlamydomonas cells is shown on the top left. A part of the chloroplast which includes the pyrenoid, starch sheath and the central thylakoid knot is shown within the schematic is enlarged. The bait proteins LCI9, RBMP2 and FZL (purple) and the notable proteins found in the respective proxiome (light blue) are linked by a black line representing their putative proximity.

Aside from informing LCI9 possible function, the LCI9 proxiome also offered some additional insights to the cell's starch metabolism. Three starch branching enzyme homologues are present in *Chlamydomonas*, they include a type 1 branching enzyme: SBE1; and two type 2 branching enzymes: SBE2 and SBE3. Recent work on the starch branching enzymes suggests that SBE2 and SBE3 are the main branching enzyme that act on starch substrates as *sbe2* and *sbe3* mutants exhibit significantly different amylose content and amylopectin chain length within purified starch granules, while the *sbe1* mutant presents a wildtype-like starch composition (Courseaux et al. 2023). In agreement to this, SBE2 and SBE3 can be identified in the LCI9 proxiome (Fig. 3C and 7), while SBE1 was even found to be depleted in LCI9-TurboID which supports that SBE2 and SBE3's role in modifying starch composition and that SBE1 association to starch might either be highly conditional or minimal.

The RBMP2 localisation at the pyrenoid tubule makes it a valuable target to explore the protein composition of this specialised membrane structure. Different from the analysis of LCI9-TurboID, PSAF-TurboID was additionally introduced as a second spatial control in conjunction with RPE1-TurboID, which we envisioned would impose higher specificity for membrane-related proteins in the RBMP2 proxiome and potentially remove the stromal proximal proteins to RBMP2. To our surprise, despite accurately identifying many pyrenoid proteins, the RBMP2 proxiome presented a striking similarity to that of RBCS2, where many presumably pyrenoid matrix-localised proteins, such as EPYC1, CSP41B and rbcL are found (Fig. 4B, 4C and 7). We reasoned that the recovery of mostly non-membrane bound proteins could stem from the TurboID tag's placement on RBMP2 C-terminal. Such C-terminal placement positioned the TurboID tag ~770 residues away from the annotated transmembrane helix, which extends the TurboID domain well in to the pyrenoid matrix, therefore driving biotin labelling of the stromal-localised proteins. A recent report suggests that the accessible surface of the intrinsically disordered domains increase their propensity for biotinylation (Minde et al. 2020). The recovery of pyrenoid proteins in the RBMP2 proxiome, many of which carry both RBMs and intrinsically disordered domains, can therefore be expected. Cre03.g172700 represents one such protein exhibiting 4 RBMs along its disordered region and was enriched in the RBCS2 and RBMP2 proxiomes. While we fail to observe any CCM-based growth phenotype in the Cre03.g172700 mutants, this does not preclude Cre03.g172700's functioning in the CCM. A mutant of the known CCM components, LCIA, has similarly lacked an observable CCM-growth phenotype despite widely recognised as a chloroplast envelope bicarbonate channel which drives bicarbonate transport from the cytosolic space into the chloroplast stroma (Wang and Spalding 2014; Yamano et al. 2015). With Cre03.g172700's unique domain arrangement, we encourage future works on this mutant particularly on characterising their pyrenoid structure which was not performed here.

We also reported a more precise localisation of the FZL protein showing that it forms discrete enrichments on the thylakoid membrane and further confirmed its presence at the pyrenoid, presumably on the pyrenoid tubules (Fig. 1D). It is attractive to hypothesise that the membrane associated with the FZL enrichment likely represents thylakoid membranes that are in an active membrane fusion process. *Chlamydomonas* thylakoid membrane fusion is unlikely to occur in such large scale without external stimuli such as heat or light stress (Zhang et al. 2022; Broderson et al. 2024). Instead, the stable localisation of FZL echoes that of thylakoid curvature proteins (CURT1s), which are inserted into the thylakoid membrane to stabilise the highly curved grana margins (Armbruster et al. 2013). We thus hypothesise that, in addition to its putative role in thylakoid membrane fusion, FZL might also

function to stabilise membrane structures. The proximity labelling of FZL-TurboID was found to yield very few proteins in the FZL proxiome. While this might be a consequence of the multiple filtering process we implemented, true proximal proteins that are closely associated with FZL should remain. Among those highlighted, only ACK1 was found to carry a predicted transmembrane domain and therefore suggests that FZL's membrane-fusion activity might not require accessory proteins (Fig. 6B, 6C and 7). FZL's association with ACK1 was particularly interesting, as *Chlamydomonas* plastoglobules were not found to associate with thylakoid membranes extensively which deviates from those in flowering plants (Austin et al. 2006; Engel et al. 2015). Combined with FZL's punctate localisation around thylakoid membranes and proximity to plastoglobule associated ACK1, it could be envisioned that plastoglobules might be transiently recruited to thylakoid membranes during FZL-driven membrane fusion to supply membrane lipids.

In this present study, we have expanded the recombination-based cloning toolkit to allow the tagging of large and complex genes with the TurboID tag. The modified plasmid includes the incorporation of the target gene's endogenous promoter that aims to drive tagged protein expression in a native-like level. Such native expression avoids perturbing the regular cellular function as ectopic expression of essential proteins can lead to cellular defects (Prelich 2012). However, we found that this low expression of the target proteins FZL and RBMP2 complicates their detection and in extension the validation of the expression strain (Fig. 1E and 2). Use of this cloning pipeline for TurboID proximity labelling will therefore necessitate careful validation prior to performing labelling experiments. Nevertheless, our results here described an advancement of the proximity labelling technique within *Chlamydomonas.* Particularly, the apparent difference between the obtained proxiome between LCI9 and RBMP2, despite their adjacency to the pyrenoid matrix, further exemplified the spatial resolution of this labelling technique. We also outlined a collection of interesting candidates where upon further characterisation will possibly provide clues to the enigmatic pyrenoid assembly process.

## Methods

### Growth and maintenance of *Chlamydomonas* strains

*Chlamydomonas* cultures were maintained on TAP-agar medium with revised Hunter's trace elements (Kropat et al. 2011). For liquid cultures, *Chlamydomonas* was grown on an orbital shaker at 140 rpm at ~21 °C under LED lights (Valoya C65 LEDs with AP673L spectrum) at ~50 µmol photons $m^{-2}$ $s^{-1}$ unless stated otherwise.

### Molecular Cloning

To construct the recombineering-based TurboID plasmids (rc-TurboID): rc-PSAF-TurboID, rc-FZL-TurboID and rc-RBMP2-TurboID, primers are designed to flank the Venus fluorophore on the pLM099 plasmid (Emrich-Mills et al. 2021) which amplifies outward in two parts to yield a fluorophore-less fragment with BsaI restriction sites. The TurboID coding sequence was amplified from pLM899 (Lau et al. 2023) to add the identical BsaI syntax. To anneal the three DNA fragments, a Golden Gate reaction with BsaI was performed on the purified DNA from the three PCR reaction. The resultant rc-TurboID backbone plasmids were first assessed by PCR spanning the junction site between joined DNA fragments, and then sent for Sanger sequencing for verification. To clone PSAF, FZL and RBMP2 genetic regions into the rc-TurboID backbone, plasmids were cloned via recombination according to Emrich-Mills et al., (2020). For genes that contain no Golden Gate enzyme restriction sites such as LCI9, their coding sequence was cloned into the L2-TurboID as described previously in (Lau et al., 2023). The L2-RPE1-TurboID plasmid was also reproduced from (Lau et al. 2023). Briefly, A complete list of primers used and the full plasmid sequence can be found in Supplemental Data Set S andSupplemental Data Set S.

### *Chlamydomonas* transformation

To generate the TurboID-expressing *Chlamydomonas* cells, BsaI was used to linearise the plasmids: L2-RPE1-TurboID, L2-LCI9-TurboID; I-SceI was used to linearise rc-PSAF-TurboID, rc-FZL-TurboID and rc-RBMP2-TurboID. *Chlamydomonas* cells grown in continuous light were harvested during mid-log phase (~$4\times10^6$ cells $mL^{-1}$) and resuspended in TAP medium supplemented with 40 mM sucrose (TAP-sucrose) to a final cell concentration of $1\times10^8$ cells $mL^{-1}$ determined using a Countess® II FL Automated Cell Counter (Thermofisher). 120 µL of the dense cell suspension was added to 2–mm electroporation cuvette with 200 ng linearised plasmid and mixed thoroughly. Using the NEPA21 electroporator (NEPA gene), square-wave electric pulses were delivered following the below setting: two poring pulse at 300 V with 8 ms pulse length, 50 ms pulse interval, and 40% decay rate. One transfer pulse at 20 V with 50 ms pulse length, 50 ms pulse interval (Yamano and Fukuzawa 2020). Next, electroporated cells were resuspended in 8 mL TAP-sucrose medium in a 15 mL falcon shaking under dark overnight to recover. Cells were then harvested by centrifugation ×1500 g, 4 minutes at room temperature and spread on TAP medium containing paromomycin antibiotic (20 µg $mL^{-1}$) for selection.

### Fluorescent protein tagging and confocal imaging

The FZL-mNeon plasmid (pLM744) was cloned via homologous recombination as described previously (Emrich-Mills et al. 2021), primers used are described in Supplemental Table 1. Strains expressing PSAF-Venus (pLM302), LCI9-mNeon (pLM381), and RBMP2-Venus

(pLM155) were reproduced from (Meyer et al. 2020; Emrich-Mills et al. 2021). Confocal imaging of the PSAF-Venus, RBMP2-Venus, LCI9-mNeon and FZL-mNeon were performed on a Zeiss LSM880 inverted microscope equipped with an Airyscan module. The excitation and emission filters of fluorophore and chlorophyll channels were set as follows: Venus/mNeon (excitation: 514 nm; emission: 520 to 550 nm); Chlorophyll (excitation: 633; emission: 610 to 650 nm). Image analysis was carried out using the FIJI software.

**Protein Extraction and immunoblotting**

Immunoblotting was used to assess the biotinylation activity as well as TurboID expression of the transformed cells. A colony of the TurboID-expressing cells was inoculated into 20 mL of liquid TAP-medium. The cell cultures were grown on a shaking platform until cells reached mid-log phase (2–4×$10^6$ cells $mL^{-1}$). Cell cultures were then harvested and resuspended in Tris-phosphate medium (TP) and into TP-supplemented with 2.5 mM Biotin substrate. The biotinylation reaction was allowed to proceed for 4 hours and the resultant cell culture were harvested by centrifugation (×17900 g, 4°C, 2 minutes) before snap-freezing in liquid nitrogen. 50 µL Modified RIPA lysis buffer (25 mM Tris–HCl pH 7.4, 300 mM NaCl, 1 mM DTT, 5 mM MgCl2, 0.1 mM PMSF, 1× EDTA-free protease inhibitor [Roche], 0.1% [w/v] SDS, 0.5% [w/v] deoxycholic acid, and 1% [v/v] Triton X-100) was added to resuspend the cell pellet. Lysis was performed using freeze-thaw method for 5 cycles using a warm water bath. Cell lysate was clarified by centrifugation (×17900 g, 4°C, 10 minutes). 6× protein loading dye was then added to the clarified lysate before denaturation for 10 minutes on heat block at 95°C. Denatured protein samples were resolved on Mini-Protean™ Tris-Glycine PAGE gels and transferred to PVDF membrane using a semi-dry transfer system (BioRad). Membrane was blocked with 3% (w/v) BSA in Tris-buffered saline with 0.1% (v/v) Tween 20 (TBST) and probed with antibodies accordingly. Antibodies were diluted in TBST as follows: Streptavidin Dylight-488 conjugate (1:4,000, Fisher Scientific #21832); anti-HA (1:1,000, Fisher Scientific 26183); and anti-tubulin (1:2,000, Sigma #T6074).

**Streptavidin-affinity purification**

Biotinylated proteins were enriched via streptavidin-affinity purification as previously mentioned (Lau et al. 2023) with minor modifications. Briefly, wildtype (cc-5325) cells and cells expressing RPE1-TurboID; PSAF-TurboID; RBMP2-TurboID, FZL-TurboID and LCI9-TurboID were inoculated into 400 mL TP-medium supplemented with 3% $CO_2$ supply on a Duran bottle with constant stirring at 160 rpm. Cells were allowed to grow until 0.5 – 1×$10^6$ cells $mL^{-1}$ and were switched to 0.04% $CO_2$ supply for at least 2 days to acclimate cells in a CCM inducing condition. After CCM induction, cells are harvested and resuspended in TP-medium to a cell concentration of $OD_{750}$ = 2.5. The suspensions were then split as triplicates

before 100 mM Biotin stock solution in DMSO was added to the cell culture to a final concentration of 2.5 mM. Biotin incubation step was allowed to proceed for 2 hours and halted by centrifugation 21,300 × g, 2 min at 4 °C. Biotin-labelled cells were rinsed 3 times with ice-cold TP medium before snap-frozen in liquid nitrogen until lysis. The cell lysis step was performed as mentioned above except protein loading dye was not added to the clarified cell lysate. Zeba Spin desalting columns (#89891, Thermo Fisher) were used to remove free biotin in lysate. The BCA protein assay (ThermoFisher) was used to measure protein concentrations after diluting protein extract in 1:10 dilution and 1.75 mg protein was added to the 50 µL Streptavidin beads that was pre-equilibrated with lysis buffer. Overnight incubation of beads with protein lysate was carried out in 4°C cold room on a rotator. Beads were then washed twice with lysis buffer for 5 min: once with 1 M KCl for 2 min; once with 0.1 M $NaCO_3$ for 1 min; once with 4 M urea in 50 mM triethylammonium bicarbonate, pH 8.5 (TEAB) for 1 min; once with 6 M urea in 50 mM TEAB for 1 min; and twice with 50 mM TEAB buffer for 5 min. The washed beads were frozen at −70 °C until submitted for mass spectrometry.

**Mass spectrometry analysis**

The 18 TurboID samples were analysed using mass spectrometry with a TimsTOF HT mass spectrometer (Bruker). On-Bead digestion of peptides was performed as previously described (Lau et al. 2023). EvoSep LC separation was performed using the 30 SPD pre-set elution method and a $C_{18}$ Performance column (8 cm x 150 μm x 1.5 μm). The nanoUPLC was interfaced to a timsTOF HT mass spectrometer (Bruker) with a CaptiveSpray ionisation source (Source). Positive PASEF-DIA (Parallel accumulation serial fragmentation – Data independent acquisition), spectra were acquired using Compass HyStar software (version 6.2, Thermo). Instrument source settings were: capillary voltage, 1,500 V; dry gas, 3 l/min; dry temperature; 180°C. The PASEF mass spectrometry method used a 1.8 second cycle time with $MS^2$ DIA windows of 50 *m/z* between *m/z* 400-1201 and a mobility range of 0.6-1.43 1/k0. The fragment scan *m/z* range was 100-1700. Resulting data were searched using DIA-NN (1.8.2.27) against the *Chlamydomonas* sequences from the *Chlamydomonas reinhardtii* genome cc-4532 v6.1 (JGI) appended with common proteomic contaminants and the TurboID protein. The search was performed at 1% false discovery rate with data filtered to require individual protein q-values as less than 0.01 for identification and a minimum of two peptides per accepted protein. DIA-quantification was performed within DIA-NN. In total there are 4610 accepted proteins groups. Statistical comparison between sample groups has been performed using an in-house installation of FragPipe-Analyst within R Shiny. Using FragPipe-Analyst pairwise comparisons were conducted between all groups with sample

minimum imputation applied. Non-adjusted p-values are used to plot volcano plots. The Complete analysis file can be found in Supplemental Data Set .

### Accession numbers

Sequence data from this article can be found in Phytozome, the Plant Comparative Genomics portal of the Department of Energy's Joint Genome Institute, under the following accession numbers: Cre02.g091750_4532; Cre02.g097250_4532; Cre03.g145787_4532: HSP22C; Cre03.g169300_4532; Cre03.g172700_4532; Cre03.g179961_4532; Cre03.g185250_4532: SSS2; Cre03.g199535_4532; Cre04.g215150_4532: SSS1B; Cre06.g259100_4532: SAGA3; Cre06.g261750_4532: RBMP1; Cre06.g270100_4532: SBE2; Cre08.g358563_4532; Cre08.g374100_4532; Cre08.g385500_4532: AMA1; Cre09.g394473_4532: LCI9; Cre09.g394510_4532; Cre09.g394547_4532; Cre09.g394621_4532: SAGA2; Cre09.g407501_4532; Cre09.g412100_4532: PSAF; Cre09.g413650_4532; Cre09.g416850_4532: RBMP2; Cre10.g430350_4532; Cre10.g435800_4532: CSP41B; Cre10.g436550_4532: EPYC1; Cre10.g440050_4532: CSP41A; Cre10.g444700_4532: SBE3; Cre12.g511900_4532: RPE1; Cre12.g530950_4532; Cre12.g801337_4532; Cre13.g570350_4532: ACK1; Cre16.g688302_4532; Cre17.g718950_4532: RCA2; Cre17.g719900_4532: PWD1; CreCp.g802313_4532: rbcL

## Acknowledgements

The authors would like to thank the University of York Biosciences Technology Facility for confocal microscopy access, advice on mass spectrometry experiments and the subsequent analysis.

## Authors contributions

L.C.M.M. guided and supervised the project; C.S.L. designed and performed the molecular cloning, TurboID expression screening and the streptavidin-affinity purification of the TurboID tagged lines as well as performed the confocal imaging of fluorescently tagged lines. L.A. performed the streptavidin-affinity purification of the LCI9-TurboID line. P.G. generated the plasmid used for the fluorescent protein tagging of FZL. A.D. performed LC-MS/MS analysis. C.S.L. analysed the data and wrote the manuscript.

## Supplemental Figures

Supplemental Figure S1 – Abundance and grouping of the submitted MS samples.

Supplemental Figure S2 – Localisation of the FZL protein tagged with mNeonGreen

Supplemental Data Set S1 – A list of all oligo sequences used in this work.

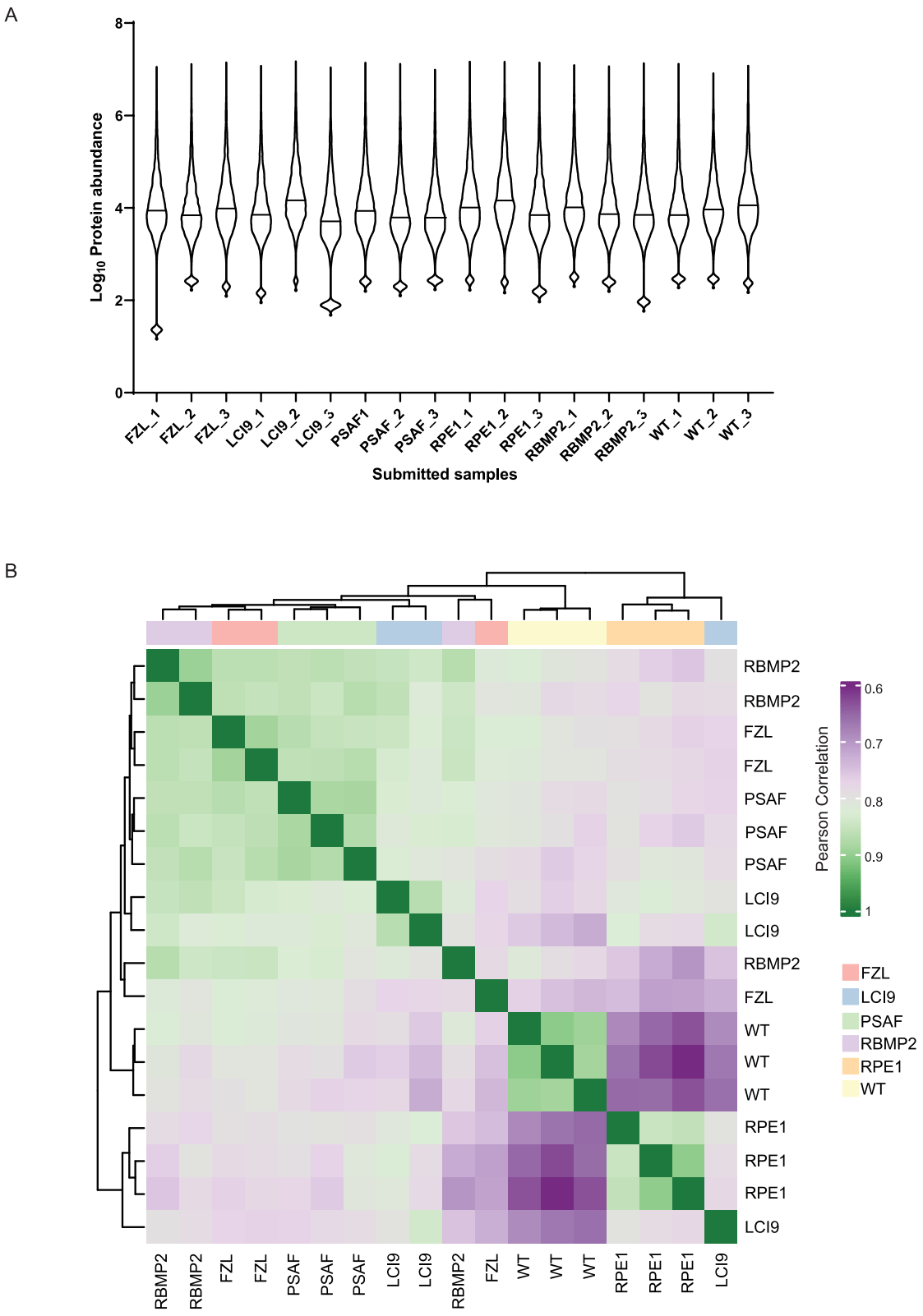
Supplemental Data Set S2 —- LCI9 proxiome.

Supplemental Data Set S3 – RBMP2 proxiome.

Supplemental Data Set S4 – FZL proxiome.

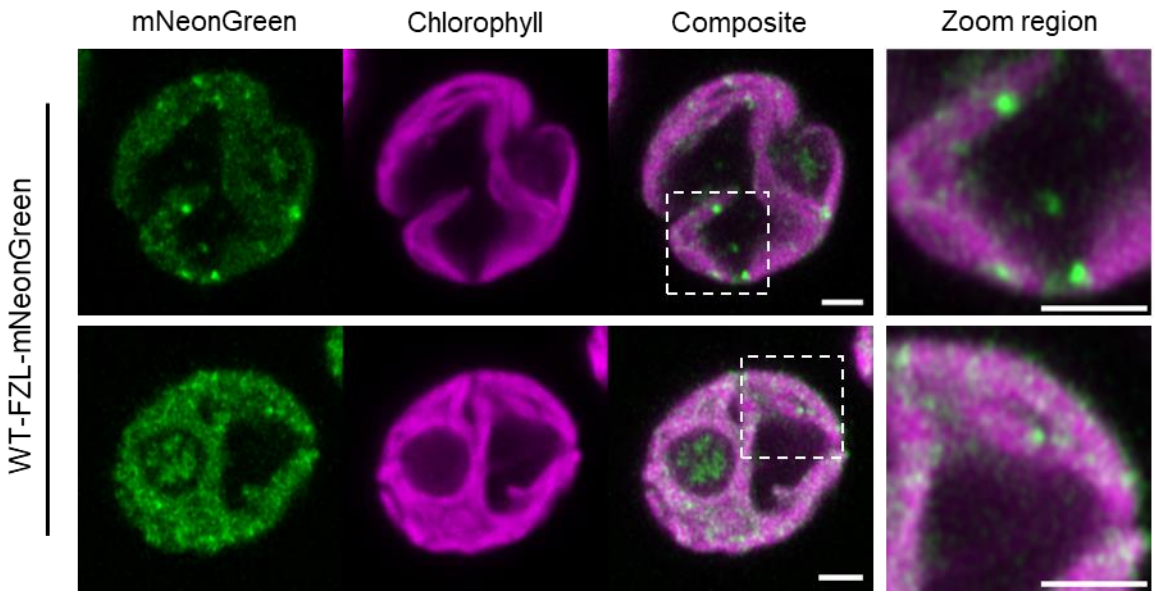
Supplemental Data Set S5 – The plasmid sequence of the rc-TurboID backbone.

Supplemental Data Set S6 – Mass spectrometry analysis of the submitted LCI9-TurboID, RBMP2-TurboID, PSAF-TurboID, FZL-TurboID and wildtype samples.




**Supplemental Figure S1.** Abundance and grouping of the submitted MS samples. (Supports Fig.3, 4 and 5). **A)** Total protein level across the three replicates of the submitted MS samples is shown as a

violin plot. The abundance of identified proteins are log10 transformed and a median line is shown as the horizontal line across each violin. **B)** Pearson correlations of all submitted samples are presented as a heatmap. The Pearson correlation coefficient is used to colour each comparison between submitted samples.




**Supplemental Figure S2.** Localisation of the FZL protein tagged with mNeonGreen. (Supports Figure 1) Recombineering plasmid pLM744 was transformed into Wildtype strains. Cells are cultivated in photoautotrophic conditions prior to imaging, the FZL signal is at both pyrenoid tubules and surrounding thylakoid membranes throughout the cell. Zoomed region highlights the bright puncta that are observed at the edges of thylakoid membrane, where membrane fusion event likely occurs. The green and magenta signals represent the mNeonGreen and Chlorophyll Channel, respectively. Scale bar = 2 µm.

**Supplemental Data Set S1.** A list of all oligo sequences used in this work.

| Oligo (oLM) | Sequence (in 5' – 3' direction) | Description |
|---|---|---|
| 427 | CGTTATCGGGGAAGAAGTGG | Forward primer for silent mutation on ccdB to remove BsaI site |
| 2843 | GATACCGGCACACTGGCC | Reverse primer for silent mutation on ccdB to remove BsaI site |
| 2840 | GAAGGTCTCAGCTTGGTACCGCTAGCCTTG | Forward primer to amplify Fragment 1 from the modified pLM099 |
| 2938 | GCAGGTCTCACATTGAAGATCCTTTGATCTTTTCTACG | Reverse primer to amplify Fragment 1 from the modified pLM099 |
| 2835 | TAAGGTCTCGAATGGTTAACTAAAAGCCAGATAACAG | Forward primer to amplify Fragment 2 from the modified pLM099 |
| 2837 | TAGGTCTCCTGCCAGATCCACCGCCGGAGCCACCCAG | Reverse primer to amplify Fragment 2 from the modified pLM099 |
| 2838 | CATGGTCTCTGGCAGCGGCAGCACCAG | Forward primer to amplify the TurboID-3×HA tag from L2-RBCS2-TurboID (pLM911) |
| 2839 | ATGGTCTCAAAGCAGCGGCATTGCACAGTCAC | Reverse primer to amplify the TurboID-3×HA tag from L2-RBCS2-TurboID (pLM911) |
| 3116 | GAAATTAATACGACTCACTATAGCAGAGGGGCCTGATACAAAGGTTTTAGAGCTAGAAATAGCAAG | Oligos containing T7 promoter and target sequence. Duplexed with oLM3118 to form the gRNA synthesis template. From (Yu et al. 2017). |
| 3118 | AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC | Oligos containing gRNA scaffold sequence. Duplexed with oLM3118 to form the gRNA synthesis template. From (Yu et al. 2017). |
| 3119 | CGTTGATCCTGCCAATAATGTT | Forward primer to screen CRISPR mutant |
| 3120 | TAGCTTCTTCTTAATGGTCGGC | Reverse primer to screen CRISPR mutant |

| | | |
|---|---|---|
| 1666 | GTTTGCGGGTTGTGACTGAAAC | Primers on the Hygromycin resistance cassette on the 5’ side amplifying outward |
| 2020 | GGAAGCGGACCGAGGACTTC | Primers on the Hygromycin resistance cassette on the 3’ side amplifying outward |

**Supplemental Data Set S2.** LCI9 proxiome. The LCI9 proxiome obtained by filtering all identified proteins which exhibits a $Log_2FC$ value >1 and p-value <0.05 when compared against the RPE1-TurboID and wildtype control. [A] Only the primary protein is listed, the complete protein group which contains multiple spliced forms or different proteins can be found in the supplemental dataset 6. [B] The protein abundance shown refers to the determined protein abundance in Arend et al., 2023 of the control strain cc-1690 (21gr).

| Protein Group | Gene name | Description | Protein abundance | $Log_2FC$ (LCI9 /RPE1) | $Log_2FC$ (LCI9 /WT) | p-value (LCI9 /RPE1) | p-value (LCI9 /WT) |
|---|---|---|---|---|---|---|---|
| Cre09.g394473_4532.1.p | LCI9 | Low-CO2-inducible protein | 0.086681458 | 6.58 | 8.36 | 4.53E+00 | 5.71E+00 |
| Cre02.g091750_4532.1.p | - | - | - | 5.63 | 5.33 | 3.62E+00 | 3.40E+00 |
| Cre12.g530950_4532.1.p | - | Putative triacylglycerol lipase | - | 3.78 | 3.87 | 1.38E+00 | 1.42E+00 |
| Cre08.g358563_4532.1.p | - | - | - | 3.39 | 1.91 | 4.08E+00 | 2.05E+00 |
| Cre09.g394510_4532.1.p | - | - | 0.015788402 | 3.39 | 4.47 | 5.04E+00 | 6.48E+00 |
| Cre03.g179961_4532.1.p | POLE1 | DNA polymerase epsilon subunit 1 | - | 3.26 | 4.76 | 1.49E+00 | 2.44E+00 |
| Cre01.g003850_4532.1.p | CYP747A1 | Cytochrome P450, CYP197 superfamily | - | 3.23 | 3.93 | 2.93E+00 | 3.68E+00 |
| Cre10.g444700_4532.1.p | SBE3 | Starch Branching Enzyme 3 | 0.14245015 | 3.18 | 6.71 | 4.21E+00 | 8.15E+00 |
| Cre13.g576600_4532.1.p | - | Conserved Protein with Similarity to Catalytic domain of Protein Tyrosine Kinases | - | 3.17 | 5.24 | 2.06E+00 | 3.78E+00 |
| Cre03.g185250_4532.1.p | SSS2 | Soluble starch synthase II | 0.202246973 | 3.08 | 3.85 | 4.28E+00 | 5.34E+00 |
| Cre03.g187150_4532.1.p | - | - | - | 2.97 | 3.56 | 5.11E+00 | 6.04E+00 |
| Cre09.g394621_4532.2.p | - | - | 0.024261679 | 2.96 | 5.29 | 2.82E+00 | 5.28E+00 |
| Cre09.g407501_4532.1.p | - | - | - | 2.91 | 4.44 | 2.06E+00 | 3.47E+00 |
| Cre02.g091750_4532.5.p | - | - | - | 2.84 | 4.3 | 5.04E+00 | 7.25E+00 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Cre06.g270100_4532.1.p | SBE2 | Starch Branching Enzyme 2 | 0.062653006 | 2.72 | 6.25 | 3.06E+00 | 7.02E+00 |
| Cre12.g504000_4532.1.p | - | - | - | 2.71 | 3.39 | 1.93E+00 | 2.57E+00 |
| Cre04.g215150_4532.3.p | SSS1B | Soluble starch synthase IB | - | 2.41 | 3.66 | 1.40E+00 | 2.43E+00 |
| Cre06.g271400_4532.1.p | - | Phosphoglycolate phosphatase/4-nitrophenylphosphatase | - | 2.39 | 2.09 | 1.60E+00 | 1.33E+00 |
| Cre09.g394547_4532.1.p | - | - | 0.063875396 | 2.06 | 2.36 | 2.24E+00 | 2.67E+00 |
| Cre10.g430350_4532.1.p | POB23 | Proteome of basal body 23 | 0.017525039 | 1.85 | 5.51 | 2.74E+00 | 7.98E+00 |
| Cre08.g385500_4532.1.p | AMA1 | Alpha-amylase 1 | 0.064976828 | 1.79 | 3.86 | 2.31E+00 | 5.43E+00 |
| Cre12.g485350_4532.1.p | - | - | - | 1.73 | 2.98 | 2.70E+00 | 4.93E+00 |
| Cre09.g400000_4532.1.p | - | - | 0.074088742 | 1.56 | 1.51 | 1.66E+00 | 1.59E+00 |
| Cre13.g571400_4532.1.p | - | - | - | 1.56 | 2.48 | 2.03E+00 | 3.55E+00 |
| Cre17.g720450_4532.1.p | - | - | 0.033171588 | 1.5 | 2.93 | 2.24E+00 | 4.82E+00 |
| Cre01.g049132_4532.1.p | - | - | 0.091031315 | 1.15 | 2.73 | 2.05E+00 | 5.44E+00 |
| Cre17.g719900_4532.1.p | PWD1 | Phosphoglucan water dikinase 1 | 0.068362828 | 1.11 | 3.57 | 1.40E+00 | 5.52E+00 |

**Supplemental Data Set S3.** RBMP2 proxiome. The RBMP2 proxiome obtained by filtering all identified proteins which exhibits a $Log_2FC$ value >1 and p-value <0.05 when compared against the PSAF-TurboID, RPE1-TurboID and wildtype control. [A] Only the primary protein is listed, the complete protein group which contains multiple spliced forms or different proteins can be found in the supplemental dataset 6. [B] The TMHMM (Transmembrane hidden Markov model) predictions are obtained from genome annotation files in *Chlamydomonas reinhardtii* V6.1. [C] The protein abundance shown refers to the determined protein abundance in Arend et al., 2023 of the control strain cc-1690 (21gr). [d] The RBM (Rubisco binding motif) prediction is obtained from Meyer et al., 2020.

| Protein Group[A] | Gene name | Description | TMHMM[b] | Protein abundance[c] | RBM[d] | $Log_2FC$ (RBMP2/PSAF) | $Log_2FC$ (RBMP2/WT) | $Log_2FC$ (RBMP2/RPE1) | p-value (RBMP2/PSAF) | p-value (RBMP2/WT) | p-value (RBMP2/RPE1) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cre12.g530950_4532.1.p | - | Putative triacylglycerol lipase | 0 | - | 0 | 6.04 | 4.17 | 4.61 | 2.53E+00 | 1.84E+00 | 1.79E+00 |
| Cre10.g430350_4532.1.p | POB23 | Proteome of basal body 23 | 0 | 0.017525 | 2 | 5.03 | 7.31 | 3.43 | 7.45E+00 | 9.48E+00 | 5.37E+00 |
| Cre09.g416850_4532.1.p | POB32 | Proteome of basal body 32 | 0 | 0.053612 | 7 | 4.54 | 4.70 | 5.04 | 4.29E+00 | 4.65E+00 | 4.77E+00 |
| Cre10.g436550_4532.1.p | EPYC1 | Essential pyrenoid component 1 | 0 | 0.830658 | 10 | 4.11 | 4.01 | 2.1 | 3.23E+00 | 3.33E+00 | 1.37E+00 |
| Cre10.g440050_4532.1.p | CSP41A | Chloroplast stem-loop-binding protein 41a | 0 | 0.126995 | 3 | 3.71 | 2.57 | 2.16 | 4.63E+00 | 3.33E+00 | 2.51E+00 |
| Cre17.g718950_4532.1.p | RCA2 | RuBisCO activase-like protein | 0 | 0.051943 | 1 | 3.41 | 3.76 | 1.83 | 4.74E+00 | 5.62E+00 | 2.32E+00 |
| Cre06.g259100_4532.1.p | POB30 | Proteome of basal body 30 | 0 | 0.06129 | 0 | 3.37 | 5.66 | 3.12 | 3.76E+00 | 6.35E+00 | 3.45E+00 |
| Cre10.g435800_4532.1.p | CSP41B | Chloroplast stem-loop-binding protein 41b | 0 | 0.210232 | 1 | 3.33 | 2.31 | 1.79 | 5.73E+00 | 4.35E+00 | 2.97E+00 |
| Cre16.g688302_4532.1.p | - | - | 0 | 0.052963 | 0 | 3.25 | 3.09 | 2.04 | 4.65E+00 | 4.49E+00 | 2.75E+00 |
| CreCp.g802313_4532.1.p | rbcL | RuBisCO large subunit | 0 | - | 0 | 3.05 | 1.73 | 1.98 | 5.29E+00 | 3.51E+00 | 3.36E+00 |
| Cre02.g143550_4532.2.p | ELIP3 | Early light-induced LHC-like protein | 0 | 0.214212 | 0 | 2.58 | 3.14 | 4.42 | 4.49E+00 | 5.79E+00 | 7.28E+00 |
| Cre03.g172700_4532.1.p | - | - | 0 | - | 4 | 2.32 | 3.50 | 3.01 | 2.52E+00 | 4.32E+00 | 3.45E+00 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cre01.g051000_4532.1.p | - | Protein-/Histone-arginine N-methyltransferase | 0 | 0.039185 | 1 | 2.23 | 1.59 | 1.52 | 2.19E+00 | 1.63E+00 | 1.32E+00 |
| Cre03.g169300_4532.1.p | - | - | 7 | - | 1 | 2.2 | 2.08 | 2.33 | 1.51E+00 | 1.35E+00 | 1.63E+00 |
| Cre06.g261750_4532.1.p | RBMP1 | RuBisCO binding membrane protein 1 | 0 | 0.079324 | 2 | 2.15 | 5.29 | 4.51 | 4.69E+00 | 9.77E+00 | 8.71E+00 |
| Cre02.g097250_4532.1.p | - | transportin-3 family protein | 0 | NA | 0 | 1.41 | 1.47 | 1.46 | 2.14E+00 | 2.72E+00 | 2.24E+00 |

**Supplemental Data Set S4.** FZL proxiome. The FZL proxiome obtained by filtering all identified proteins which exhibits a $Log_2FC$ value >1 and p-value <0.05 when compared against the PSAF-TurboID, RPE1-TurboID and wildtype control. [A] Only the primary protein is listed, the complete protein group which contains multiple spliced forms or different proteins can be found in the supplemental dataset 6. [B] The TMHMM (Transmembrane hidden Markov model) predictions are obtained from genome annotation files in *Chlamydomonas reinhardtii* V6.1. [C] The protein abundance shown refers to the determined protein abundance in Arend et al., 2023 of the control strain cc-1690 (21gr). [d] The RBM (Rubisco binding motif) prediction is obtained from Meyer et al., 2020.

| Protein Group[A] | Gene name | Description | TMHMM[b] | Protein abundance[c] | RBM[d] | $Log_2FC$ (FZL /PSAF) | $Log_2FC$ (FZL /RPE1) | $Log_2FC$ (FZL /WT) | p-value (FZL /PSAF) | p-value (FZL /RPE1) | p-value (FZL /WT) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cre14.g616600_4532.1.p | FZL1 | Plastid-localized | 0 | 0.033398 | 0 | 8.48 | 7.96 | 8.34 | 9.79E+00 | 9.40E+00 | 9.68E+00 |
| Cre08.g374100_4532.1.p | - | - | 0 | - | 0 | 8.38 | 3.79 | 7.49 | 6.27E+00 | 2.73E+00 | 5.67E+00 |
| Cre09.g413650_4532.1.p | - | - | 0 | NA | 1 | 4.49 | 4.52 | 4.12 | 1.79E+00 | 1.80E+00 | 1.60E+00 |
| Cre02.g119500_4532.1.p | - | - | 0 | - | 0 | 3.6 | 3.97 | 3.86 | 1.54E+00 | 1.75E+00 | 1.68E+00 |
| Cre06.g259100_4532.1.p | POB30 | Proteome of basal body 30 | 0 | 0.06129 | 0 | 2.78 | 2.53 | 5.15 | 3.02E+00 | 2.69E+00 | 5.77E+00 |
| Cre03.g199535_4532.1.p | - | Stress-enhanced protein 3 | 0 | 0.549445 | 0 | 2.41 | 1.48 | 2.23 | 3.43E+00 | 1.89E+00 | 3.13E+00 |
| Cre12.g801337_4532.1.p | - | - | 0 | - | 0 | 1.92 | 3.67 | 2.78 | 3.42E+00 | 6.49E+00 | 5.05E+00 |
| Cre06.g261750_4532.1.p | RBMP1 | RuBisCO binding membrane protein 1 | 0 | 0.079324 | 2 | 1.9 | 4.25 | 5.11 | 4.11E+00 | 8.36E+00 | 9.47E+00 |
| Cre12.g486209_4532.1.p | - | - | 0 | - | 3 | 1.72 | 2.12 | 1.4 | 2.97E+00 | 3.79E+00 | 2.32E+00 |
| Cre13.g570350_4532.1.p | AKC4 | conserved expressed ABC-1 like kinase | 1 | 0.042816 | 0 | 1.67 | 1.41 | 2.67 | 2.09E+00 | 1.66E+00 | 3.67E+00 |
| Cre03.g145787_4532.1.p | HSP22C | Heat shock protein 22C | 0 | NA | 0 | 1.61 | 1.94 | 2.04 | 2.59E+00 | 3.24E+00 | 3.43E+00 |
| Cre12.g525650_4532.1.p | SUFS1 | Group II cysteine desulfurase | 0 | 0.026647 | 0 | 1.39 | 1.4 | 1.38 | 1.46E+00 | 1.48E+00 | 1.45E+00 |

**Supplemental Data Set S5.** The plasmid sequence of the rc-TurboID backbone.

TGGTCTCTAATGGTTAACTAAAAGCCAGATAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAAT
ATATACTGATATGTATACCCGAAGTATGTCAAAAAGAGGTATGCTATGAAGCAGCGTATTACAGTGACAGTTG
ACAGCGACAGCTATCAGTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCA
GAATGAAGCCCGTCGTCTGCGTGCCGAACGCTGGAAAGCGGAAAATCAGGAAGGGATGGCTGAGGTCGCC
CGGTTTATTGAAATGAACGGCTCTTTTGCTGACGAGAACAGGGGCTGGTGAAATGCAGTTTAAGGTTTACAC
CTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTACAGAGTGATATTATTGACACGCCCGGGCGACG
GATGGTGATCCCCCTGGCCAGTGCACGTCTGCTGTCAGATAAAGTCCCCCGTGAACTTTACCCGGTGGTGC
ATATCGGGGATGAAAGCTGGCGCATGATGACCACCGATATGGCCAGTGTGCCGGTATCCGTTATCGGGGAA
GAAGTGGCTGATCTCAGCCACCGCGAAAATGACATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAA
ATGTCAGGCTCCCTTATACACAGCCAGTCTGCAGGTCGACCATAGTGTTAACGGAGATCTGGGTGGCTCCG
GCGGTGGATCTGGCAGCGGCAGCACCAGCGGCAGCGGCAGCACCGGCAAGGTGAGTCGACGAGCAAGCC
CGGCGGATCAGGCAGCGTGCTTGCAGATTTGACTTGCAACGCCCGCATTGTGTCGACGAAGGCTTTTGGCT
CCTCTGTCGCTGTCTCAAGCAGCATCTAACCCTGCGTCGCCGTTTCCATTTGCAGGACAACACCGTGCCCCT
GAAGCTGATCGCCCTGCTGGCCAACGGCGAGTTCCACAGCGGCGAGCAGCTGGGCGAGACGCTGGGCAT
GAGCCGCGCCGCCATCAACAAGCACATCCAGACCCTGCGCGACTGGGGCGTGGACGTGTTCACCGTGCCC
GGCAAGGGCTACAGCCTGCCCGAGCCCATCCCCCTGCTGAACGCCAAGCAGATCCTGGGCCAGCTGGACG
GCGGCAGCGTGGCCGTGCTGCCCGTGGTGGACAGCACCAACCAGTACCTGCTGGACCGCATCGGCGAGC
TGAAGAGCGGCGACGCCTGCATCGCCGAGTACCAGCAGGCCGGCCGCGGCAGCCGCGGCCGCAAGTGGT
TCAGCCCCTTCGGCGCCAACCTGTACCTGAGCATGTTCTGGCGCCTGAAACGTGGCCCCGCCGCTATTGGC
CTGGGCCCCGTGATCGGCATCGTGATGGCCGAGGCCCTGCGCAAGCTGGTAAGTCTTTCTGTGTCGCGGG
GTTCTGGGCGTTCGCATGCGCAACAGTGTCGCACGGTCGCTCTTGCAGCACAGTCACTACAGATAGTCCAA
GTCCGACGCATGGCGATCGGGCAACTGCGATTTGCACATGCGGCAAGGGATCTCTAGCTCGGGCTGGCGA
AGCCTTCAGGACATGGAGCGCTGTCCAGCAGCTGGTTGGTGATGCTCTATCCTAAATTGCCCCTCCCACAC
ACCCTTACTTGCTTTCCAGGGCGCCGACAAGGTGCGCGTGAAGTGGCCCAACGACCTGTACCTGCAGGAC
CGCAAGCTGGCCGGCATCCTGGTGGAGCTGGCCGGCATCACCGGCGACGCCGCCCAGATCGTGATCGGC
GCCGGCATCAACGTGGCCATGCGCCGCGTGGAGGAGAGCGTGGTGAACCAGGGCTGGATCACCCTGCAG
GAGGCCGGCATCAACCTGGACCGCAACACCCTGGCCGCCACCCTGATCCGCGAGCTGCGCGCCGCCCTG
GAGCTGTTCGAGCAGGAGGGCCTGGCCCCCTACCTGCCCCGCTGGGAGAAGCTGGACAACTTCATCAACC
GCCCCGTGAAGCTGATCATCGGCGACAAGGAGATCTTCGGCATCAGCCGCGGCATCGACAAGCAGGGCGC
CCTGCTGCTGGAGCAGGACGGCGTGATCAAGCCCTGGATGGGCGGCGAGATCAGCCTGCGCAGCGCCGA
GAAGGGTTCGTCTGGTTACCCCTACGACGTGCCCGACTACGCTGGCTCCTACCCGTACGACGTCCCCGACT
ACGCTGGCAGCTACCCTTACGACGTGCCGGACTACGCTTAAGCTTAGCAGCTGGACCGCCTGTACCATGGA
GAAGAGCTTTACTTGCCGGGATGGCCGATTTCGCTGATTGATACGGGATCGGAGCTCGGAGGCTTTCGCGC
TAGGGGCTAGGCGAAGGGCAGTGGTGACCAGGGTCGGTGTGGGGTCGGCCCACGGTCAATTAGCCACAG
GAGGATCAGGGGGAGGTAGGCACGTCGACTTGGTTTGCGACCCCGCAGTTTTGGCGGACGTGCTGTTGTA
GATGTTAGCGTGTGCGTGAGCCAGTGGCCAACGTGCCACACCCATTGAGATGACCAACCAACTTACTGGCA
ATATCTGCCAATGCCATACTGCATGTAATGGCCAGGCCATGTGAGAGTTTGCCGTGCCTGCGCGCGCCCCG
GGGGCGCAGTTTAGCTGACCAGCCGTGGGATGATGCACGCATTTGCAAGGACAGGGTAATCACAGCAGCA
ACATGGTGGGCTTAGGACAGCTGTGGGTCAGTGGACGGACGGCAGGGGAGGGACGGCGCAGCTCGGGAG
ACAGGGGGAGACAGCGTGACTGTGCAATGCCGCTGCTTGGTACCGCTAGCCTTGACATGATTGGTGCGTAT
GTTTGTATGAAGCTACAGGACTGATTTGGCGGGCTATGAGGGCGGGGGAAGCTCTGGAAGGGCCGCGATG
GGGCGCGCGGCGTCCAGAAGGCGCCATACGGCCCGCTGGCGGCACCCATCCGGTATAAAAGCCCGCGAC
CCCGAACGGTGACCTCCACTTTCAGCGACAAACGAGCACTTATACATACGCGACTATTCTGCCGCTATACAT

AACCACTCAGCTAGCTTAAGATCCCATCAAGCTTGCATGCCGGGCGCGCCAGAAGGAGCGCAGCCAAACCA
GGATGATGTTTGATGGGGTATTTGAGCACTTGCAACCCTTATCCGGAAGCCCCCTGGCCCACAAAGGCTAG
GCGCCAATGCAAGCAGTTCGCATGCAGCCCCTGGAGCGGTGCCCTCCTGATAAACCGGCCAGGGGGCCTA
TGTTCTTTACTTTTTTACAAGAGAAGTCACTCAACATCTTAAAATGGCCAGGTGAGTCGACGAGCAAGCCCG
GCGGATCAGGCAGCGTGCTTGCAGATTTGACTTGCAACGCCCGCATTGTGTCGACGAAGGCTTTTGGCTCC
TCTGTCGCTGTCTCAAGCAGCATCTAACCCTGCGTCGCCGTTTCCATTTGCAGGATGGCCACTCCGCCCTC
CCCGGTGCTGAAGAATTTCGAAGCATGGACGATGCGTTGCGTGCACTGCGGGGTCGGTATCCCGGTTGTG
AGTGGGTTGTTGTGGAGGATGGGGCCTCGGGGGCTGGTGTTTATCGGCTTCGGGGTGGTGGGCGGGAGTT
GTTTGTCAAGGTGGCAGCTCTGGGGGCCGGGGTGGGCTTGTTGGGTGAGGCTGAGCGGCTGGTGTGGTT
GGCGGAGGTGGGGATTCCCGTACCTCGTGTTGTGGAGGGTGGTGGGGACGAGAGGGTCGCCTGGTTGGT
CACCGAAGCGGTTCCGGGGCGTCCGGCCAGTGCGCGGTGGCCGCGGGAGCAGCGGCTGGACGTGGCGG
TGGCGCTCGCGGGGCTCGCTCGTTCGCTGCACGCGCTGGACTGGGAGCGGTGTCCGTTCGATCGCAGTCT
CGCGGTGACGGTGCCGCAGGCGGCCCGTGCTGTCGCTGAAGGGAGCGTCGACTTGGAGGATCTGGACGA
GGAGCGGAAGGGGTGGTCGGGGGAGCGGCTTCTCGCCGAGCTGGAGCGGACTCGGCCTGCGGACGAGG
ATCTGGCGGTTTGCCACGGTGACCTGTGCCCGGACAACGTGCTGCTCGACCCTCGTACCTGCGAGGTGAC
CGGGCTGATCGACGTGGGGCGGGTCGGCCGTGCGGACCGGCACTCCGATCTCGCGCTGGTGCTGCGCGA
GCTGGCCCACGAGGAGGACCCGTGGTTCGGGCCGGAGTGTTCCGCGGCGTTCCTGCGGGAGTACGGGCG
CGGGTGGGATGGGGCGGTATCGGAGGAAAAGCTGGCGTTTTACCGGCTGTTGGACGAGTTCTTCTGAGGG
ACCTGATGGTGTTGGTGGCTGGGTAGGGTTGCGTCGCGTGGGTGACAGCACAGTGTGGACGTTGGGATCC
CCGCTCCGTGTAAATGGAGGCGCTCGTTGATCTGAGCCTTGCCCCCTGACGAACGGCGGTGGATGGAAGA
TACTGCTCTCAAGTGCTGAAGCGGTAGCTTAGCTCCCCGTTTCGTGCTGATCAGTCTTTTTCAACACGTAAAA
AGCGGAGGAGTTTTGCAATTTTGTTGGTTGTAACGATCCTCCGTTGATTTTGGCCTCTTTCTCCATGGGCGG
GCTGGGCGTATTTGAAGCGTAGGGATAACAGGGTAATTGCACTGAAATCTAGAAATATTTTATCTGATTAATA
AGATGATCTTCTTGAGATCGTTTTGGTCTGCGCGTAATCTCTTGCTCTGAAAACGAAAAAACCGCCTTGCAG
GGCGGTTTTTCGAAGGTTCTCTGAGCTACCAACTCTTTGAACCGAGGTAACTGGCTTGGAGGAGCGCAGTC
ACCAAAACTTGTCCTTTCAGTTTAGCCTTAACCGGCGCATGACTTCAAGACTAACTCCTCTAAATCAATTACC
AGTGGCTGCTGCCAGTGGTGCTTTTGCATGTCTTTCCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG
CGCAGCGGTCGGACTGAACGGGGGGTTCGTGCATACAGTCCAGCTTGGAGCGAACTGCCTACCCGGAACT
GAGTGTCAGGCGTGGAATGAGACAAACGCGGCCATAACAGCGGAATGACACCGGTAAACCGAAAGGCAGG
AACAGGAGAGCGCACGAGGGAGCCGCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGC
CACCACTGATTTGAGCGTCAGATTTCGTGATGCTTGTCAGGGGGGCGGAGCCTATGGAAAAACCAGGTGGC
ACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCAT
GAATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCA
TATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCC
TGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGT
TATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCA
GACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGT
GATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAA
CCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAA
TGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGT
CGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACC
TTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTG
CCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGA
GCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATT

GTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGG
ATCTTCGGAGTGAGACCT

**Supplemental Data Set S6.** Mass spectrometry analysis of the submitted LCI9-TurboID, RBMP2-TurboID, PSAF-TurboID, FZL-TurboID and wildtype samples. **(The dataset is submitted as a separate file due to their large size)**

## References


**Abreu IN, Aksmann A, Bajhaiya AK, Benlloch R, Giordano M, Pokora W, Selstam E, and Moritz T**. Changes in lipid and carotenoid metabolism in Chlamydomonas reinhardtii during induction of CO2-concentrating mechanism: Cellular response to low CO2 stress. Algal Res. 2020:**52**(102099):102099.

**Arasaki K, Takagi D, Furuno A, Sohda M, Misumi Y, Wakana Y, Inoue H, and Tagaya M**. A new role for RINT-1 in SNARE complex assembly at the*trans*-Golgi network in coordination with the COG complex. Mol Biol Cell. 2013:**24**(18):2907–2917.

**Arend M, Zimmer D, Xu R, Sommer F, Mühlhaus T, and Nikoloski Z**. Proteomics and constraint-based modelling reveal enzyme kinetic properties of Chlamydomonas reinhardtii on a genome scale. Nat Commun. 2023:**14**(1):4781.

**Armbruster U, Labs M, Pribil M, Viola S, Xu W, Scharfenberg M, Hertle AP, Rojahn U, Jensen PE, Rappaport F, et al.** Arabidopsis CURVATURE THYLAKOID1 proteins modify thylakoid architecture by inducing membrane curvature. Plant Cell. 2013:**25**(7):2661–2678.

**Austin JR II, Frost E, Vidi P-A, Kessler F, and Staehelin LA**. Plastoglobules are lipoprotein subcompartments of the chloroplast that are permanently coupled to thylakoid membranes and contain biosynthetic enzymes. Plant Cell. 2006:**18**(7):1693–1703.

**Bottomley MJ, Collard MW, Huggenvik JI, Liu Z, Gibson TJ, and Sattler M**. The SAND domain structure defines a novel DNA-binding fold in transcriptional regulation. Nat Struct Biol. 2001:**8**(7):626–633.

**Broderson M, Niyogi KK, and Iwai M**. Macroscale structural changes of thylakoid architecture during high light acclimation in Chlamydomonas reinhardtii. Photosynth Res. 2024. https://doi.org/10.1007/s11120-023-01067-1

**Courseaux A, George O, Deschamps P, Bompard C, Duchêne T, and Dauvillée D**. BE3 is the major branching enzyme isoform required for amylopectin synthesis in Chlamydomonas reinhardtii. Front Plant Sci. 2023:**14**. https://doi.org/10.3389/fpls.2023.1201386

**Dodge JD**. The fine structure of chloroplasts and pyrenoids in some marine dinoflagellates. J Cell Sci. 1968:**3**(1):41–47.

**Doerr A**. DIA mass spectrometry. Nat Methods. 2015:**12**(1):35–35.

**Eitzinger N, Wagner V, Weisheit W, Geimer S, Boness D, Kreimer G, and Mittag M**. Proteomic analysis of a fraction with intact eyespots of Chlamydomonas reinhardtii and assignment of protein methylation. Front Plant Sci. 2015:**6**. https://doi.org/10.3389/fpls.2015.01085

**Emrich-Mills TZ, Yates G, Barrett J, Girr P, Grouneva I, Lau CS, Walker CE, Kwok TK, Davey JW, Johnson MP, et al.** A recombineering pipeline to clone large and complex genes in Chlamydomonas. Plant Cell. 2021:**33**(4):1161–1181.

**Engel BD, Schaffer M, Kuhn Cuellar L, Villa E, Plitzko JM, and Baumeister W**. Native architecture of the Chlamydomonas chloroplast revealed by in situ cryo-electron tomography. Elife. 2015:**4**. https://doi.org/10.7554/elife.04889

**Farah J, Rappaport F, Choquet Y, Joliot P, and Rochaix JD**. Isolation of a psaF-deficient mutant of Chlamydomonas reinhardtii: efficient interaction of plastocyanin with the photosystem I reaction center is mediated by the PsaF subunit. EMBO J. 1995:**14**(20):4976–4984.

**Fei C, Wilson AT, Mangan NM, Wingreen NS, and Jonikas MC**. Modelling the pyrenoid-based CO2-concentrating mechanism provides insights into its operating principles and a roadmap for its engineering into crops. Nat Plants. 2022:**8**(5):583–595.

**Findinier J, Delevoye C, and Cohen MM**. The dynamin-like protein Fzl promotes thylakoid fusion and resistance to light stress in Chlamydomonas reinhardtii. PLoS Genet. 2019a:**15**(3):e1008047.

**Findinier J, Laurent S, Duchêne T, Roussel X, Lancelon-Pin C, Cuiné S, Putaux J-L, Li-Beisson Y, D'Hulst C, Wattebled F, et al.** Deletion of BSG1 in Chlamydomonas reinhardtii leads to abnormal starch granule size and morphology. Sci Rep. 2019b:**9**(1):1–13.

**Flori S, Jouneau P-H, Finazzi G, Maréchal E, and Falconet D**. Ultrastructure of the Periplastidial Compartment of the Diatom Phaeodactylum tricornutum. Protist. 2016:**167**(3):254–267.

**Gao H, Sage TL, and Osteryoung KW**. FZL, an FZO-like protein in plants, is a determinant of thylakoid and chloroplast morphology. Proc Natl Acad Sci U S A. 2006:**103**(17):6759–6764.

**He S, Chou H-T, Matthies D, Wunder T, Meyer MT, Atkinson N, Martinez-Sanchez A, Jeffrey PD, Port SA, Patena W, et al.** The structural basis of Rubisco phase separation in the pyrenoid. Nat Plants. 2020:**6**(12):1480–1490.

**Kerr ID, Hutchison E, Gerard L, Aleidi SM, and Gelissen IC**. Mammalian ABCG-transporters, sterols and lipids: To bind perchance to transport? Biochim Biophys Acta Mol Cell Biol Lipids. 2021:**1866**(3):158860.

**Kreis E, König K, Misir M, Niemeyer J, Sommer F, and Schroda M**. TurboID reveals the proxiomes of *Chlamydomonas* proteins involved in thylakoid biogenesis and stress response. Plant Physiol. 2023:**193**(3):1772–1796.

**Lau CS, Dowle A, Thomas GH, Girr P, and Mackinder LCM**. A phase-separated CO2-fixing pyrenoid proteome determined by TurboID in *Chlamydomonas reinhardtii*. Plant Cell. 2023:**35**(9):3260–3279.

**LeKieffre C, Spero HJ, Russell AD, Fehrenbacher JS, Geslin E, and Meibom A**. Assimilation, translocation, and utilization of carbon between photosynthetic symbiotic dinoflagellates and their planktic foraminifera host. Mar Biol. 2018:**165**(6). https://doi.org/10.1007/s00227-018-3362-7

**Mackinder LCM, Chen C, Leib RD, Patena W, Blum SR, Rodman M, Ramundo S, Adams CM, and Jonikas MC**. A Spatial Interactome Reveals the Protein Organization of the Algal CO2-Concentrating Mechanism. Cell. 2017:**171**(1):133-147.e14.

**Markina ZV, Orlova TY, Vasyanovich YA, Vardavas AI, Stivaktakis PD, Vardavas CI, Kokkinakis MN, Rezaee R, Ozcagli E, and Golokhvast KS**. Porphyridium purpureum microalga physiological and ultrastructural changes under copper intoxication. Toxicol Rep. 2021:**8**:988–993.

**Meyer MT, Genkov T, Skepper JN, Jouhet J, Mitchell MC, Spreitzer RJ, and Griffiths H**. Rubisco small-subunit α-helices control pyrenoid formation in *Chlamydomonas*. Proc Natl Acad Sci U S A. 2012:**109**(47):19474–19479.

**Meyer MT, Itakura AK, Patena W, Wang L, He S, Emrich-Mills T, Lau CS, Yates G, Mackinder LCM, and Jonikas MC**. Assembly of the algal $CO_2$-fixing organelle, the pyrenoid, is guided by a Rubisco-binding motif. Sci Adv. 2020:**6**(46). https://doi.org/10.1126/sciadv.abd2408

**Minde D-P, Ramakrishna M, and Lilley KS**. Biotin proximity tagging favours unfolded proteins and enables the study of intrinsically disordered regions. Commun Biol. 2020:**3**(1):1–13.

**Mitrousis G, Olia AS, Walker-Kopp N, and Cingolani G**. Molecular basis for the recognition of snurportin 1 by importin β. J Biol Chem. 2008:**283**(12):7877–7884.

**Oka M and Yoneda Y**. Importin α: functions as a nuclear transport factor and beyond. Proc Jpn Acad Ser B Phys Biol Sci. 2018:**94**(7):259–274.

**Pérez-Amador MA, Abler ML, De Rocher EJ, Thompson DM, van Hoof A, LeBrasseur ND, Lers A, and Green PJ**. Identification of BFN1, a bifunctional nuclease induced during leaf and stem senescence in Arabidopsis. Plant Physiol. 2000:**122**(1):169–180.

**Prelich G**. Gene Overexpression: Uses, mechanisms, and interpretation. Genetics. 2012:**190**(3):841–854.

**Rottet S, Besagni C, and Kessler F**. The role of plastoglobules in thylakoid lipid remodeling during plant development. Biochim Biophys Acta Bioenerg. 2015:**1847**(9):889–899.

**Seung D, Schreier TB, Bürgy L, Eicke S, and Zeeman SC**. Two plastidial coiled-coil proteins are essential for normal starch granule initiation in Arabidopsis. Plant Cell. 2018:**30**(7):1523–1542.

**Shimakawa G, Okuyama A, Harada H, Nakagaito S, Toyoshima Y, Nagata K, and Matsuda Y**. Pyrenoid-core CO2-evolving machinery is essential for diatom photosynthesis in elevated CO2. Plant Physiol. 2023. https://doi.org/10.1093/plphys/kiad475

**Spraggon G, Pantazatos D, Klock HE, Wilson IA, Woods VL Jr, and Lesley SA**. On the use of DXMS to produce more crystallizable proteins: Structures of the *T. maritima* proteins TM0160 and TM1171. Protein Sci. 2004:**13**(12):3187–3199.

**Su X, Ma J, Pan X, Zhao X, Chang W, Liu Z, Zhang X, and Li M**. Antenna arrangement and energy transfer pathways of a green algal photosystem-I–LHCI supercomplex. Nat Plants. 2019:**5**(3):273–281.

**Taylor NMI, Manolaridis I, Jackson SM, Kowal J, Stahlberg H, and Locher KP**. Structure of the human multidrug transporter ABCG2. Nature. 2017:**546**(7659):504–509.

**Wang L, Patena W, Van Baalen KA, Xie Y, Singer ER, Gavrilenko S, Warren-Williams M, Han L, Harrigan HR, Hartz LD, et al.** A chloroplast protein atlas reveals punctate structures and spatial organization of biosynthetic pathways. Cell. 2023:**186**(16):3499-3518.e14.

**Wang L, Yamano T, Takane S, Niikawa Y, Toyokawa C, Ozawa S-I, Tokutsu R, Takahashi Y, Minagawa J, Kanesaki Y, et al.** Chloroplast-mediated regulation of $CO_2$-concentrating mechanism by $Ca^{2+}$-binding protein CAS in the green alga *Chlamydomonas reinhardtii*. Proc Natl Acad Sci U S A. 2016:**113**(44):12586–12591.

**Wang Y and Spalding MH**. Acclimation to very low $CO_2$: Contribution of limiting $CO_2$ inducible proteins, LCIB and LCIA, to inorganic carbon uptake in *Chlamydomonas reinhardtii*. Plant Physiol. 2014:**166**(4):2040–2050.

**Wijeyesakere SJ, Richardson RJ, and Stuckey JA**. Crystal structure of patatin-17 in complex with aged and non-aged organophosphorus compounds. PLoS One. 2014:**9**(9):e108245.

**Yamano T and Fukuzawa H**. Transformation of the Model Microalga Chlamydomonas reinhardtii Without Cell-Wall Removal. Methods Mol Biol. 2020:**2050**:155–161.

**Yamano T, Sato E, Iguchi H, Fukuda Y, and Fukuzawa H**. Characterization of cooperative bicarbonate uptake into chloroplast stroma in the green alga *Chlamydomonas reinhardtii*. Proc Natl Acad Sci U S A. 2015:**112**(23):7315–7320.

**Yu J, Baek K, Jin E, and Bae S**. DNA-free genome editing of Chlamydomonas reinhardtii using CRISPR and subsequent mutant analysis. Bio Protoc. 2017:**7**(11). https://doi.org/10.21769/bioprotoc.2352

**Zhan Y, Dhaliwal JS, Adjibade P, Uniacke J, Mazroui R, and Zerges W**. Localized control of oxidized RNA. J Cell Sci. 2015:**128**(22):4210–4219.

**Zhan Y, Marchand CH, Maes A, Mauries A, Sun Y, Dhaliwal JS, Uniacke J, Arragain S, Jiang H, Gold ND, et al.** Pyrenoid functions revealed by proteomics in Chlamydomonas reinhardtii. PLoS One. 2018:**13**(2):e0185039.

**Zhang N, Mattoon EM, McHargue W, Venn B, Zimmer D, Pecani K, Jeong J, Anderson CM, Chen C, Berry JC, et al.** Systems-wide analysis revealed shared and unique responses to moderate and acute high temperatures in the green alga Chlamydomonas reinhardtii. Commun Biol. 2022:**5**(1). https://doi.org/10.1038/s42003-022-03359-z

# Chapter 4 – PINS2 is important for pyrenoid starch sheath formation

## Introduction to the chapter

The $CO_2$ delivery mechanism in *Chlamydomonas* requires the coordination of different components in the pyrenoid organelle. After the Ci-delivery mechanism delivers $HCO_3^-$ into the thylakoid lumen, a carbonic anhydrase CAH3 is responsible for the conversion of $HCO_3^-$ into $CO_2$ for release into the pyrenoid matrix. At this last step, a large amount of $CO_2$ is released to saturate Rubisco, which allows the cell to suppress photorespiration and achieve an increase in photosynthetic output. However, this large release of $CO_2$ similarly results in its leakage from the pyrenoid matrix that, if not controlled, will lead to futile cycling of the $CO_2$ and crippling of the $CO_2$ concentration mechanism.

The *Chlamydomonas* starch sheath that surrounds the pyrenoid matrix has long been proposed as a diffusion barrier, which limits the $CO_2$ leakage from pyrenoid matrix. The leakage barrier function is also proposed to be aided by a putative carbonic anhydrase, LCIB, which actively re-locates to the pyrenoid periphery upon experiencing $C_i$-limiting conditions (Yamano et al. 2022). Despite its importance, not much is known about the formation of the starch sheath. This leads to the research that began in chapters 2 and 3, which aimed to provide a strong foundation for us to identify pyrenoid proteins that potentially participate in the starch sheath synthesis process. During the writing process of Chapter 2, we had cultivated a particular interest in the PINS2 protein due to its similarity to another protein, LCI9, through bioinformatics analysis. We quickly discovered that Cre09.g394473 (PINS1), Cre09.g394510 (PINS2) and Cre09.g394547 (PINS3) are likely homologs, all sharing a similar domain arrangement.

Previously, a role for the pyrenoid starch sheath formation was suggested for PINS1 (or LCI9) due to its pyrenoid periphery localisation and association with canonical starch-related proteins such as SBE3/PFK1/PFK2 (Mackinder et al. 2017). In Chapter 2, we had similarly localised PINS2 to this pyrenoid starch periphery localisation. PINS3 was cloned and localised to the same starch sheath gap location by an alumni postdoctoral research associate, Dr. Gary Yates (data not shown). However, multiple attempts to transform the PINS3-Venus plasmid into *Chlamydomonas* has failed to generate fluorescent colonies and was therefore not discussed in the present chapter. Previously, the identification of a SNARE-domain annotation at the C-terminal of PINS2 drove us to localise this protein in Chapter 2, as their placement at the pyrenoid starch sheath periphery also marked the entrance of the pyrenoid tubules, which is a unique membrane structure formed by the conversion of regularly sheet-like thylakoids into tubule-like structures. The SNARE-domain annotation therefore placed the PINS protein as a prime target which could drive this sheet-

to-tubule membrane change via a SNARE based mechanism. However, we later find that we could not detect a SNARE motif on the PINS2 sequence using multiple motif searching webserver (MOTIF and InterPro; (Kanehisa 1997; Paysan-Lafosse et al. 2023)) suggesting the t-SNARE annotation might possibly be a mis-annotation. Nonetheless, the coiled-coil domain identified on the PINS proteins still sparked our interest as they parallel the PTST proteins which directly influence starch granule initiation (Seung et al. 2015, 2017) and thus could potentially reshape the stromal starch granules into the discoidal pyrenoid starch sheath through their interaction with starch. To explore this possibility, we decided to focus on characterising the PINS2 gene by generating multiple CRISPR-knock out lines assessing their starch and thylakoid ultrastructure.

### Declaration of authorship

All experimental work presented below was carried out by Chun Sing Lau with the following exception:

Liat Adler grew the cells for the transmission electron microscopy experiment in Figure 5

Seb grant generated two of the lines: WT-LCIB-Venus and WT-LCIB-mCherry used in Figure 6 and performed the confocal imaging experiment in Figure 6.

All figures are assembled by Chun Sing Lau.

## The starch sheath peripheral coiled-coil protein PINS2 is important for pyrenoid starch sheath formation

### Authors

Chun Sing Lau[1], Seb Grant[1], Liat Adler[2], Luke C M Mackinder[1]

[1]The University of York, Centre of Novel Agricultural Products, Department of Biology, United Kingdom
[2]Carnegie Institute of Science, Biosphere Sciences and Engineering, United States

### Abstract

Starch synthesis is a tightly controlled process essential for coordinating cellular energy reserves in response to the dynamic environments that plants and algae face every day. In *Chlamydomonas reinhardtii*, a unique form of transitory starch reserve, called pyrenoid starch, appears in parallel with the induction of the cell's $CO_2$ concentration mechanism (CCM). Unlike the lenticular starch granules commonly seen in plant leaves, pyrenoid starch exhibits a cup-shaped morphology and encircles the pyrenoid. This unique morphology not only delimits the site of pyrenoid tubule entry into the pyrenoid matrix but also functions as a diffusion barrier, lowering $CO_2$ leakage and improving CCM efficiency *in vivo*. Despite their

apparent importance, the starch sheath assembly process is relatively unclear. Here, we identified a starch-related protein, PINS2, which was previously localised to the starch sheath periphery. PINS2 belongs to a three-gene cluster, all sharing the same domain architecture of a starch-binding domain and a coiled-coil domain, which shows similarity to the Arabidopsis PTST protein critical for granule initiation. Through CRISPR-mediated mutant generation, we found that the *pins2* mutant showed an aberrant starch sheath that is thinned and contains protrusions which overlap each other. Additionally, we discovered that the putative carbonic anhydrase LCIB fails to re-localise to the pyrenoid periphery in a similar fashion to that observed in starch-less *Chlamydomonas* mutants. We hypothesized that the PINS protein forms a multimeric assembly at the pyrenoid starch sheath periphery, where their starch-binding domain restricts starch synthesis locally to maintain the starch sheath gap. The characterization of the *pins2* mutant represents a step towards understanding the enigmatic starch sheath assembly process in *Chlamydomonas reinhardtii*.

## Introduction

Found in nearly all eukaryotic algal lineages (Badger et al. 1998; Barrett et al. 2021), the pyrenoid is a structurally diverse organelle consisting of highly condensed Rubisco (Ribulose-1,5-biphosphate Carboxylase/Oxygenase) within the chloroplast. Pyrenoids are responsible for ~30% of the global carbon cycling marking its striking geochemical importance (Field et al. 1998; Mackinder et al. 2016). Pyrenoid organisation and composition are best characterised in the green algal model species *Chlamydomonas reinhardtii (*Chlamydomonas hereafter). The Chlamydomonas pyrenoid consists of a Rubisco-rich pyrenoid matrix, a starch sheath which encircles the pyrenoid matrix; as well as matrix-traversing pyrenoid tubules stemming from the neighbouring thylakoid membranes. The *Chlamydomonas* pyrenoid matrix condenses most of its Rubisco via phase separation driven by the multivalent interaction of an intrinsically disordered linker protein called EPYC1 with the Rubisco small subunit (RBCS) (Mackinder et al. 2016; Freeman Rosenzweig et al. 2017). Pyrenoid assembly principles appear to be broadly conserved in diverse algae lineages, as shown by the recent discovery of the disordered Rubisco linker protein PYCO1 in diatoms (Oh et al. 2023), which shows a similar architecture such as repetitive disordered regions interspaced with  Rubisco interacting motifs (He et al. 2020; Meyer et al. 2020). While the principle of this matrix formation is well characterised, the precise molecular mechanism that underpin the assembly principles of the other structural components in the *Chlamydomonas* pyrenoid remains poorly understood. One such component is the formation of a starch sheath surrounding the pyrenoid matrix.

In *Chlamydomonas*, the pyrenoid starch sheath forms under CCM-inducible conditions concomitantly with the increase in pyrenoid matrix size (Ramazanov et al. 1994). The

pyrenoid starch exhibits a cup-shaped morphology in contrast to the lenticular stromal starch when cells are grown in High-$CO_2$ (5%) condition (Ramazanov et al. 1994). Despite their difference in morphology, the isolated starch content in Low and High-$CO_2$ conditions displays little difference in total amount (Izumo et al. 2011) suggesting that the starch synthesis enzymes underlying their formation are likely similar. It is hypothesised that the formation of starch granules surrounding the pyrenoid could be a direct consequence of the steep 3-PGA gradient generated by the Rubisco activity at the pyrenoid matrix (Ball et al. 1991). However, their unique morphology is possibly regulated by an additional suite of starch-associated proteins which are conditionally expressed during CCM inducing conditions. Indeed, one example is SAGA1 (StArch Granules Abnormal 1),where in its absence the starch sheath is thinned and elongated, as well as yielding multiple smaller pyrenoids (Itakura et al. 2019). It is hypothesised that SAGA1 increases the starch sheath affinity to the Rubisco matrix via two domains: Rubisco binding motifs (RBM) and carbohydrate-binding module 20 (CBM20), which in turn regulates the starch sheath morphology. However, these aberrant starch sheaths still deviate from the typical stromal starch granules suggesting that additional proteins are at play in shaping the pyrenoid starch sheath. Recent work in *Arabidopsis* found multiple proteins containing coiled-coil domains to be important in regulating starch granule synthesis and in extension granule morphology. PTST (Protein targeting to starch) proteins carries both a CBM48 domain and a coiled-coil structure. PTST1 was found to direct the GBSSI (Granule-bound starch synthase 1) towards starch granules via electrostatic interaction driven by their respective coiled-coil structure (Seung et al. 2015). PTST2 and PTST3 are instead thought to be crucial for starch granule initiation through their interaction with SS4 (Starch synthase 4; (Seung et al. 2017)). Also, immuno-precipitation of PTST2 later identified two additional coiled-coil proteins, MRC (Myosin-Resembling Chloroplast Protein) and MFP1 (MAR Binding Filament-like Protein 1) involved in starch granule initiation. A similar example of a starch-related coiled-coil protein has also been found in *Chlamydomonas*. *Chlamydomonas* cells experiencing nutrient deprivation such as nitrogen starvation condition are known to rapidly accumulate transitory (stromal) starch while pyrenoid starch sheath is degraded (Blaby et al. 2013). Mutation in the *BSG1* (Bimodal starch granule 1) gene encoding a coiled-coil protein with CBM20 domain was found to cause cells to excessively accumulate pyrenoid starch under nitrogen starvation (Findinier et al. 2019). These examples together highlight a possible conserved function for CBM domain containing coiled-coil proteins participating in starch metabolism.

Despite its prevalence in pyrenoid-containing eukaryotic algae, the participation of the starch sheath in the *Chlamydomonas* CCM had been called to question. Not only does metabolic flux modelling of *Chlamydomonas* dispute the starch sheath's proposed function as a $CO_2$

diffusion barrier (Küken et al. 2018), a starch-less mutant which carries no pyrenoid starch sheath appeared to exhibit wildtype like oxygen evolution measurement suggesting CCM-induction is not compromised (Villarejo et al. 1996). However, recent work finds a small but noticeable difference in $C_i$ affinity when comparing starch-less and wildtype mutants. More importantly, critical CCM components such as LCIB and CAS1 are found to mis-localise in strains with aberrant starch sheaths (Toyokawa et al. 2020; Shimamura et al. 2023). In the present work, we follow up on the RBCS and LCI9 proximity labelling results that identified potential additional starch-associated proteins in the *Chlamydomonas* pyrenoid and found that Cre09.g394473 (*PINS1/LCI9*), Cre09.g394510 (*PINS2*) and Cre09.g394547 (*PINS3*) form a three gene cluster that shares a domain architecture comprising of a starch-binding domain and a coiled-coil domain. Additionally, they were all shown to localise to the pyrenoid starch sheath gap, which further hints at their participation in the maintenance of pyrenoid starch sheath morphology. To elucidate their function, we have chosen PINS2 as a primary target and generated two independent CRISPR knock-out lines. Detailed investigation into their starch morphology suggests that PINS2 modulates starch sheath assembly possibly by limiting excessive granule growth at the starch sheath edge.

## Results

### Cre09.394510 encodes one of three starch-binding proteins in a gene cluster

A recent study exploring the pyrenoid composition through proximity labelling has identified a previously uncharacterised pyrenoid component Cre09.g394510 (Lau et al. 2023). Cre09.g394510 contains two predicted domains, Carbohydrate-binding module family 20 (CBM20) known to bind starch (Christiansen et al. 2009) followed by a predicted coiled-coil region that is homologous to the structure maintenance of chromosome family (SMC) (Harvey et al. 2002). Cre09.g394510 was shown to localise to the pyrenoid starch sheaths interface, a localisation pattern shared with a Low-$CO_2$ inducible protein called LCI9 (Yamano et al. 2008; Mackinder et al. 2017). Due to their apparent similarity, we searched for additional genes in the *Chlamydomonas* genome v5.6 using keywords such as “Starch” and “CBM” to curate a list of starch-related proteins (Supplemental Data Set S1) and performed coiled coil domain prediction on the previously uncharacterised proteins using PCOILS on (Zimmermann et al. 2018) (Supplemental Fig. S1). Through this search we have additionally identified Cre09.g394547 and Cre09.g394473 (LCI9) as genetic neighbours of Cre09.g394510’s and share similar domain arrangement, making them likely homologues (Figure 1A and 1B; Supplemental Fig. S2). For simpler reference, we hereafter renamed LCI9, Cre09.g394510 and Cre09.g394547 as PINS1 (Pyrenoid interface starch protein, Cre09.g394473), PINS2 (Cre09.g394510) and PINS3 (Cre09.g394547) respectively based on their chromosome position. Like many CCM-related genes within the pyrenoid such as

EPYC1, CAS1 and SAGA1, position-specific iterated blast (PSI-blast) search of the PINS protein against non-redundant protein sequences suggests that the PINS protein is not widely conserved, as protein homologues are only found in closely related *Chlamydomonas* species. We next performed sequence alignment on PINS1 – 3 and found the three proteins share ~30% sequence similarity (Supplemental Fig. S2). While the domain arrangement is largely conserved between all three proteins, PINS2 exhibit two unique features including only a single CBM20 domain and an extended coiled-coil region (Fig. 1B – 1D). Despite the low sequence conservation, presence of their annotated domain alongside with the extended coiled-coil region is still reflected in their Alphafold modelling (Supplemental Fig. S3A). The modelled PINS's CBM20 domains are largely similarly to the known starch binding CBM20 structure 1AC0 from *Aspergillus niger* (Sorimachi et al., 1997). Notably, the two critical residues: W590 and W543 found in the starch binding pocket 1 (BdS1) hypothesised to bind to amylose in 1AC0 is conserved in the PINS modelled structure (Fig. 1D; (Sorimachi et al. 1997; Ngo et al. 2019)). However, an additional beta strand, located away from the starch binding pocket can also be found. This beta strand at the C-terminal of the CBM20 domain connects the CBM20 domain to the coiled-coil domain in the modelled structure. A non-conserved long disordered region (residue 222 – 278) is also present in the middle of PINS2's CBM20 domain between two β-strands and is situated away from the starch binding pockets, which might serve an alternative function (Fig. 1D, pink arrow).

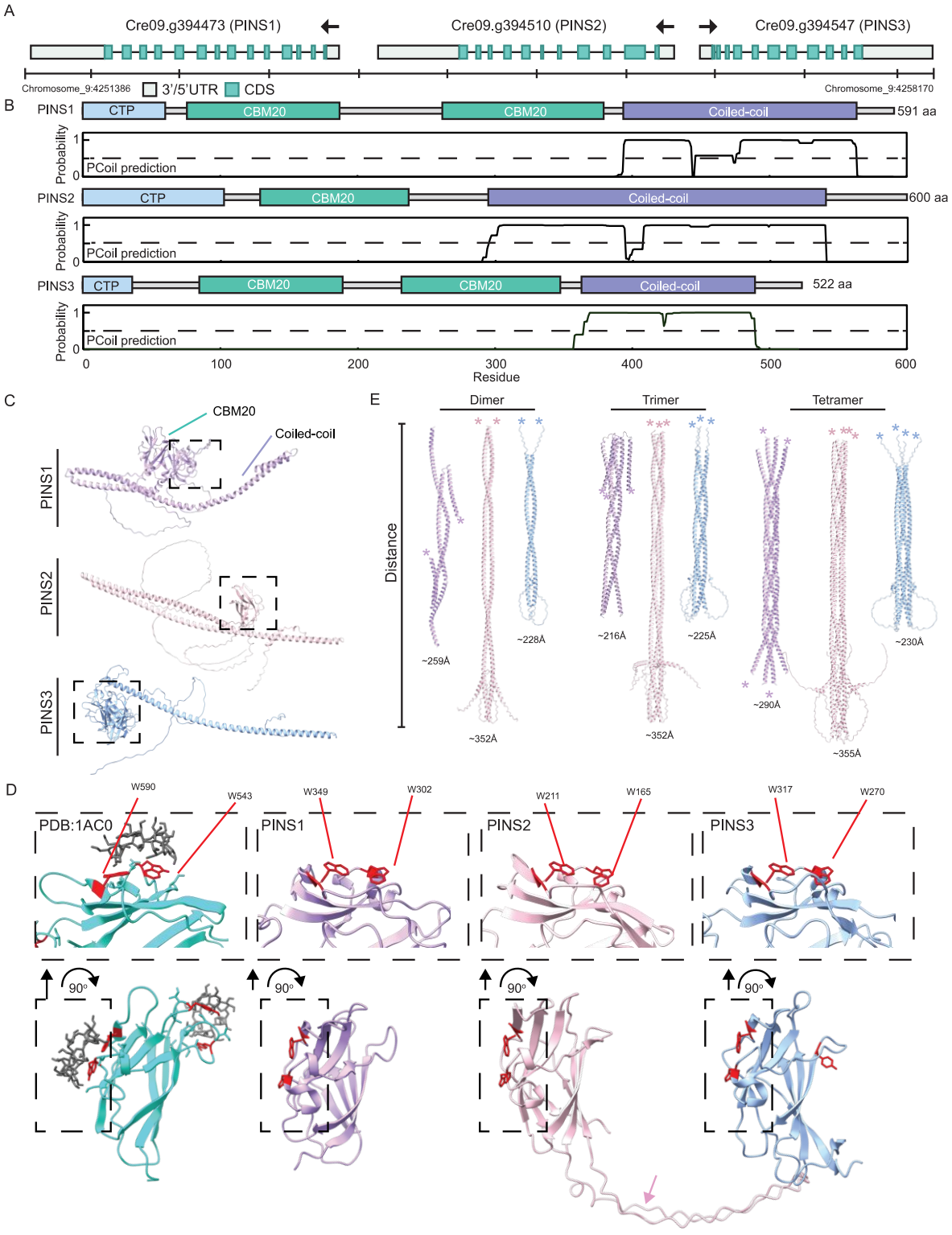



**Figure 1.** Genetic analysis of the PINS gene cluster. **A)** Corresponding chromosome position of the PINS1 (Cre09.g394473), PINS2 (Cre09.g394510) and PINS3 (Cre09.g394547), Chromosome position from Chromosoe_9:4251386 – Chromosome_9:4258170 is displayed. The marked ticks on axis are separated by 2000 bp. The Black arrows above each locus refer to the transcription direction for each gene. **B)** Domain arrangement of the PINS protein family. CTP, Chloroplast transit peptide; CBM20, Carbohydrate binding module 20. The probability to form coiled-coil structure is predicted via

the Pcoil webserver (Gruber et al. 2006; Gabler et al. 2020) using a scanning window-length of 28 residue. CBM20 Domain analysis of the three PINS protein. **C)** The Alphafold model of the PINS1; PINS2; and PINS3. The dashed rectangle highlights the CBM20 domain on the protein model. **D)** Analysis of the CBM20 region of PINS1 – 3 against the *Aspergillus niger* Bsd1. The three PINS model structures were aligned to the *Aspergillus niger* CBM20 domain using the Matchmaker function in ChimeraX. Highlighted are the two critical residues on Aspergillus niger Bsd1 W590 and W543 (red) as well as the corresponding residue in PINS1 – 3 structures. **E)** Alphafold modelling of the PINS1–3 coiled-coil domain region as homomeric dimers, trimers, and tetramers via the Alphafold Colab book. The position of the CBM20 domains is represented by asterisks coloured respective to the protein model. The distance between the two furthest atoms was then measured and noted below each structure. The protein model with the highest pLDDT scores are displayed.

Coiled-coil domains in proteins like SNARE complexes and tropomyosin mediate multimer assembly and are critical for their function (Stein et al. 2009; Li et al. 2010). To explore the propensity of multimer assembly in PINS1 – 3, their coiled-coil region was isolated and modelled as homomeric dimers, trimers and tetramers representing the most observed oligomeric states. When modelled as dimers, the two helices of PINS2 and PINS3 twist to form a parallel coiled-coil with a standard a-d core geometry, while PINS1 yields an antiparallel coiled-coil where its N-terminal bends to form a partial trimer connected with a middle dimer region (Fig. 1E, purple asterisk). In contrast, the trimer and tetramer models of PINS1-3 forms stereotypical parallel coiled-coil architecture with differing supercoil periodicity (Fig. 1E). PINS2, particularly, seemed to exhibit uneven periodicity along its axis. In agreement with the longer predicted coiled-coil domain in PINS2, the distance across the coiled-coil axis is measured as ~352 – 355 Å (35 nm), longer than that of PINS1 and PINS3 by ~100 Å (Fig. 1E).

**PINS1 and PINS2 proteins have differential locations at the starch sheath contact sites.**

We previously reported PINS2-Venus forms fluorescent puncta surrounding the pyrenoid and suggested the observed puncta are indicative of a starch sheath gaps localisation (Lau et al. 2023). To confirm PINS2's precise placement, we co-expressed PINS2-Venus alongside the starch marker protein STA2-mCherry (Fig. 2A). We found that PINS2-Venus is sandwiched between the STA2-mCherry confirming the former starch sheath gap localisation (Fig. 2B). This is not their exclusive location however, as plotting the intensity profile across pyrenoid shows that PINS2 additionally lines the starch sheath that is facing the pyrenoid matrix (Fig. 2C). The pixel intensity at the starch sheath gap is calculated to be ~3 times higher than that at the starch-matrix interface (Supplemental Fig. S3). Therefore,

the observed signal enrichment at the starch sheath gap is unlikely to stem from the signal addition coming from both sides of the starch sheath lining alone.

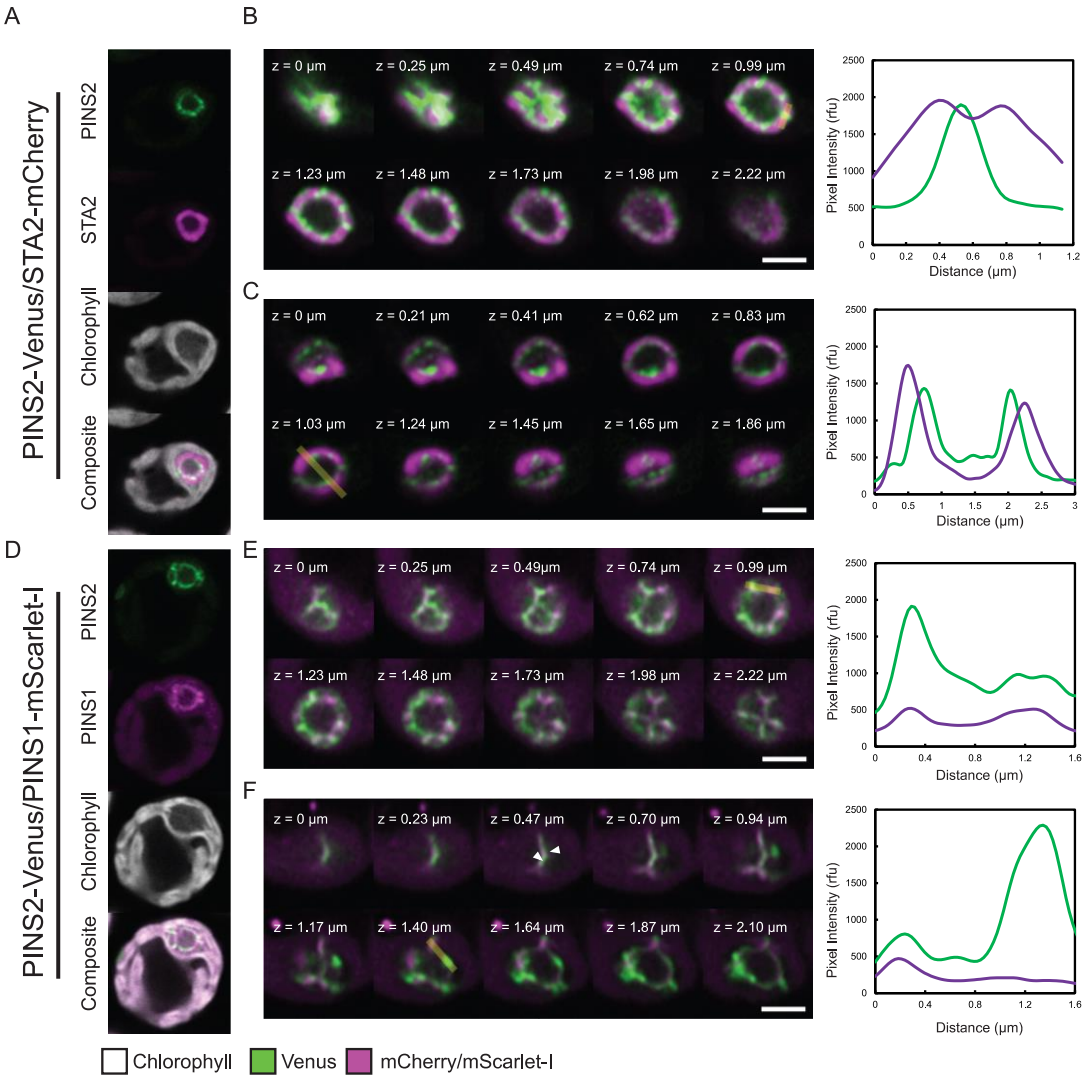



**Figure 2.** Localisation of the PINS1 and PINS2 via confocal imaging. **A)** Representative image of cells co-expressing PINS2-Venus along with the starch marker STA2-mCherry. **B – C)** Representative Z-section across the pyrenoid volume was imaged on the PINS2/STA2 dual-tagged line, a line (yellow) was drawn across the neighbouring starch sheath plates **(B)** or through the vertical axis of the starch sheath **(C)**. The corresponding intensity profiles are plotted on the right panel where green and purple line represent the measured PINS2-Venus and STA2-mCherry intensity respectively. **D)** Representative image of cells co-expressing PINS2-Venus with the PINS1-mScarlet-I. **E – F)**, Z-section across the pyrenoid volume in the PINS1/PINS2 dual-tagged line, a line is drawn across two presumed gaps between starch plates which shows enrichment in both PINS protein **E)** or enrichment of only PINS2 **(F)**. The corresponding intensity profiles are plotted on the right panel where green and

purple line represents the measured PINS2-Venus and PINS1-mScarlet-I intensity respectively. Scale bar = 2 µm.

Due to their similarly described localisation pattern, we wondered if PINS1 and PINS2 are enriched identically at the starch sheath gaps. We therefore co-expressed PINS1-mScarlet-I and PINS2-Venus in wildtype background. Confocal imaging of the dual-tagged lines showed that while both PINS proteins are targeted to the starch sheath gap, varying degree of enrichment can be observed at the same gap location. The measured enrichment between PINS protein seemed stochastically distributed and showed little correlation with each other. Gap position where either both PINS1 and PINS2 are highly enriched (Fig. 2D; 2E); or only one of the PINS proteins is apparent can be found (Fig. 2F). To capture the three-dimensional features of the pyrenoid, a Z-section across the pyrenoid volume was imaged. The top-section of the pyrenoid outlines the starch sheath gaps and clearly demonstrates the differential signal enrichment between PINS1 and PINS2 can be observed even on the same contact site between starch sheaths (Fig. 2F; white arrows). Additionally, the pyrenoid-facing starch lining localisation observed in PINS2 is either not found or to a much lesser extent in PINS1 which gives additional evidence to deviation in their enrichment mechanism.

**PINS2 deficient cells exhibit a CCM-related growth phenotype**

Due to their clear association with the pyrenoid starch sheath, we set out to explore whether these proteins are involved in the starch sheath formation. We introduced mutation to the *PINS2* locus via CRISPR-Cas9 mediated gene editing technology. Briefly, we adopted the method developed by Greiner et al., 2017. Cas9-gRNA protein complexes (RNP) were tested on purified DNA to confirm efficiency before co-transformed alongside the amplified hygromycin (*AphVII*) resistance cassette into wildtype *Chlamydomonas* strain (CC-5325) via electroporation (Figure 3A, Supplemental Fig. S4). The resulting transformants were then screened to identify introduction of the resistance cassette to the locus via pooled-PCRs using primers which bind to the flanking region of target region and the resistance cassette (Fig. 3A and 3B). We identified multiple colonies which carry the resistance gene. The edited locus was then sequenced (Fig. 3B). We found that in *pins2(d8),* a truncated hygromycin resistance cassette is inserted in the sense direction 4 bp upstream of the PAM recognition site, while in *pins2(f5)* line, a complete hygromycin resistance cassette in the antisense direction 4 bp upstream of the PAM recognition site. The latter strain was seen to have introduced a premature stop codon in-frame which likely disrupted its translation (Fig. 3B, black arrow) and was therefore used as the background line for complementation. To complement the *PINS2* locus, 3816 bp upstream of *PINS2* locus was cloned to the pLM099

via the recombineering method (Emrich-Mills *et al.,* 2021), expressing the PINS2 protein with a Venus fluorophore. The constructed plasmid was transformed into *pins2-f5* colony for complementation. Expression of PINS2-Venus was then screened by fluorescence and further verified via immunoblotting along with confocal imaging to ensure PINS2-Venus are localised to the correct location (Supplemental Fig. S5). The two resultant confirmed lines are then designated as *pins2*(*f5)*-C1 and *pins2(f5)*-C2. To see if the deletion of PINS2 resulted in a CCM-related growth defect, the two identified CRISPR-KO line *pins2(f5)* and *pins2(d8)* were grown alongside their genetic background CMJ030 (Wildtype) under various $CO_2$ conditions (Fig. 3C). We additionally screened their growth response in different medium pH to shift the $CO_2$: $HCO_3^-$ balance to create increased pressure on the CCM. Here we observed that the two *pins2* knock-out lines exhibit a CCM-responsive growth defect, where lines show impaired growth in $CO_2$ limiting environment (0.4% and 0.01% $CO_2$) while growing indistinguishably compared to wildtype when supplemented with $CO_2$ (3%). This $CO_2$ limitation to growth is exacerbated at higher pH (pH 8.2), where $CO_2$ availability is further limited. Interestingly, the use of complemented lines only yields small growth recovery at very low $CO_2$ and still fails to grow at high pH condition. One explanation is that the addition of a fluorescent protein tag at its C-terminus possibly impedes its ability to drive interaction through its C-terminal coiled-coil domain.

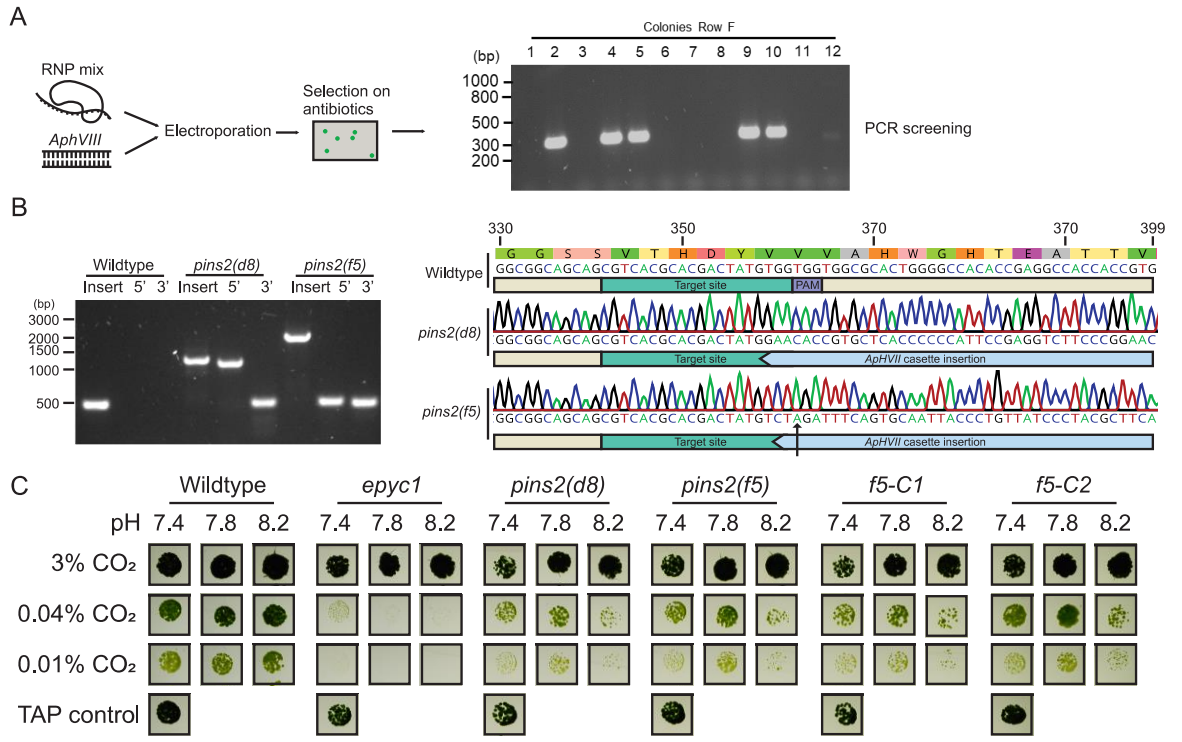



**Figure 3.** CCM-dependent phenotype of the *pins2(f5)* mutant. **A)** Schematic representation of CRISPR-Cas9 RNP knock-in experiment performed. A representative screening PCR is displayed. **B)**

PCR of the edited site in isolated CRISPR mutant lines and the wildtype background. Sequencing results of the amplified junction on the edited locus is shown on the right. The black arrow on pins2(f5) sequencing reads indicates an in frame stop codon. **C)** Wildtype, *epyc1*, *pins2(d8)*, *pins2(f5)* and *pins2(f5)* expressing PINS2-Venus (*f5-C1* and *f5-C2*) were grown under three different $CO_2$ concentration in autotrophic conditions. Spots grown under inorganic carbon supplement (TAP control) are used as a positive control. Each spots represents ~1000 cells. All images presented here under the same conditions (e.g. pH, $CO_2$ level) originated from the same plate and are re-arranged. Images of the annotated plate in full is shown on Supplemental Figure S6.

**The *pins2* mutant shows diverse pyrenoid starch organisation and has a swelled thylakoid tubule network.**

To assess whether the starch sheath region is affected by the PINS2 deletion, we introduced the fluorescently tagged starch marker protein STA2 (Starch synthase 2) and SBE3 (Starch branching enzyme 3) to both the wildtype background and the *pins2(f5)* mutant (Fig. 4A, Supplemental Fig. S7). Taking confocal images of pyrenoids in both lines, the total STA2-Venus signals were used as a proxy for starch sheath thickness due to their exclusive starch sheath localisation and overall brightness. We found that the total starch amount in *pins2(f5)* mutant is lower than the wildtype line irrespective of the overall cell sizes (Fig. 4A, right panel). The *pins2(f5)* mutant also display an aberrant starch sheath organisation, where overlapping starch sheath (Fig. 4A, white arrow in zoomed region) can be found surrounding the pyrenoid matrix. Our results suggest that *PINS2* deletion impairs overall starch synthesis and possibly perturbs the starch sheath assembly around the pyrenoid matrix.

With the reduced overall starch area in *pins2* mutant and PINS proteins' localisation to the pyrenoid starch sheath gaps, we explored the starch sheath in *pins2(f5)* in greater detail through transmission electron microscopy. To explore this, wildtype, *pins2(f5)*, *pins2(d8)* were grown in autotrophic conditions supplemented with 3% $CO_2$. Cells were then sampled after altering $CO_2$ supplementation to air-level (0.04% $CO_2$) to fully induce the CCM for 48 hours. Roughly more than 35 pyrenoids were imaged for each line. In comparison to the wildtype, both *pins2* mutants exhibited greater starch sheath variation in size and arrangement (Fig. 4B). Notably, many starch sheaths are shown to be thinner than Wildtype. Starch sheath overlaps in the *pins2* mutants can also be observed (Fig. 4B), which supports the phenotype observed in fluorescence imaging of STA2-Venus in *pins2(f5)*. Aside from the apparent differences in starch sheath morphology, the pyrenoid tubule network within *pins2* mutants was shown to be much more extensive and seem to occupy a larger volume of the pyrenoid matrix (Fig. 4B). The increased volume of pyrenoid tubules is likely not a consequence of photosynthetic damage, as neither the pyrenoid tubule structure nor thylakoid membrane neighbouring the pyrenoid appear swollen (Figure 4C) which is commonly observed in cells experiencing high light stress (Majeran et al. 2001). This

together shows that not only is the PINS2 protein required for the correct starch sheath formation but could also play a role in regulating thylakoid tubule occupancy within the pyrenoid matrix.

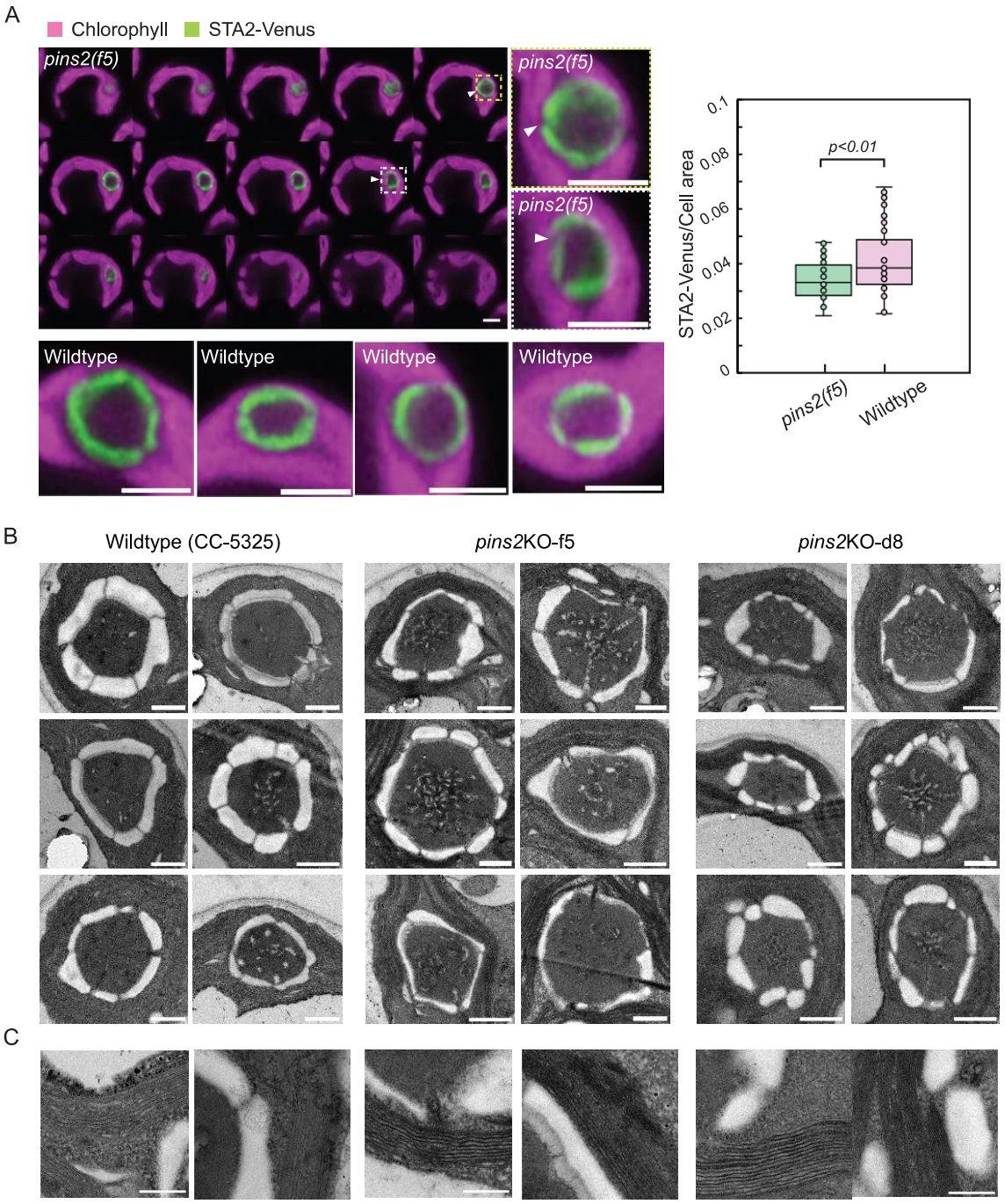



**Figure 4.** Transmission electron micrographs comparison between the pyrenoids in *pins2* (d8 and f5) mutants against their wildtype background (CC-5325). **A)** The Starch sheath visualised by STA2-Venus. White arrow indicates the starch sheath aberration observed in the z-section of *pins2(f5)* which is further highlighted with in the zoom region. Representative images of wildtype expressing STA2-Venus are also shown. To compare the total starch area per cell size between WT and

*pins2(f5)*, an area is drawn around the STA2-fluorescence, chlorophyll is then inferred as cell size. n = 29 – 31, p-value = 0.006785. Scale bar represents 2 µm. **B)** Representative micrograph of the pyrenoid in the two *pins2* mutants (f5 and d8) and wildtype cells. The three strains were grown under autotrophic conditions and supplemented with 0.04% $CO_2$ to induce the CCM. The starch sheath is represented by the white granule structure enclosing the electron dense pyrenoid while the irregular network-like structures within the pyrenoid matrix represents the pyrenoid tubules. Micrographs are only shown from cells containing an intact pyrenoid and with no obvious whole cell defects. Additional micrographs of pyrenoids are included in Supplemental Fig. S8. Scale bar represents 500 nm. **C)** Representative micrograph of thylakoids in the two *pins2* mutants and wildtype cells. Scale bar represents 500 nm.

### *Pins2* mutant fails to aggregate LCIB at the correct location

Under very low $CO_2$ condition (<0.02% $CO_2$), LCIB responsible for converting stromal $CO_2$ to $HCO_3$ changes localisation from diffuse in the chloroplast stroma to a puncta-like formation surrounding the pyrenoid (Yamano et al. 2010, 2022). Their re-localisation is shown to be intimately linked with starch sheath morphology, where the tagged LCIB-Clover signal formed large aggregates at the basal region of the chloroplast in cells lacking an intact starch sheath due to a deletion of the starch synthase gene ISA1 (ISOAMYLASE 1, Cre03.g155001) (Toyokawa et al. 2020). We therefore examined whether LCIB re-localisation is perturbed in the *pins2* mutant that also exhibits starch sheath aberrations. We grew *pins2(f5)* and wildtype expressing LCIB-Venus photoautotrophically and supplemented them with 0.01% $CO_2$ to acclimate cells to the very low $CO_2$ condition for 24 hours prior to imaging. As slight differences in cell state or population could impact their photosynthetic performance leading to discrepancies in the $CO_2$ concentration within culture despite efforts to keep the culturing setup identical, we decided to co-cultivate *pins2(f5)*-LCIB-Venus cells with WT-LCIB-mCherry which allows us to monitor LCIB-mCherry localisation from wildtype within the same culture, therefore acting as an internal control. We found that in wildtype cells, the LCIB-Venus signal forms small puncta, which encircles the pyrenoid organelle upon acclimation to VLC condition (Fig. 5). While in the *pins2(f5)* mutant lines the LCIB-Venus signal formed a single large puncta that failed to re-localise directly surrounding the pyrenoid and was instead found at the basal region of chloroplast like the phenotype described of *isa1* and *sta11*. The internal control LCIB-mCherry signals exhibited a localisation pattern parallel to wildtype, suggesting that the failure of LCIB to re-localise in *pins2(f5)* is not a defect due to differences in $CO_2$ concentration between the cultures. These results give additional support that correct starch sheath formation, particularly at the starch sheath edge, is important for LCIB re-localisation towards the pyrenoid.

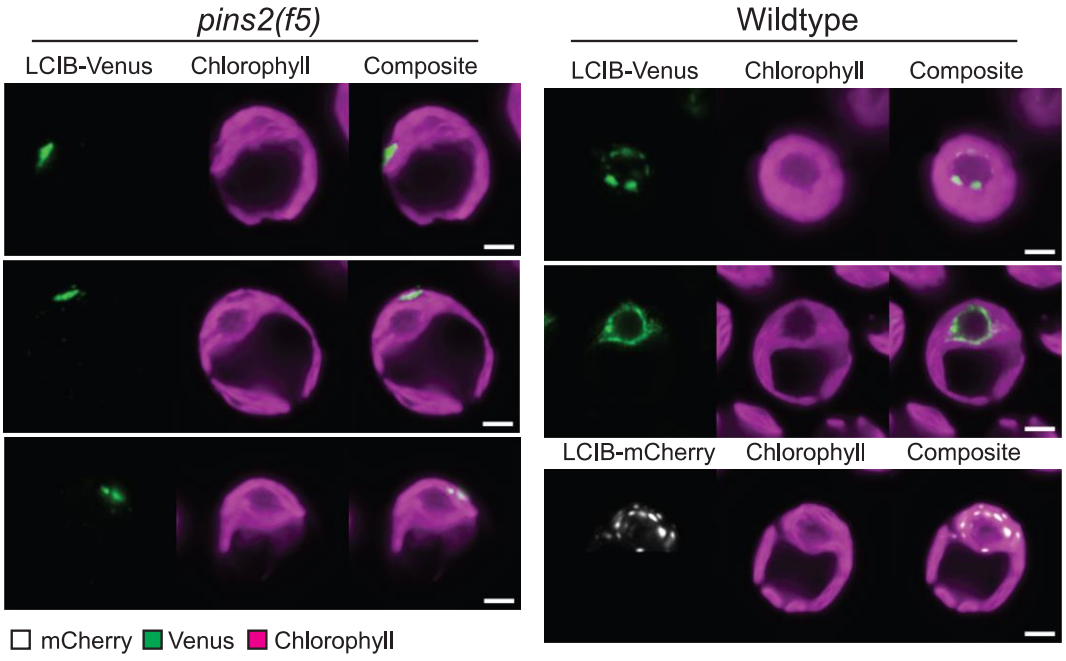



**Figure 5.** Comparison of the LCIB localisation at very low $CO_2$ condition in *pins2(f5)* and wildtype. Two independent transformants of *pins2(f5)* of LCIB-Venus was assessed. LCIB-mCherry line was co-cultivated with *pins2(f5)*-LCIB-Venus in the same flask as an internal culture control. This facilitates the direct comparison of LCIB localisation under VLC condition. Scale bar = 2 µm.

## Discussion

### PINS2 protein might line the starch sheath edge to regulate pore size

In this work, we examined the phenotype caused by a mutation in the gene *PINS2*. We further identified this gene as one of three genetic neighbours that carries a CBM20 – coiled coil domain arrangement. Based on their clear starch association, we examined their starch phenotype in detail and found that *pins2* mutants exhibit aberrant starch sheath formation where starch granules are often of irregular thickness and overlap with each other. Despite previously annotated as glucan 1,4-alpha-glycosidases, none of the three proteins are found to contain domains predicted to catalytically break down starch molecules suggesting that these proteins likely influence the pyrenoid starch sheath formation indirectly. Indeed, the ectopic over-expression of the CBM20 domain alone in Barley yields a compound-type starch granules formed seemingly from the incomplete fusion of multiple smaller starch granules (Zhong et al. 2020). These complex granules contrast the normally large round morphology and suggest that CBM20 domain competitively binds to starch surface perturbing the regular starch-protein association of starch synthases. Furthermore, *Ostreococcus tauri* starch synthase (OsttaSSIII-B) maximal activity on soluble starch is reduced by ~80% via the addition of CBM20CP protein, a CBM20-domain containing coiled-

coiled protein (Hedin et al. 2022). These together argue that the CBM20 domain likely inhibits starch granule synthesis. However, *PINS1-3* transcript levels during diurnal cycle clearly aligns with starch synthesis genes, which are upregulated in the beginning of the light period (Supplemental Fig. 9; (Strenkert et al. 2019)). Additionally, affinity purification of the PINS1 protein alongside with proximity labelling of this protein identified starch synthesis related proteins SBE3 as a close interactor (Mackinder et al. 2017). These seemingly contradictory data might be resolved in view of PINS localisation at the starch sheath gaps (Strenkert et al. 2019). We propose that as starch granule forms and grows during the light period, the starch sheath enclosing the pyrenoid matrix is enlarged. The peripheral localised PINS saturates the edge ends of starch. The enrichment of PINS and their CBM20 domains at the starch sheath edge limits starch granule growth towards their neighbouring granules and prevent their fusion, allowing starch granules to grow without severing the pyrenoid tubules entry into the pyrenoid matrix that is essential for $CO_2$ delivery within the pyrenoid. Proteins carrying a starch binding module and coiled-coil domains like *at*PTST1, *at*PSTS2, MRC etc. are known to be critical for starch granule synthesis. Through electrostatic interaction driven by their coiled-coil domain with interacting starch synthases, they either direct starch substrates towards starch synthases or localise starch synthases to existing starch granules which allow granule initiation (Seung et al. 2015, 2017, 2018). However, PTST1 has been shown be weakly surrounding starch granules in the chloroplast stroma, exhibiting a striking contrast to the PINS's starch edge localisation. Their different partition to the starch sheath surface is likely a property of their different coiled-coil domain, as another protein containing CBM48-domain, FLO6, are also shown to coat the starch granule structure in rice (Peng et al. 2014). More evidently, Cre09.g415600 similarly shares the CBM20 Coiled-coil domain arrangement, but fluorescent-protein tagging of this protein is shown to uniformly coat the starch sheath with little visible enrichment to the starch sheath gaps (Wang et al. 2023). Coiled-coil domains have the propensity for multimeric oligomerisation. Interestingly, modelling of this coiled-coil domains in homomultimeric formation often stretches the coiled-coil domain to form a long helix structure extending from ~200 nm to ~350 nm depending on the PINS proteins (Fig. 1E). This length span is coincidentally parallel to the distance measured across the starch sheath gaps. The peripherally enriched PINS might therefore physically restrain the merging of starch granules even though when appressed. Allowing the pyrenoid tubules to maintain their established network that traverses the pyrenoid without being severed by the growing starch sheath. Interestingly, confocal imaging of the PINS2 in parallel with PINS1 clearly demonstrates that both proteins do not exclusively co-localise together. Instead, small PINS enrichment can be seen along the same starch sheath gap which points to differential recruitment of these proteins to the starch sheath edge based on their coiled-coil properties. Together with the

multimer properties, this might allow starch gaps of different dimensions to appear, resulting in the larger gaps that maintains the pyrenoid tubule traversions through the starch sheath, and the smaller gaps found at starch plate-starch plate interfaces.

**Impact of starch morphology on the pyrenoid tubule network**

In *Chlamydomonas,* thylakoid membranes are arranged as sheets that form an interconnected complex network (Engel et al. 2015). When the thylakoid membrane reaches the pyrenoid, they merge and transform into circular tubules which extend into the pyrenoid matrix through the gaps between the starch sheath. These pyrenoid tubules then further fuse and contort into a reticulated membrane network within the pyrenoid matrix. The formation of pyrenoid tubule structure is at least partly influenced by the pyrenoid starch sheath, as starch-less *Chlamydomonas* mutants show a much more minimal pyrenoid tubule network (Villarejo et al. 1996; Toyokawa et al. 2020). However, *saga1* mutant cells possess multiple pyrenoids with abnormal starch sheaths, and only one of these pyrenoids contain a typical pyrenoid tubule network. This shows that the starch sheath alone does not recruit pyrenoid tubules. In the *pins2* mutant line, we saw an increased complexity of pyrenoid tubules in the pyrenoid matrix. The development of pyrenoid tubules into the convoluted pyrenoid network likely requires insertion of proteins into the membrane or formation of protein scaffolds to stabilise the curved membrane structure like the Arabidopsis CURT proteins (Armbruster et al. 2013). One hypothesis is that the rise in tubule network volume could be an aftermath of starch sheath malformation perturbing the targeting or functioning of these tubule-shaping proteins. It is possible that the PINS proteins are key to generating the local environment for these membrane proteins for their action. The pyrenoid tubules therefore quickly transformed into the more disorganised structure when PINS2 is deficient, leading to the increased network volume. Additionally, the restriction of pyrenoid tubules by the malformed starch sheath might limit metabolite transport and protein targeting towards the pyrenoid leading to knock-on effects that caused the observed morphological change. For example, low external $CO_2$ concentration is shown to trigger a $Ca^{2+}$ signal spike at the pyrenoid (Wang et al. 2016), it is possible that the compensatory ion import (such as $K^+$) from bulk stroma into the pyrenoid tubule is limited by the severed tubule. This could lead to an ion imbalance in the thylakoid lumen which causes thylakoid swelling. Recent reports also found multiple new proteins targeted to the pyrenoid tubules that likely functions in DNA/RNA processing and protein folding (Lau et al. 2023; Wang et al. 2023). The limited transport imposed by the severed pyrenoid tubules could lead to an overaccumulation of toxic side products which results in the swelled thylakoid network observed. Future work on visualising the pyrenoid tubules by tagging proteins such as PSAF or PSAH that are known to localise on the thylakoid membrane will give valuable information in addressing these hypotheses.

**PINS2 is required for correct recruitment of LCIB to the pyrenoid periphery**

The pyrenoid starch sheath forms a physical barrier surrounding the pyrenoid that is proposed to reduce $CO_2$ diffusion out of the pyrenoid matrix when the CCM is induced (Ramazanov et al. 1994). The reduction of $CO_2$ leakage is further supported by the carbonic anhydrase complex LCIB/LCIC which drives the preferential conversion of stromal $CO_2$ into $HCO_3^-$ that is supplied to the $HCO_3^-$ channels BST1-3 (Yamano et al. 2010; Mukherjee et al. 2019; Kasili et al. 2023). In sub air-level $CO_2$ (<0.02%), LCIB is shown to aggregate around the pyrenoid forming puncta surrounding the pyrenoid tubule entry position (Yamano et al. 2010; Mackinder et al. 2017). Their specific placement at the pyrenoid facilitate their function in preventing $CO_2$ leakage when the external $CO_2$ concentration is further limited. Recent work has shown that the presence of the pyrenoid starch sheath is key to the re-localisation of LCIB proteins as starchless mutants were shown to cause LCIB to aggregate at the basal area of the chloroplast (Toyokawa et al. 2020). One proposed mechanism of LCIB's dynamic localisation points to a protein present on the starch sheath, which directly recruits LCIB once the external very low $CO_2$ condition triggers their re-localisation. Our work here shows that the *pins2* mutant similarly fails to re-localise LCIB near the pyrenoid matrix. This failure to re-localise LCIB as well as PINS2's localisation at the starch sheath gap, where $CO_2$ leakage prevention is most necessary support PINS2 as a prime candidate that recruits LCIB to coat the starch sheath gaps. However, LCIB/LCIC is not found as interactors for PINS1 proteins nor enriched in the PINS1 proxiome, due to the close association between PINS1 and PINS2, this likely infers that PINS2 might facilitate LCIB re-localisation via alternative processes, such as the modulation of the starch sheath gaps for other proteins that recruit LCIB. Regardless, the discovery of the *PINS2* mutation resulting in mis-localisation of LCIB affirms that the pyrenoid starch is a crucial component of the CCM, and more precisely points to the starch sheath gaps as a key to LCIB re-localisation. Proximity labelling could help further elucidate the mechanism in which PINS2 participates in LCIB re-localisation to the pyrenoid.

**Future perspectives**

A major hurdle of engineering an efficient pyrenoid-based CCM in crop plants lies in minimising the $CO_2$ leakage stemming from the increased $CO_2$ delivery into pyrenoid matrix. The $CO_2$ which re-diffuses from the pyrenoid matrix effectively "short circuits" the $CO_2$ delivery system and thus creates an energetically wasteful futile cycle without providing the benefits of the CCM. To efficiently implement a pyrenoid-based CCM, a diffusion barrier must be built surrounding the pyrenoid to limit this $CO_2$ leakage (Fei et al. 2022). Expression of SAGA1 and SAGA2 in *Arabidopsis thaliana* containing a proto-pyrenoid (Atkinson et al. 2020, 2023) was shown to modify starch granules to a more plate-like morphology and drive

these starch granules to encircle the proto-pyrenoid. The next step towards engineering the physical barrier is to further shape the starch granules as shells to completely coat the proto-pyrenoid. To aid with this process, a more complete view of the pyrenoid starch sheath formation must be established. We envisioned that future works in characterizing proteins such as Cre09.g415600, PINS1–3, and PTST that contain both starch-binding and coiled-coil domains will be key to elucidate the pyrenoid starch sheath formation mechanism and to engineer it into plants. Furthermore, fine-tuning the expression levels of starch related proteins might also be an important consideration as disrupting native starch-protein association can lead to wider changes in cell metabolic state (Zhong et al. 2020). Our report here highlights PINS2 as a novel CCM-related protein which modulates the pyrenoid starch sheath formation in *Chlamydomonas reinhardtii* and provides new evidence for the pyrenoid starch sheath gap in facilitating $CO_2$ recapture by enabling accurate LCIB re-localisation. Altogether this work shows the PINS family as promising candidates to be introduced in higher plant systems in the effort of assembling a fully functional pyrenoid-based CCM.

## Methods

### Gene and protein sequence analysis of PINS1 – 3

The genomic and protein sequence of PINS1 (Cre09.g394473); PINS2 (Cre09.g394510); and PINS3 (Cre09.g394547) are used for sequence alignment in Geneious (Sievers et al. 2011). To identify related protein homologues, their protein sequences were submitted for NCBI position site specific blasting (PSI-Blast search) against the non-redundant protein database, all protein hits are documented in Supplemental Table 2. To uncover additional proteins carrying similar domain arrangement, keywords such as “CBM”; “CBM20”, “starch”, “carbohydrate” were searched on the Phytozome (JGI portal) on the *Chlamydomonas reinhardtii* v5.5 genome assembly. The resulting annotation and proteins are collated on Supplemental Data Set S1.

### Alpha fold Modelling and structure analysis of PINS1 – 3

To assess their CBM20 domain, Alpha fold model structure of PINS1, PINS2 and PINS3 was compared against the solved structure 1AC0 (Sorimachi et al. 1997). Alignment of their structure was performed by the Matchmaker function in ChimeraX (Meng et al. 2023) using 1AC0 as the reference structure. For multimeric assessment, the coiled-coil sequences were obtained from the Pcoil prediction webserver (Gruber et al. 2006), 400 aa – 570 aa in PINS1; 300 aa – 550 aa PINS2; and 350 aa – 550 aa in PINS3 were submitted to the Alphafold Colab book (Jumper et al. 2021; Mirdita et al. 2022) and modelled as homomeric dimers, trimers and tetramers. Only the structure with the highest pLDDT score were used in the comparison. Distance across the modelled coiled-coil domain was measured by picking the

two atoms furthest apart along the axis of coiled-coil structure using the distance function in ChimeraX. All protein models are visualised in UCSF ChimeraX (Meng et al. 2023).

**Growth and maintenance of *Chlamydomonas* strains**

*Chlamydomonas* cultures were maintained on TAP-agar medium with revised Hunter's trace elements (Kropat et al. 2011). For liquid cultures, *Chlamydomonas* are grown on orbital shaker at 140 rpm at ~21 °C under LED lights (Valoya C65 LEDs with AP673L spectrum) at ~50 µmol photons $m^{-2}$ $s^{-1}$.

**Imaging of fluorescently tagged lines**

To fluorescently tagged PINS1 and PINS2, genetic region including their native promoter region from of *PINS1*, *PINS2* are cloned via recombineering cloning to yield PINS1-mScarlet-I (pLM383) and PINS2-Venus (pLM547); (Emrich-Mills et al. 2021), see Supplemental Data set 1 for all primers used). Plasmids encoding LCIB-Venus (pLM012), STA2-Venus(pLM016), STA2-mCherry (pLM025) and SBE3-Venus (pLM020) were reproduced from Mackinder et al., 2016. To obtain fluorescent colonies, PINS1-mScarlet-I/PINS2-Venus were linearised with I-SceI; EcoRV for STA2-mCherry and LCIB-Venus; and DraI for SBE3-mCherry. *Chlamydomonas* cells grown in continuous light regime were harvested during mid-log phase (~$4\times10^6$ cells $mL^{-1}$) and resuspended in TAP medium supplemented with 40 mM Sucrose to a final cell concentration of $1\times10^8$ cells $mL^{-1}$ using the Countess® II FL Automated Cell Counter (Thermofisher). 120 µL Cells suspension was added to 2–mm electroporation cuvette with 200 ng linearised plasmid. Using the NEPA21 electroporator (NEPA gene), electric pulses were delivered following the below setting: Two poring pulse at 300 V with 8 ms pulse length, 50 ms pulse interval, and 40% decay rate. One transfer pulse at 20 V with 50 ms pulse length, 50 ms pulse interval. (Yamano and Fukuzawa 2020). Next, electroporated cells are resuspended in 8 mL TAP-sucrose medium in a 15 mL falcon shaking under dark overnight to recover. Cells were then harvested by centrifugation ×1500 g, 4 minutes at room temperature and spread on TAP medium containing the corresponding antibiotics for selection. Colonies appearing at approximately 10 days were picked to new TAP-Agar plates to screen for fluorescence using the Typhoon 8610 scanner. Fluorescent colonies with confirmed fluorescence are inoculated in TAP medium until mid-log phase in continuous light on 6-well plates. Cell cultures were resuspended in TP-medium overnight prior to imaging. Cells are laid onto µ-Slide 18 Well Chamber slides (Ibidi) and immobilised with 1.5% low melting point agarose (Invitrogen). Confocal imaging was performed on Zeiss LSM880 inverted microscope equipped with the Airyscan module. The excitation and emission filters of fluorophore and chlorophyll channels were set as follows: Venus/mNeon (excitation: 514 nm; emission: 520 to 550 nm);

Chlorophyll (excitation: 633; emission: 610 to 650 nm). Image analysis is carried out using the FIJI software. Bleach corrections using the simple ratio method (Miura 2020) was used to correct the Z-section display in Figure 3.

**CRISPR-Cas9 mutant generation**

To obtain the Cas9 protein, pET-28b-Cas9-His (Addgene, #47327) was transformed into BL21 DE3 (Rosetta) *E. coli* competent cells. Starter culture of Cas9 expressing cells is inoculated into 2 Litres of LB culture and grown until $OD_{600}$ reaches 0.6 – 0.8. IPTG was then added to a final concentration of 0.4 mM to induce Cas9 expression at 18°C for 16 hours. Cell harvested via centrifugation (×20000 g, 10 minutes, at 4°C) and resuspended in lysis buffer (20 mM HEPES·NaOH pH7.5, 300 mM NaCl, 50 mM Imidazole, 1 mM PMSF, 1x Roche cOmplete™ Protease Inhibitor, 1 mM β-mercaptoethanol) before sonication. Resultant lysate is clarified via centrifugation (×40000 g, 40 minutes, at 4°C) and purified using a HisTrap™ High performance Nickel column (Cytiva) on an ÅKTA™ Purifier chromatography system (ÅKTA). The Cas9 Protein was eluted using a gradient method with the elution buffer (20 mM HEPES·NaOH pH7.5, 300 mM NaCl, 500 mM Imidazole, 1 mM β-mercaptoethanol). The eluted fractions containing Cas9 was detected by SDS-PAGE and was further purified through Gel filtration using the HiLoad 16/600 Superdex 200 PG (Cytiva, 28989335). The purified Cas9 protein at a final concentration of 8 µg per mL are exchanged into storage buffer (20 mM HEPES·NaOH, pH 7.5, 150 mM NaCl, 1 mM DTT, 0.1 mM EDTA 50% Glycerol) and snap-freeze in liquid nitrogen and stored in -70°C freezer until use. gRNA target regions are designed using the CHOPCHOP webserver (Labun et al. 2019), high ranking targets on exons at the starting half of targeted gene are selected. Self-complementary oligos containing the T7 promoter and the gRNA scaffold sequence was designed according to (Yu et al. 2017) (see Supplemental Data set S3 for all primers used). *In vitro* gRNA synthesis is performed using the HiScribe T7 Quick High Yield RNA synthesis Kit (New England Biology, E2050S) and purified using Monarch RNA clean-up columns (New Engand Biology, T2057L) according to the manufacturer's protocol. To allow selection of transformed cells, DNA fragments encoding for *AphVII* (Hygromycin) resistance are amplified from pLM162 and purified using the QIAquick PCR purification kit (Qiagen).

To obtain CRISPR Knock out lines at the *PINS2* locus, a RNP electroporation approach according to Greiner et al., 2017 was followed with minor modification. Briefly, *Chlamydomonas* cultures are grown under synchronised light condition (12 hours light / 12 hours dark). Cultures are placed into 40°C incubator shaking at 220 rpm for 30 minutes just before entering the dark phase. Cells are then resuspended in TAP medium supplemented with 40 mM Sucrose. Purified Cas9 protein and the gRNA are mixed at a molar ratio of 1:3 and is incubated in 30°C for 15 minutes. The Cas9/gRNA mixture alongside with 2000 ng of

purified *AphVII* cassette are added to 150 uL of harvested cell in 4-mm electroporation cuvettes. Electroporation using Gene Pulser II (Biorad) is performed using following settings: 800 V, 25 µF, 1000 Ω. Cell suspension is then recovered overnight in TAP-Sucrose medium at room temperature before plated on TAP-agar plates containing hygromycin B (10 µg $mL^{-1}$). Colonies appearing after ~12 days are picked to grow on 96-well plates in TAP medium. 10 µL Chlamydomonas culture is mixed with 50 µL of 10 mM EDTA, pH 8 and boil for 10 minutes to obtain a crude DNA extraction. To identify edited lines, primers flanking the targeted site is used and amplicons shown to be different from wildtype lines are isolated and sequence verified.

**Transmission electron microscopy**

The protocol used for processing *Chlamydomonas* cells for TEM broadly followed that gifted by Martin Jonikas group (personal communication). 400 mL cells are grown in liquid TP-medium bubbled with 3% $CO_2$ to mid-log phase. Cultures are then switched to bubbling with 0.04% $CO_2$ gas mixture overnight before harvesting via centrifugation (×1000 g, 4 minutes, at room temperature). Following centrifugation (1000g, 5 min at all following steps), cells were fixed in 2.5% Glutaraldehyde in TP-medium for 1 hour at room temperature on a rotator. Fixed cells were washed (minimum 10 minutes) before incubation in 1% $OsO_4$, 1.5% (w/v) K3[Fe(CN)6] on ice , in dark, for 1- 2 hours. The osmicated samples are washed with water as above resuspension in low melting point agar in distilled water (3%) to give a final agar concentration of 1-2%. Agar blocks (max 2 $mm^2$) containing cells were dehydrated through a graded ethanol series (30%, 50%, 70%, 90% 10 minutes at each concentration, followed by two 15-minute incubations in 100% ethanol), before infiltration with Spurr replacement resin (TAAB Laboratories Equipment Ltd, England); Resin: Ethanol, 1:2. 1:1. 2:1. Followed by three changes of 100% resin. Each resin change was a minimum of 1 hour and at least one 100% incubation was overnight. Samples were embedded in flat embedding moulds and resin was polymerised at 70°C for 48 hours. Ultrathin sections (70 nm) are cut using a Leica EM UC7 ultramicrotome and collected on uncoated copper grids (200 mesh). Sections were poststained with 2% Uranyl acetate in 50% Ethanol (in dark) followed by Reynold's lead citrate (Reynold 1963) in $CO_2$-depleted chamber, both steps carried out for 5 minutes. Images are acquired using a Jeol JEM-1400 Transmission electron microscope operating at 120kV (Jeol USA).

**Complementation and growth assay**

To assess growth phenotype of the generated strains, *pins2(f5)* was transformed with pLM547 as described above to obtain a complemented line. Wildtype, *pins2(f5)*, *pins2(d8)* and the complemented line (f5-C1 and f5-C2) were inoculated in TAP medium and grown on

shaker at 140 rpm. Once reaching mid-log phase, cells were resuspended in TP medium and cell concentration was measured using Countess® II FL Automated Cell Counter (Thermofisher). Cell suspensions were serially diluted to 100 cells $\mu L^{-1}$, 100 cells $\mu L^{-1}$ and 1 cells $\mu L^{-1}$ in TP medium. 10 µL of cell suspension at each dilution was then spotted on to a TP Agar plates (1.5% Agar) and TAP agar plates for control. Culture plates were then placed under a growth chamber under 400 µmol photons $m^{-2}$ $s^{-1}$ for 7 days. Images of each plate are taken using the PhenoBooth+ Colony counter (Singer).

**Protein extraction and immunoblotting**

Complementation of the *pins2* mutant was verified using immunoblotting. For protein extraction, $1\times10^7$ *Chlamydomonas* cells grown in TAP medium were harvested by centrifugation (×1000 g, 4 minutes, 4°C) once cell culture has reached mid-log phase. Cell pellets were resuspended in 100 µL protein extraction buffer (25 mM Tris–HCl pH 7.4, 300 mM NaCl, 1 mM DTT, 5 mM MgCl2, 0.1 mM PMSF, 1× EDTA-free protease inhibitor [Roche], 0.1% [w/v] SDS, 0.5% [w/v] deoxycholic acid, and 1% [v/v] Triton X-100) and lysed by five freeze-thaw cycle using liquid nitrogen. Lysed cell suspension was centrifuged (×17000 g, 10 minutes, 4°C) to obtain the clarified protein extract. 20 µL of 6 × protein loading dye was added to the protein extract and was denatured on heat block for 10 minutes at 95°C. 20 µL of protein extract was loaded onto Mini-PROTEAN® TGX™ (BIORAD) Precast PAGE gels and transferred to PVDF membrane via a semi-dry transfer system. Anti-FLAG primary antibody (1:1000, Sigma #F1804) was used to detect the PINS2-Venus expression at 1:1000 dilution. Anti-Mouse Alexa Fluor plus 555 (Invitrogen #A32727) was used as secondary antibody to visualise detected FLAG signal using the Typhoon 5 scanner (ThermoFisher).

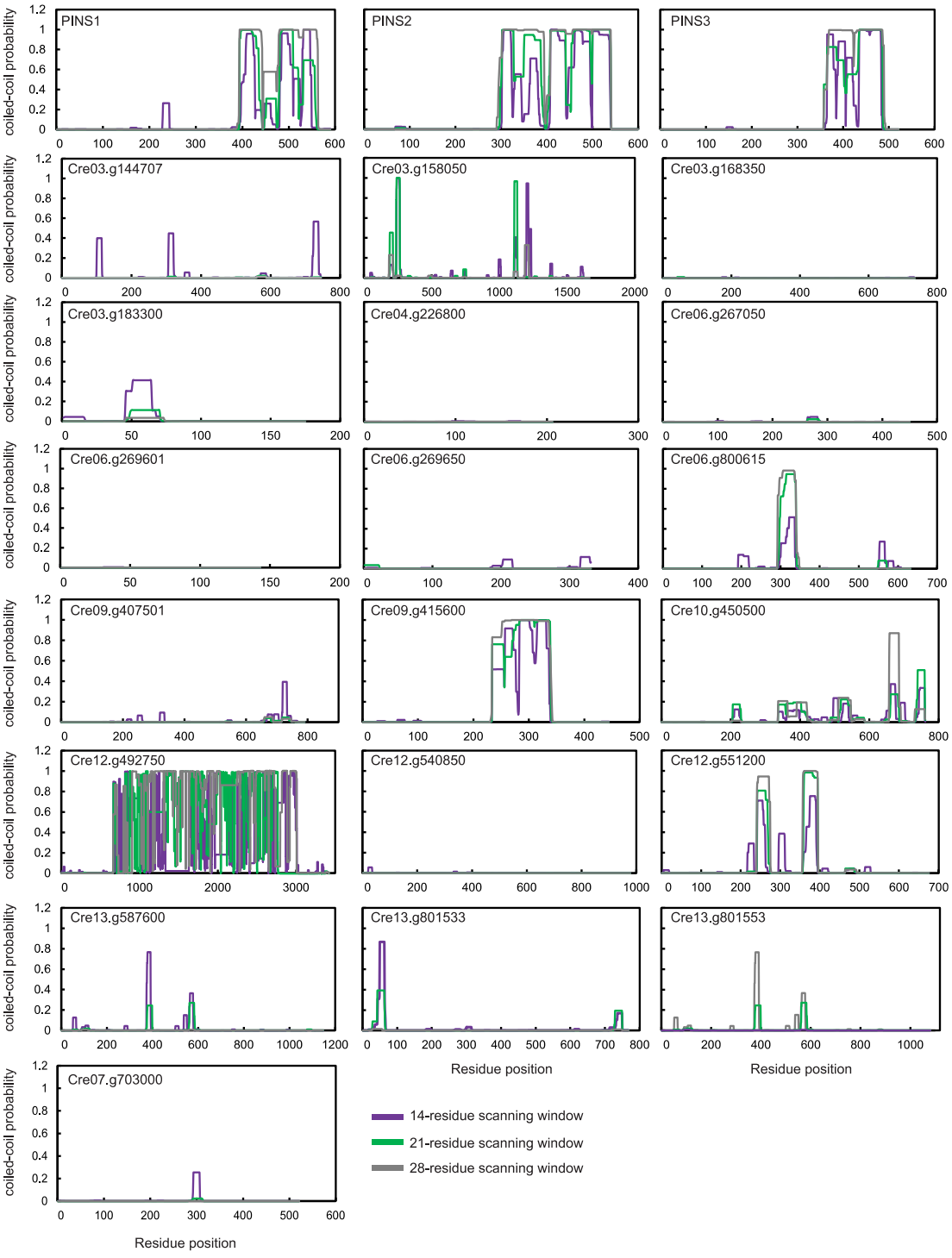
**Accession numbers**

Sequence data from this article can be found in Phytozome, the Plant Comparative Genomics portal of the Department of Energy's Joint Genome Institute, under the following accession numbers: Cre09.g394473: PINS1/LCI9; Cre09.g394510: PINS2; Cre09.g394547: PINS3; Cre17.g721500: STA2; Cre10.g452800: LCIB; Cre06.g307500: LCIC; Cre11.g467712: SAGA1; Cre09.g394621: SAGA2.

**Author contribution**

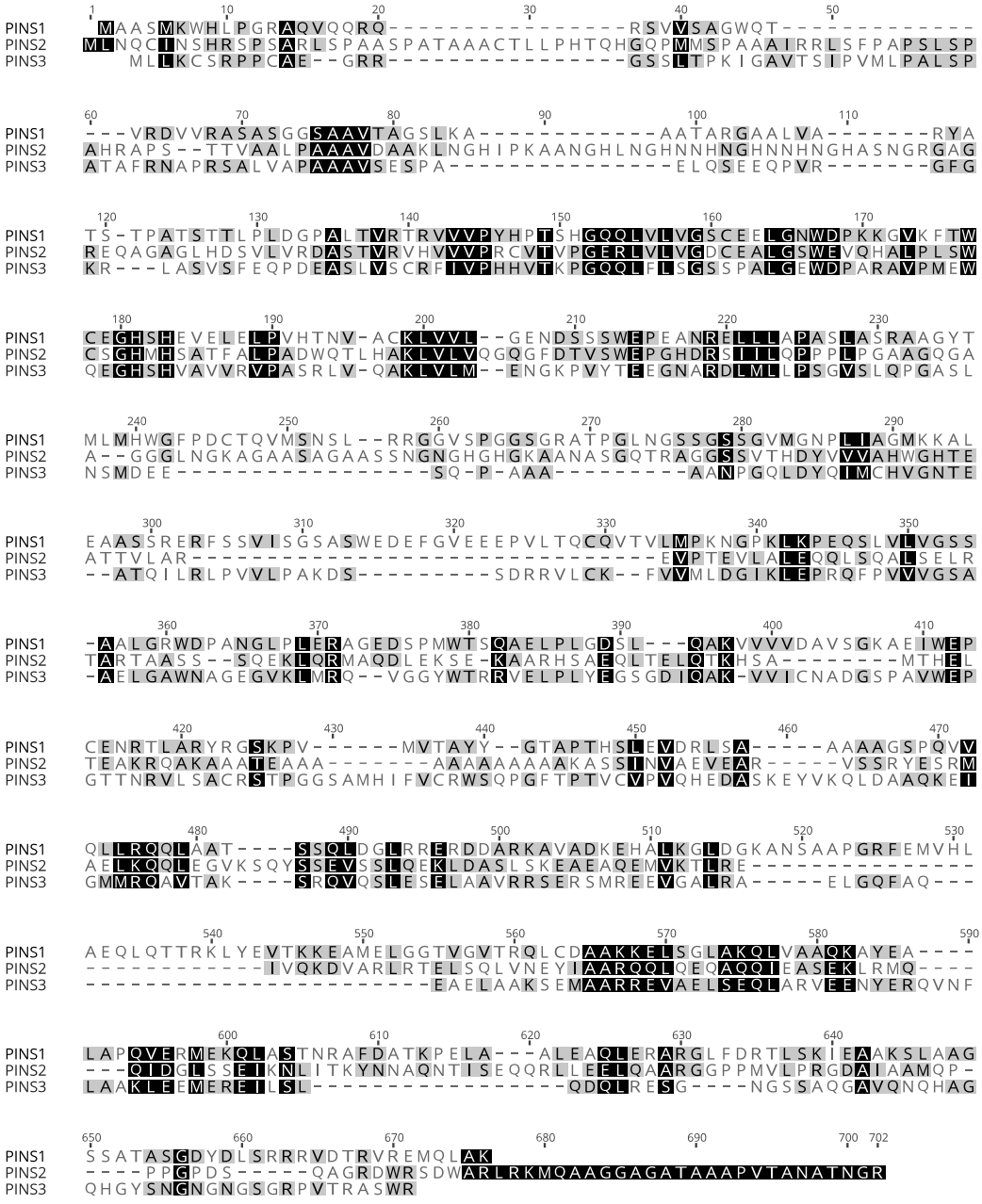
L.C.M.M. guided and supervised the project; C.S.L. designed and performed the CRISPR mutant generation, growth phenotyping. C.S.L. and L.A. performed the transmission electron microscopy experiment. C.S.L. and S.G. performed the fluorescent protein tagging and confocal imaging. C.S.L. analysed the data and wrote the manuscript.

**Supplemental Figures**

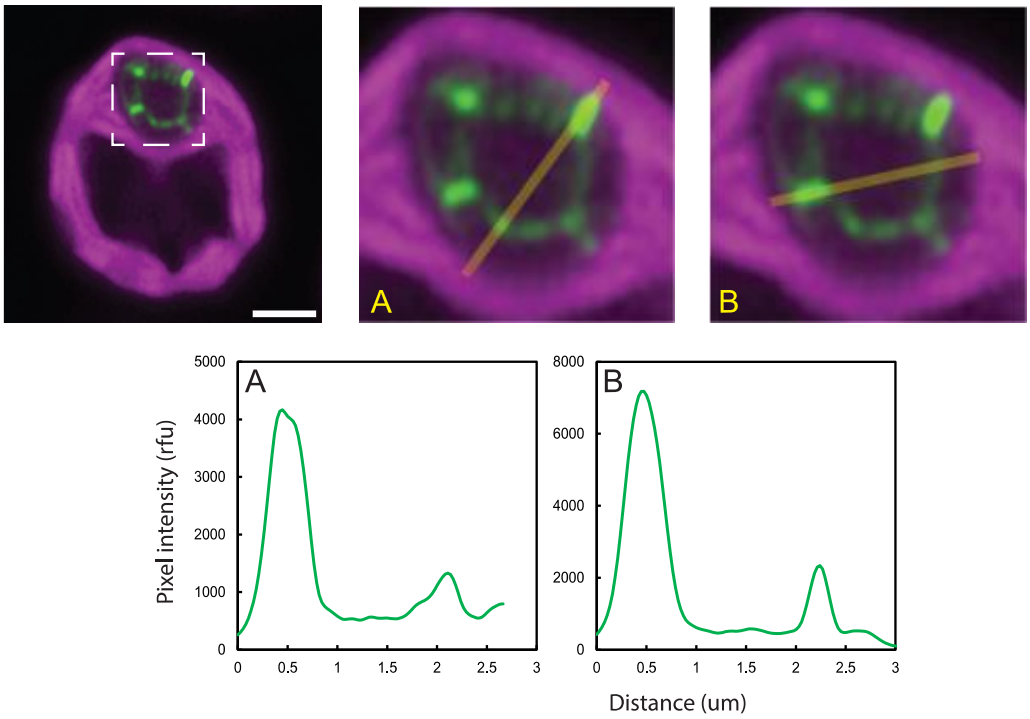



**Supplemental Figure S1.** Coiled-coil domain prediction of uncharacterised starch-related proteins. (Supports Figure 1) Protein sequence of uncharacterised starch-related proteins highlighted in Supplemental Data Set 1 were submitted to the Pcoil webserver (Gabler et al., 2020; Gruber et al., 2006) for coiled-coil probability predictions. Here the probability is plotted on the residue position of each protein sequence, the three residue-scanning windows are displayed in purple (14-residue), green (21-residue) and grey (28-residue). PINS proteins can be seen to display a consistent pattern of coiled-coil domain on its C-terminal, while the other uncharacterised starch-related protein, except

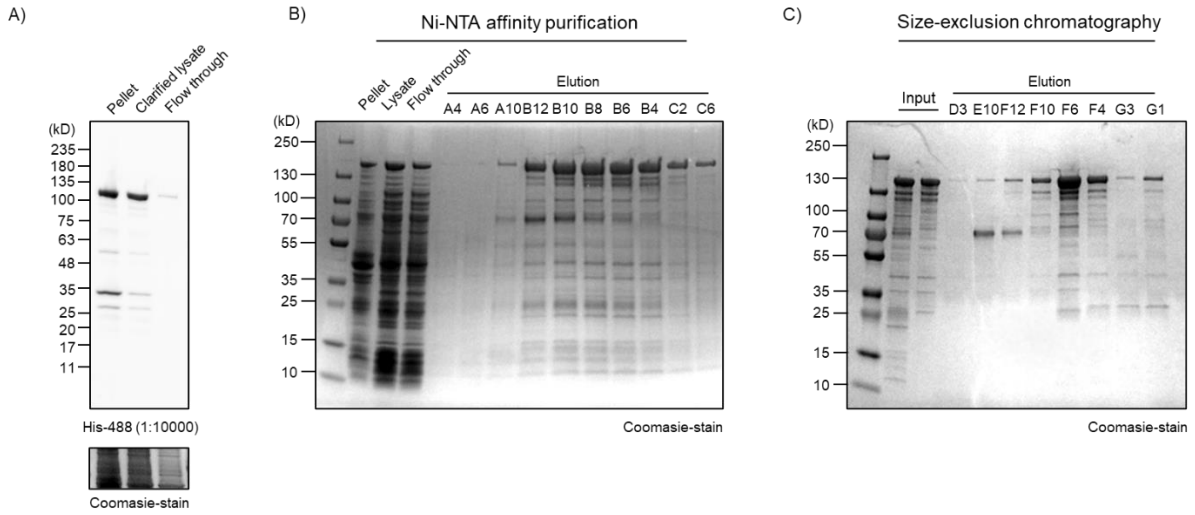
for Cre09.g415600 and Cre12.g492750 carries a coiled-coil domain in the central position, do not exhibit such pattern. Particularly, Cre12.g492750 is predicted to possess multiple short coiled-coil domain of ~20-100 residues distributed along its sequence.



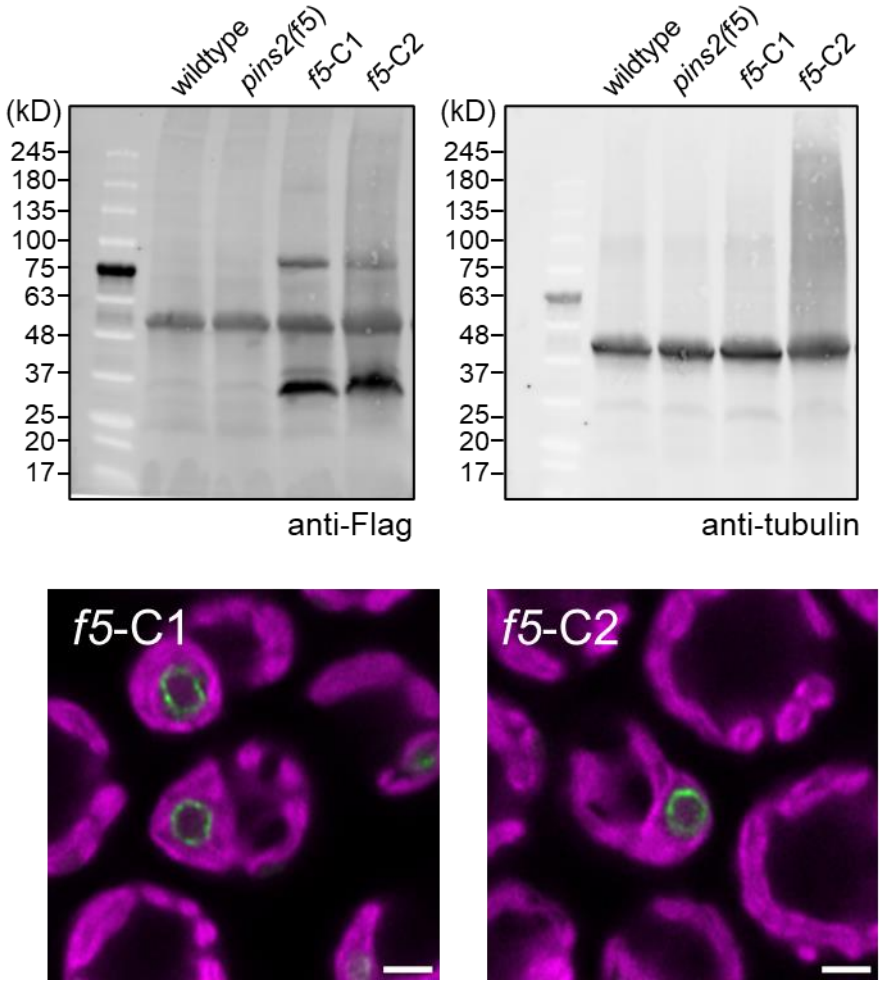

**Supplemental Figure S2**. Multiple sequence alignment of the PINS1 – 3. (Supports Figure 1) Clustal Omega 1.2.2 sequence alignment of the three proteins are performed with Geneious Prime. Hyphen denotes open gap; Grey background denotes agreement between any two sequences; Black background denotes identical residue across all three sequences.



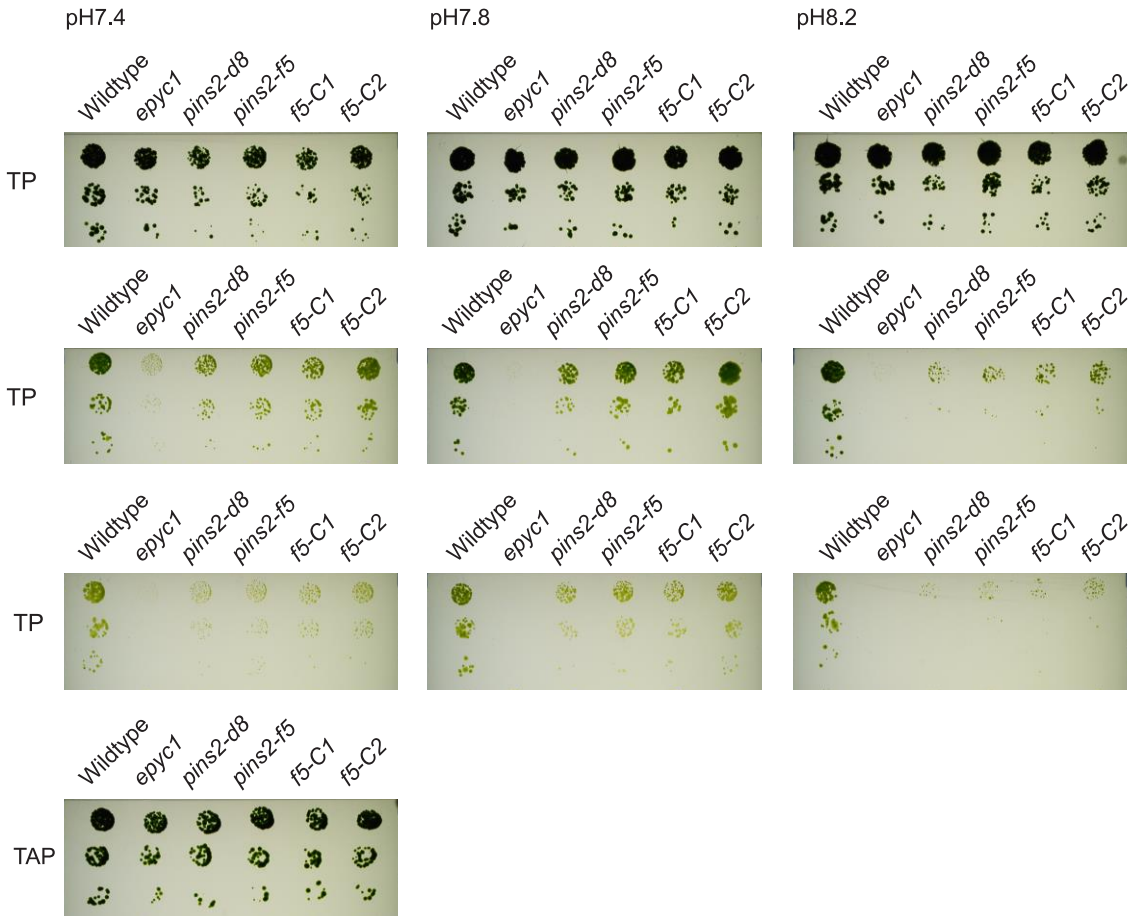

**Supplemental Figure S3.** The line plot profile of PINS2-Venus expressed in wildtype lines. (Supports Figure 2) Two lines (A and B) are drawn across the starch sheath edge to the pyrenoid facing side of starch sheath to compare the PINS2 signal from the two sites. The maximum value of the two peaks is noted on the line plot.



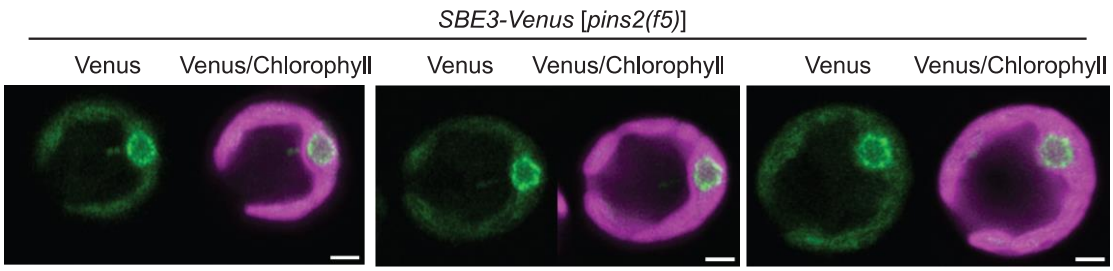

**Supplemental Figure S4.** Purification of the Cas9 protein for CRISPR-Knockout experiments. (Supports Figure 3) **A)** The chemically competent E. coli BL21 Rosetta (DE3) line was transformed with pET27b-Cas9-His. E. coli line expressing the Cas9 protein was incubated with 0.4 mM IPTG at 18°C for 20 hours to induce Cas9 expression. At the end of the induction period, cells were pelleted by centrifugation and lysed by sonication. The pellet and clarified lysate were probed against anti-His antibody. Cas9 protein (~158 kD) present at the clarified lysate lane suggests soluble Cas9 protein were expressed. **B)** Nickel affinity purification of the Cas9 protein, fractions which exhibit high UV absorbance was analysed with SDS-PAGE. The strong signal at 160 kD size likely represents the Cas9 protein. The eluted fractions (B12 – C2) was shown to contain the most Cas9 protein and was subsequently collected for further purification. **C)** Gel filtration of the Cas9 protein, fractions containing Cas9 protein collected from Ni-NTA purification were subjected to further purification via size-exclusion chromatography.



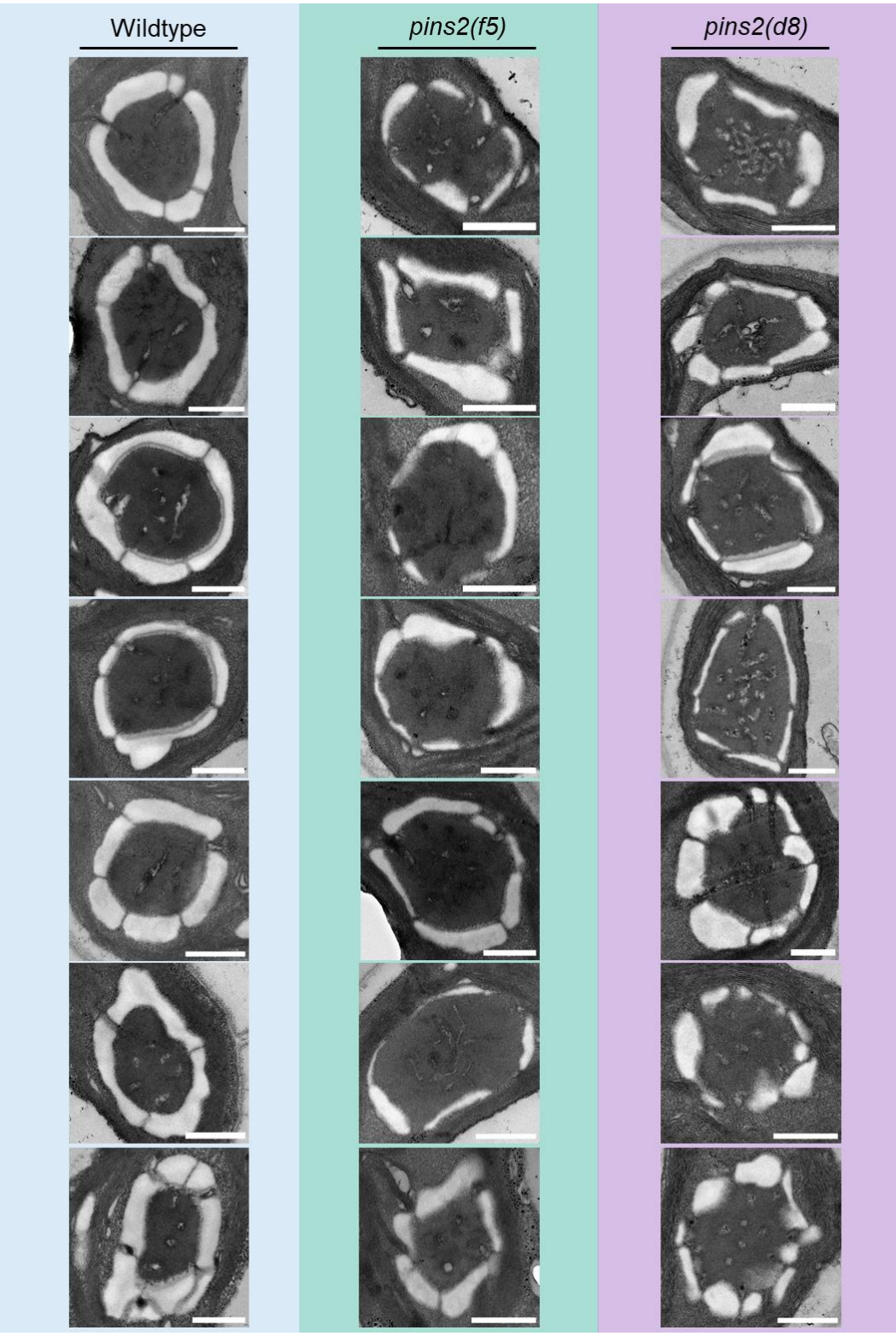

**Supplemental Figure S5.** Complementation of PINS2-Knockout line F5. (Supports Figure 3) The *pins2-KO(f5)* has been transformed with pLM547. The protein extract from the two complemented line *f5-C1* (strain9) and *f5-C2* (strain 11) and its mutant background are blotted against Flag-anitbody, and tubulin was used as a loading control. The lower panel shows the localization of PINS2-Venus that is confirmed through confocal imaging, it shows that the mesh-like localisation of PINS2 is retained in the complemented line.



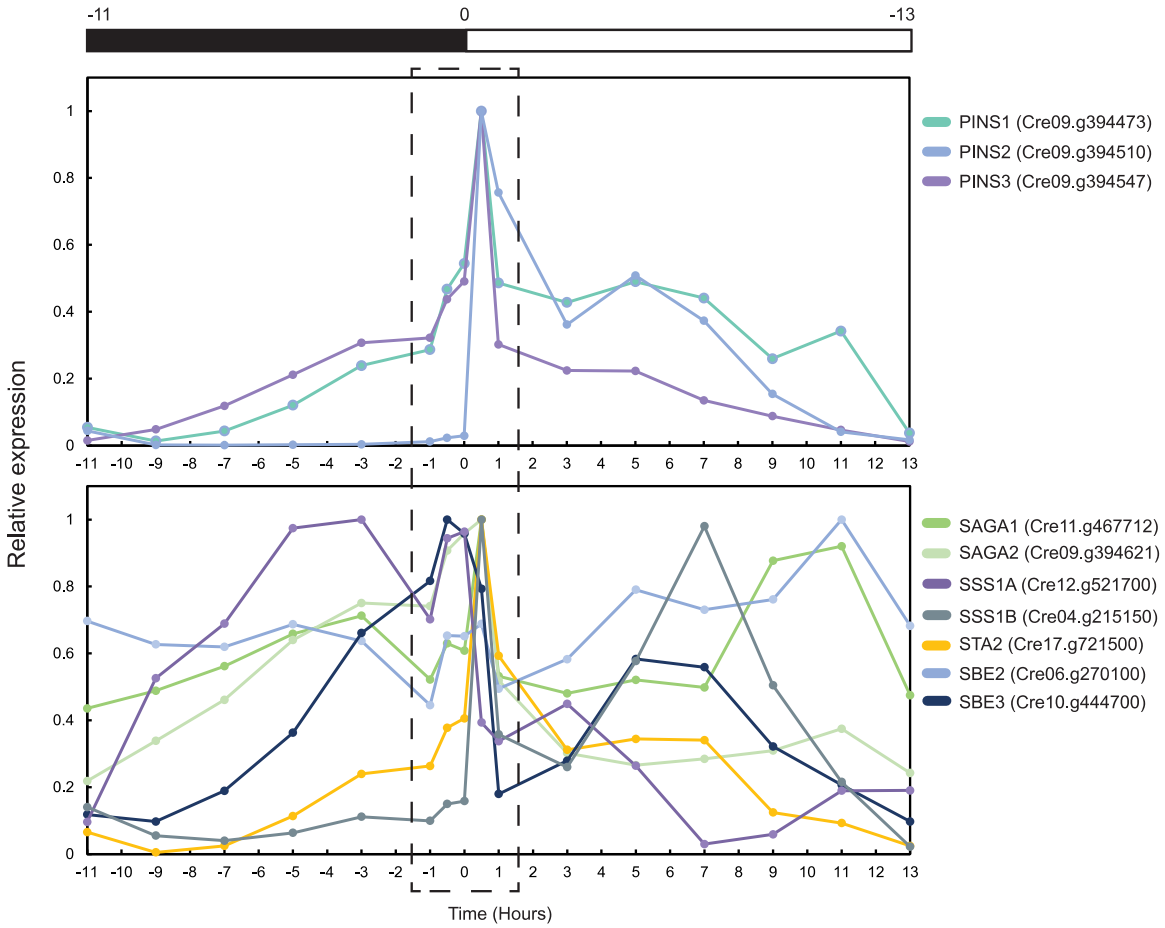

**Supplemental Figure S6.** images of the full plate used to assemble Figure 3. (Supports Figure 3) Only the top half of the plates is shown here as the lower half represents other unrelated mutants that were screened in the same plate. The spots representing 1000 cells were cropped and used to assemble Figure 2. For growth assay, agar plates were incubated in gas-tight chambers supplied with the defined $CO_2$ concentrations (High – 3%, Low – 0.04%, Very low – 0.01%) in 400 µmols photons $m^{-2}$ $s^{-1}$. Colonies are grown on TAP plates for loading control, these plates were incubated on shelfs with ~10 µmols photons $m^{-2}$ $s^{-1}$.




**Supplemental Figure S7.** Localisation of the starch synthesis protein SBE3 in *pins2(f5)* mutant. (Supports Figure 2) pLM019 (SBE3-Venus; Mackinder et la., 2017) was transformed into *pins2(f5)*. Cells are cultivated in photoautotrophic conditions prior to imaging, the SBE3 signal is found largely at pyrenoid location reminiscent of a uniform starch sheath localisation which aligns with previous report. The green and magenta signals represent the Venus and Chlorophyll Channel, respectively. Scale bar = 2 µm.




**Supplemental Figure S8.** Additional TEM images of Wildtype, *pins2(f5)* and *pins2(d8)*. (Supports Figure 4) Scale bar represents 500 nm.




**Supplemental Figure S9.** Expression level of PINS1 – 3. Comparison of the expression profiles between PINS1 – 3 (top panel) and starch synthesis proteins (bottom panel) is shown here. The plotted data were reproduced from (Strenkert et al. 2019) where the expression profiles are obtained throughout a day-night cycle, represented by the black-and-white bar on the top panel. The white colour denotes day and black colour denotes night. The relative expression was calculated by normalising the genes expression level against its corresponding detected maximum. The dashed rectangle highlights the time period where all three PINS proteins are maximally expressed. This time period corresponds well to that of the starch synthesis proteins (SSS1A, SSS1B, STA2, SBE2 and SBE3) and starch-related pyrenoid proteins (SAGA1 and SAGA2).

**Supplemental Dataset S1.** A list of annotated starch-related genes on the Chlamydomonas genome. Using key words “CBM” and “Starch”, a list of starch-related genes are curated below using the Chlamydomonas genome V6.1 on Phytozome (JGI portal). The corresponding gene symbol and Arabidopsis homolog are taken from the gene annotation file. The Carbohydrate binding module (if present) is highlighted on the CBM column. Comments on the proposed gene function are taken from the gene description on the Phytozome portal. For uncharacterised genes (Asterik), analysis on their annotated domains on the gene sequence is used if no description has been provided for this gene.

| Locus | Gene symbol | Gene name | Arabidopsis homolog | CBM | Comments |
|---|---|---|---|---|---|
| *Cre03.g144707 | - | - | AT1G70820 | | Phosphoglucomutase, drives the interconversion between Glucose-6-Phosphate and Glucose-1-Phosphate |
| *Cre03.g158050 | - | - | AT4G33500 | CBM20 | Phosphoprotein phosphatase 2C-related |
| *Cre03.g168350 | - | - | AT5G11010 | - | P-loop containing nucleoside triphosphate hydrolase, predicted to participate in glycogen synthesis pathway |
| *Cre03.g183300 | - | - | - | CBM20 | Contain CBM20 domain at the C-terminal and no predicted catalytic domains |
| *Cre04.g226800 | - | - | - | CBM20 | Contain CBM20 domain at the middle section and no predicted catalytic domains |
| *Cre06.g267050 | - | - | - | CBM21 | Contain CBM21 domain at the C-terminal and no predicted catalytic domains |
| *Cre06.g269601 | - | - | | CBM20 | Contain CBM20 domain at the N-terminal and no predicted catalytic domains |
| *Cre06.g269650 | - | - | - | CBM20 | Contain CBM20 domain at the N-terminal and a predicted C-terminal SAP domain which mediates DNA/RNA interactions |
| *Cre06.g800615 | - | - | - | - | Contain multiple starch synthase catalytic domain, likely an uncharacterised starch synthase protein |
| *Cre09.g407501 | - | - | | CBM20 | Contain CBM20 domain at the N-terminal and no predicted catalytic domains |
| *Cre09.g415600 | - | - | - | CBM20 | Contain CBM20 domain at the N-terminal and no predicted catalytic domains |
| *Cre10.g450500 | - | - | - | CBM20 | Contain CBM20 domain at the N-terminal and short anti-freeze protein motif, no predicted catalytic domains |
| *Cre12.g492750 | - | - | - | CBM20 | Contain CBM20 domain at the N-terminal and short anti-freeze protein motif, no predicted catalytic domains |
| *Cre12.g540850 | - | - | - | CBM10 | Contain short CBM10 domain interspaced through sequence, no predicted catalytic domains |
| *Cre12.g551200 | - | - | - | CBM20 | Contain CBM20 domain at the N-terminal and short anti-freeze protein motif, no predicted catalytic domains |
| *Cre13.g587600 | - | - | - | CBM25 | Annotated with Alpha/beta hydrolase domain. |
| *Cre13.g801533 | - | - | - | | Contains a glycosyl transferase domain |
| *Cre13.g801553 | - | - | - | CBM25 | Annotated with Alpha/beta hydrolase domain. |
| *Cre17.g703000 | - | - | - | CBM20 | Contain CBM20 domain at the N-terminal and no predicted catalytic domains |
| Cre13.g567950 | AGP1 | ADP-glucose pyrophosphorylase large subunit | AT5G19220 | - | ADP-glucose pyrophosphorylase large subunit 1# Glucose-1-phosphate adenylyltransferase / ADP-glucose synthase# regulatory subunit catalyzes the formation of the glucosyl nucleotide from ATP and glucose-1-phosphate# contains nucleotidyl transferase domain# involved in starch metabolism |

| | | | | | |
|---|---|---|---|---|---|
| Cre16.g683450 | AGP2 | ADP-glucose pyrophosphorylase large subunit | - | - | ADP-glucose pyrophosphorylase large subunit 2# Glucose-1-phosphate adenylyltransferase / ADP-glucose synthase# regulatory subunit catalyzes the formation of the glucosyl nucleotide from ATP and glucose-1-phosphate# contains nucleotidyl transferase domain# involved in starch metabolism |
| Cre07.g331300 | AGP3 | ADP-glucose pyrophosphorylase large subunit | - | - | ADP-glucose pyrophosphorylase large subunit 3# Glucose-1-phosphate adenylyltransferase / ADP-glucose synthase# regulatory subunit catalyzes the formation of the glucosyl nucleotide from ATP and glucose-1-phosphate# contains nucleotidyl transferase domain# involved in starch metabolism |
| Cre03.g188250 | AGP4 | ADP-glucose pyrophosphorylase small subunit | AT5G48300.1 | - | ADP-glucose pyrophosphorylase large subunit 4# Glucose-1-phosphate adenylyltransferase / ADP-glucose synthase# regulatory subunit catalyzes the formation of the glucosyl nucleotide from ATP and glucose-1-phosphate# contains nucleotidyl transferase domain# involved in starch metabolism |
| Cre08.g385500 | AMA1 | Alpha-amylase 1 | - | - | Belongs to glycoside hydrolase family 13; catalyses hydrolysis of (1-4)-alpha-D-glucosidic linkages in polysaccharides so as to remove successive alpha-maltose residues from the non-reducing ends of the chains in the conversion of starch to maltose, arabidopsis are named AMY |
| Cre08.g362450 | AMA2 | Alpha-amylase 2 | - | - | Belongs to glycoside hydrolase family 13; catalyses hydrolysis of (1-4)-alpha-D-glucosidic linkages in polysaccharides so as to remove successive alpha-maltose residues from the non-reducing ends of the chains in the conversion of starch to maltose |
| Cre08.g384750 | AMA3 | Alpha-amylase 3 | - | - | Belongs to glycoside hydrolase family 13; catalyses hydrolysis of (1-4)-alpha-D-glucosidic linkages in polysaccharides so as to remove successive alpha-maltose residues from the non-reducing ends of the chains in the conversion of starch to maltose |
| Cre06.g307150 | AMB1 | Beta-amylase 1 | AT3G23920 | - | Belongs to glycoside hydrolase family 14# alpha 1,4-glucan maltohydrolase (EC 3.2.1.2), hydrolyzes the alpha 1,4-glucosidic linkages of starch from the nonreducing ends# |
| Cre06.g270350 | AMB2 | Beta-amylase 2 | - | - | Belongs to glycoside hydrolase family 14# alpha 1,4-glucan maltohydrolase (EC 3.2.1.2), hydrolyzes the alpha 1,4-glucosidic linkages of starch from the nonreducing ends.# null-allele mutant was isolated (PMID 29743196) |
| Cre01.g044100 | AMB3 | Beta-amylase 3 | - | - | Belongs to glycoside hydrolase family 14# alpha 1,4-glucan maltohydrolase (EC 3.2.1.2), hydrolyzes the alpha 1,4-glucosidic linkages of starch from the nonreducing ends. Arabidopsis genes use BAM root# |
| Cre02.g091750 | BSG1 | Bimodal Starch granule 1 | - | CBM20 | - |
| Cre03.g181500 | DPE1 | Disproportionating enzyme 1 | AT5G64860 | - | alpha-1,4-glucanotransferase# transfers mostly maltosyl residues from a donor oligosaccharide to the nonreducing end of an acceptor. Described by Wattebled et al. 23 [PMID:12746519]. Known formally as disproportionating enzyme (D-enzyme)# Identified as STA11 in a screen for starchless mutants# |
| Cre02.g095126 | DPE2 | Disproportionating enzyme 2 | AT2G40840 | CBM20 | Alpha-1,4-glucanotransferase# Transfers mostly maltosyl residues from a donor oligosaccharide to the nonreducing end of an acceptor# probable cytosolic isoform# Known formally as disproportionating enzyme (D-enzyme)# |
| Cre07.g337300 | DYRKP1 | Dual-Specificity Tyrosine-Regulated Protein Kinase involved in starch degradation | AT1G73460 | - | Acts as a negative regulator of the sink capacity of photosynthetic cells |
| Cre06.g257050 | GBSS1B | Granule bound starch synthase IB | - | - | Granule bound starch synthase# GBSS-like protein# catalyses the transfer of a glucosyl moiety of the ADP-glc to the non-reducing end of a pre-existing alpha 1-4 linked chain. This sequence matches strongly the GBSSI sequences from higher plant. But it is highly different from the Chlamydomonas GBSSIA (STA2 gene), and there is no C-terminal extension. |

| | | | | | |
|---|---|---|---|---|---|
| Cre07.g319300 | GWD1 | Alpha-glucan water dikinase 1 | AT1G10760 | - | R1 Protein# Phosphorylates the C-3 and C-6 positions of alpha glucans with the beta phosphate of ATP# |
| Cre07.g332300 | GWD2 | Alpha-glucan water dikinase 2 | - | - | R1 protein# phosphorylates the C-3 and C-6 positions of alpha glucans with the beta phosphate of ATP# |
| Cre03.g155001 | ISA1 | Isoamylase, starch debranching enzyme 1 | AT2G39930 | CBM48 | Misplaced, loosely branched chains are trimmed by this DBE from a maturing structure to allow polysaccharide crystallization/ Posewitz et al. 24 [PMID: 1526933] showed that the STA7 locus corresponds to this gene. Mutants at this locus are starch-deficient# |
| Cre17.g698850 | ISA2 | Isoamylase-type starch debranching enzyme 2 | - | | Isoamylase-type starch debranching enzyme, isoform 2# Identified in a screen for starchless mutants# |
| Cre03.g207713 | ISA3 | Isoamylase, starch debranching enzyme 3 | AT4G09020 | CBM48 | Starch debranching enzyme# Misplaced, loosely branched chains are trimmed by this DBE from a maturing structure to allow polysaccharide crystallization# |
| Cre03.g175400 | PGI1 | Phosphoglucose isomerase 1 | AT5G42740 | - | PGI, phosphohexose isomerase, glucose-6-phosphate isomerase [EC:5.3.1.9]# belongs to sugar isomerase (SIS) family# ts-lethal mutant was isolated |
| Cre06.g278210 | PGM1 | Phosphoglucomutase 1 | AT5G51820 | - | Converts D-glucose 1-phosphate into D-glucose 6-phosphate, and participates in both the breakdown and synthesis of glucose# the reaction mechanism involves phosphoryl transfer from a phosphoserine to the substrate to create a biophosphorylated sugar, followed by a phosphoryl transfer from the substrate back to the enzyme# Identified in a screen for starchless mutants# Note that PGM = phosphogluco mutase and GPM = phosphoglycerate mutase# |
| Cre01.g012600 | PGM2 | Phosphoglucomutase 2 | - | - | Converts D-glucose 1-phosphate into D-glucose 6-phosphate, and participates in both the breakdown and synthesis of glucose# the reaction mechanism involves phosphoryl transfer from a phosphoserine to the substrate to create a biophosphorylated sugar, followed by a phosphoryl transfer from the substrate back to the enzyme# Note that PGM = phosphogluco mutase and GPM = phosphoglycerate mutase# |
| Cre10.g422450 | PGSIP6 | Plant glycogenin-like starch initiation protein 6 | AT5G18480 | - | - |
| Cre07.g336950 | PHOA2 | Starch phosphorylase 2, PhoA | - | - | gi 400790 sp Q00766 PHS1_DICDI GLYCOGEN PHOSPHORYLASE 1 (GP1)# starch phosphorylase (EC 2.4.1.1)# GenBank: ABB88569# |
| Cre12.g552200 | PHOB1 | Starch phosphorylase PhoB | AT3G46970 | - | Identified as STA4 in a screen for starchless mutants# |
| Cre08.g363874 | PPDK | Phosphoenolpyruvate Dikinase related | - | - | - |
| Cre09.g387100 | PTST1 | Protein targeting to starch 1 | AT5G39790 | CBM48 | Conserved in the Green Lineage |
| Cre11.g476650 | PUL1 | Pullulanase-type starch debranching enzyme 1 | AT5G04360 | CBM48 | - |
| Cre17.g719900 | PWD1 | Phosphoglucan water dikinase 1 | AT5G26570 | CBM20 | Contains starch-bnding domain# involved in starch metabolism |
| Cre11.g467712 | SAGA1 | Starch granule Abnormal 1 | - | CBM20 | Contains a starch binding motif and regulate starch sheath morphology# localizes to multiple puncta and streaks in the pyrenoid and physically interacts with the small and large subunits of RuBisCO# |
| Cre09.g394621 | SAGA2 | Starch granule Abnormal 2 | - | CBM20 | - |

| | | | | | |
|---|---|---|---|---|---|
| Cre06.g289850 | SBE1 | Starch Branching Enzyme 1 | - | CBM48 | Catalyses the formation of alpha-1,6 linkages within the polymer by cleaving a preexisting linear chain and transferring the fragment at the non reducing end of the cleaved glucan to an alpha-1,6 position |
| Cre06.g270100 | SBE2 | Starch Branching Enzyme 2 | AT2G36390 | CBM48 | Catalyses the formation of alpha-1,6 linkages within the polymer by cleaving a preexisting linear chain and transferring the fragment at the non reducing end of the cleaved glucan to an alpha-1,6 position |
| Cre10.g444700 | SBE3 | Starch Branching Enzyme 3 | AT2G36390 | CBM48 | Catalyses the formation of alpha-1,6 linkages within the polymer by cleaving a preexisting linear chain and transferring the fragment at the non reducing end of the cleaved glucan to an alpha-1,6 position |
| Cre08.g373450 | SBE4 | Starch Branching Enzyme 4 | - | CBM48 | catalyses the formation of alpha-1,6 linkages within the polymer by cleaving a preexisting linear chain and transferring the fragment at the non reducing end of the cleaved glucan to an alpha-1,6 position |
| Cre12.g521700 | SSS1A | Soluble starch synthase IA | - | - | ADP-glucose alpha-1,4 glucane alpha-4-glucanotransferase |
| Cre04.g215150 | SSS1B | Soluble starch synthase IB | AT5G24300 | - | ADP-glucose alpha-1,4 glucane alpha-4-glucanotransferase |
| Cre03.g185250 | SSS2 | Soluble starch synthase II | AT3G01180 | - | ADP-glucose alpha-1,4 glucane alpha-4-glucanotransferase |
| Cre06.g282000 | SSS3A | Soluble starch synthase III | - | CBM25 | Starch/Glycogen Synthase ADPG - SS3 |
| Cre13.g579598 | SSS3B | Soluble starch synthase III | AT1G11720 | CBM25 | Soluble starch synthase with similarity to plant starch synthase III (e.g., SS III from Vigna GenBank CAB4374.1)# ADP-glucose alpha-1,4 glucan alpha-4-glucanotransferase# catalyses the transfer of aglucosyl moety of ADP-glucose to the non-reducing end of a pre-existing alpha 1,4 linked chain# |
| Cre16.g665800 | SSS4 | Soluble starch synthase IV | - | - | ADP-glucose alpha-1,4 glucane alpha-4-glucanotransferase# catalyses the transfer of the glucosyl moiety of the ADP-glc to the non-reducing end of a pre-existing alpha 1,4 linked chain# (AAC17971) |
| Cre16.g663850 | SSS5 | Soluble starch synthase V | AT4G18240 | - | - |
| Cre17.g721500 | STA2 | Granule-bound starch synthase IA | AT1G32900 | - | Identified as STA2 in a screen for starchless mutants# |

**Supplemental Dataset S2.** Statistical test of image analysis in Figure 2B.

Two-Sample t-test Assuming Equal Variances was performed on the measured area of STA2-Venus in the *pins2(f5)* and wildtype strain.

| | *pins2(f5)* | Wildtype |
|---|---|---|
| Mean | 0.034215 | 0.041636 |
| Variance | 5.36E-05 | 0.000152 |
| Observations | 29 | 31 |
| Pooled Variance | 0.000105 | |
| Hypothesized Mean Difference | 0 | |
| Df | 58 | |
| t Stat | -2.80775 | |
| P(T<=t) one-tail | 0.003393 | |
| t Critical one-tail | 1.671553 | |
| P(T<=t) two-tail | 0.006785 | |
| t Critical two-tail | 2.001717 | |

**Supplemental Dataset S3.** Oligo Sequence used.

| Oligo name | Sequence | Description |
|---|---|---|
| oLM3373 | GAAATTAATACGACTCACTATAG CGTCACGCACGACTATGTGGGT TTTAGAGCTAGAAATAGCAAG | Forward primer to generate DNA template for *in vitro* transcription |
| oLM3118 | AAAAAAGCACCGACTCGGTGCC ACTTTTTCAAGTTGATAACGGAC TAGCCTTATTTTAACTTGCTATTT CTAGCTCTAAAAC | Reverse primer to generate DNA template for *in vitro* transcription |
| oLM3381 | AAGGCTGCTAATGCGTCG | Forward primers to check *PINS2* editing used in mutants screening |
| oLM3382 | ATACCATGGATGCCAAGCTC | Reverse primers check *PINS2* editing used in mutants screening |
| oLM3483 | AATGGTGCGAATGGGTGAATG | Alternative forward primers to check *PINS2* editing used in junction amplification |
| oLM1666 | GTTTGCGGGTTGTGACTGAAAC | Forward internal primer at hygromycin resistance cassette |
| oLM2020 | GGAAGCGGACCGAGGACTTC | Reverse internal primer at hygromycin resistance cassette |

## Reference


**Armbruster U, Labs M, Pribil M, Viola S, Xu W, Scharfenberg M, Hertle AP, Rojahn U, Jensen PE, Rappaport F, et al.** Arabidopsis CURVATURE THYLAKOID1 proteins modify thylakoid architecture by inducing membrane curvature. Plant Cell. 2013:**25**(7):2661–2678.

**Atkinson N, Mao Y, Chan KX, and McCormick AJ**. Condensation of Rubisco into a proto-pyrenoid in higher plant chloroplasts. Nat Commun. 2020:**11**(1):6303.

**Atkinson N, Stringer R, Mitchell SR, Seung D, and McCormick AJ**. SAGA1 and SAGA2 promote starch formation around proto-pyrenoids in Arabidopsis chloroplasts. bioRxiv. 2023. https://doi.org/10.1101/2023.11.25.568654

**Badger MR, Andrews TJ, Whitney SM, Ludwig M, Yellowlees DC, Leggat W, and Price GD**. The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based $CO_2$-concentrating mechanisms in algae. Can J Bot. 1998:**76**(6):1052–1071.

**Ball S, Marianne T, Dirick L, Fresnoy M, Delrue B, and Decq A**. A Chlamydomonas reinhardtii low-starch mutant is defective for 3-phosphoglycerate activation and orthophosphate inhibition of ADP-glucose pyrophosphorylase. Planta. 1991:**185**(1):17–26.

**Barrett J, Girr P, and Mackinder LCM**. Pyrenoids: CO2-fixing phase separated liquid organelles. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2021:**1868**(5):118949.

**Blaby IK, Glaesener AG, Mettler T, Fitz-Gibbon ST, Gallaher SD, Liu B, Boyle NR, Kropat J, Stitt M, Johnson S, et al.** Systems-level analysis of nitrogen starvation-induced modifications of carbon metabolism in a Chlamydomonas reinhardtii starchless mutant. Plant Cell. 2013:**25**(11):4305–4323.

**Christiansen C, Abou Hachem M, Janeček Š, Viksø-Nielsen A, Blennow A, and Svensson B**. The carbohydrate-binding module family 20 – diversity, structure, and function. FEBS J. 2009:**276**(18):5006–5029.

**Emrich-Mills TZ, Yates G, Barrett J, Girr P, Grouneva I, Lau CS, Walker CE, Kwok TK, Davey JW, Johnson MP, et al.** A recombineering pipeline to clone large and complex genes in Chlamydomonas. Plant Cell. 2021:**33**(4):1161–1181.

**Engel BD, Schaffer M, Kuhn Cuellar L, Villa E, Plitzko JM, and Baumeister W**. Native architecture of the Chlamydomonas chloroplast revealed by in situ cryo-electron tomography. Elife. 2015:**4**. https://doi.org/10.7554/elife.04889

**Fei C, Wilson AT, Mangan NM, Wingreen NS, and Jonikas MC**. Modelling the pyrenoid-based CO2-concentrating mechanism provides insights into its operating principles and a roadmap for its engineering into crops. Nat Plants. 2022:**8**(5):583–595.

**Field CB, Behrenfeld MJ, Randerson JT, and Falkowski P**. Primary production of the biosphere: Integrating terrestrial and oceanic components. Science. 1998:**281**(5374):237–240.

**Findinier J, Laurent S, Duchêne T, Roussel X, Lancelon-Pin C, Cuiné S, Putaux J-L, Li-Beisson Y, D'Hulst C, Wattebled F, et al.** Deletion of BSG1 in Chlamydomonas reinhardtii leads to abnormal starch granule size and morphology. Sci Rep. 2019:**9**(1):1–13.

**Freeman Rosenzweig ES, Xu B, Kuhn Cuellar L, Martinez-Sanchez A, Schaffer M, Strauss M, Cartwright HN, Ronceray P, Plitzko JM, Förster F, et al.** The Eukaryotic CO2-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization. Cell. 2017:**171**(1):148-162.e19.

**Gabler F, Nam S-Z, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, and Alva V**. Protein sequence analysis using the MPI Bioinformatics Toolkit. Curr Protoc Bioinformatics. 2020:**72**(1). https://doi.org/10.1002/cpbi.108

**Gruber M, Söding J, and Lupas AN**. Comparative analysis of coiled-coil prediction methods. J Struct Biol. 2006:**155**(2):140–145.

**Harvey SH, Krien MJE, and O'Connell MJ**. Structural maintenance of chromosomes (SMC) proteins, a family of conserved ATPases. Genome Biol. 2002:**3**(2):REVIEWS3003.

**He S, Chou H-T, Matthies D, Wunder T, Meyer MT, Atkinson N, Martinez-Sanchez A, Jeffrey PD, Port SA, Patena W, et al.** The structural basis of Rubisco phase separation in the pyrenoid. Nat Plants. 2020:**6**(12):1480–1490.

**Hedin N, Velazquez MB, Barchiesi J, Gomez-Casati DF, and Busi MV**. CBM20CP, a novel functional protein of starch metabolism in green algae. Plant Mol Biol. 2022:**108**(4–5):363–378.

**Itakura AK, Chan KX, Atkinson N, Pallesen L, Wang L, Reeves G, Patena W, Caspari O, Roth R, Goodenough U, et al.** A Rubisco-binding protein is required for normal pyrenoid number and starch sheath morphology in *Chlamydomonas reinhardtii*. Proc Natl Acad Sci U S A. 2019:**116**(37):18445–18454.

**Izumo A, Fujiwara S, Sakurai T, Ball SG, Ishii Y, Ono H, Yoshida M, Fujita N, Nakamura Y, Buléon A, et al.** Effects of granule-bound starch synthase I-defective mutation on the morphology and structure of pyrenoidal starch in Chlamydomonas. Plant Sci. 2011:**180**(2):238–245.

**Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al.** Highly accurate protein structure prediction with AlphaFold. Nature. 2021:**596**(7873):583–589.

**Kanehisa M**. Linking databases and organisms: GenomeNet resources in Japan. Trends Biochem Sci. 1997:**22**(11):442–444.

**Kasili RW, Rai AK, and Moroney JV**. LCIB functions as a carbonic anhydrase: evidence from yeast and Arabidopsis carbonic anhydrase knockout mutants. Photosynth Res. 2023. https://doi.org/10.1007/s11120-023-01005-1

**Küken A, Sommer F, Yaneva-Roder L, Mackinder LC, Höhne M, Geimer S, Jonikas MC, Schroda M, Stitt M, Nikoloski Z, et al.** Effects of microcompartmentation on flux distribution and metabolic pools in Chlamydomonas reinhardtii chloroplasts. Elife. 2018:**7**. https://doi.org/10.7554/eLife.37960

**Labun K, Montague TG, Krause M, Torres Cleuren YN, Tjeldnes H, and Valen E**. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Res. 2019:**47**(W1):W171–W174.

**Lau CS, Dowle A, Thomas GH, Girr P, and Mackinder LCM**. A phase-separated CO2-fixing pyrenoid proteome determined by TurboID in *Chlamydomonas reinhardtii*. Plant Cell. 2023:**35**(9):3260–3279.

**Li XE, Holmes KC, Lehman W, Jung H, and Fischer S**. The shape and flexibility of tropomyosin coiled coils: implications for actin filament assembly and regulation. J Mol Biol. 2010:**395**(2):327–339.

**Mackinder LCM, Chen C, Leib RD, Patena W, Blum SR, Rodman M, Ramundo S, Adams CM, and Jonikas MC**. A Spatial Interactome Reveals the Protein Organization of the Algal CO2-Concentrating Mechanism. Cell. 2017:**171**(1):133-147.e14.

**Mackinder LCM, Meyer MT, Mettler-Altmann T, Chen VK, Mitchell MC, Caspari O, Freeman Rosenzweig ES, Pallesen L, Reeves G, Itakura A, et al.** A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. Proc Natl Acad Sci U S A. 2016:**113**(21):5958–5963.

**Majeran W, Olive J, Drapier D, Vallon O, and Wollman F-A**. The Light Sensitivity of ATP Synthase Mutants of*Chlamydomonas reinhardtii*. Plant Physiol. 2001:**126**(1):421–433.

**Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, and Ferrin TE**. UCSF ChimeraX: Tools for structure building and analysis. Protein Sci. 2023:**32**(11). https://doi.org/10.1002/pro.4792

**Meyer MT, Itakura AK, Patena W, Wang L, He S, Emrich-Mills T, Lau CS, Yates G, Mackinder LCM, and Jonikas MC**. Assembly of the algal $CO_2$-fixing organelle, the pyrenoid, is guided by a Rubisco-binding motif. Sci Adv. 2020:**6**(46). https://doi.org/10.1126/sciadv.abd2408

**Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, and Steinegger M**. ColabFold: making protein folding accessible to all. Nat Methods. 2022:**19**(6):679–682.

**Miura K**. Bleach correction ImageJ plugin for compensating the photobleaching of time-lapse sequences. F1000Res. 2020:**9**(1494):1494.

**Mukherjee A, Lau CS, Walker CE, Rai AK, Prejean CI, Yates G, Emrich-Mills T, Lemoine SG, Vinyard DJ, Mackinder LCM, et al.** Thylakoid localized bestrophin-like proteins are essential for the $CO_2$ concentrating mechanism of *Chlamydomonas reinhardtii*. Proc Natl Acad Sci U S A. 2019:**116**(34):16915–16920.

**Ngo ST, Tran-Le PD, Ho GT, Le LQ, Bui LM, Vu BK, Thu Phung HT, Nguyen H-D, Vo T-S, and Vu VV**. Interaction of carbohydrate binding module 20 with starch substrates. RSC Adv. 2019:**9**(43):24833–24842.

**Oh ZG, Ang WSL, Poh CW, Lai S-K, Sze SK, Li H-Y, Bhushan S, Wunder T, and Mueller-Cajar O**. A linker protein from a red-type pyrenoid phase separates with Rubisco via oligomerizing sticker motifs. Proc Natl Acad Sci U S A. 2023:**120**(25):e2304833120.

**Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, et al.** InterPro in 2022. Nucleic Acids Res. 2023:**51**(D1):D418–D427.

**Peng C, Wang Y, Liu F, Ren Y, Zhou K, Lv J, Zheng M, Zhao S, Zhang L, Wang C, et al.** *FLOURY ENDOSPERM6* encodes a CBM48 domain-containing protein involved in compound granule formation and starch synthesis in rice endosperm. Plant J. 2014:**77**(6):917–930.

**Ramazanov Z, Rawat M, Henk M, Mason C, Matthews S, and Moroney J**. The induction of the CO2-concentrating mechanism is correlated with the formation of the starch sheath around the pyrenoid of Chlamydomonas reinhardtii. Planta. 1994:**195**(2). https://doi.org/10.1007/bf00199681

**Seung D, Boudet J, Monroe J, Schreier TB, David LC, Abt M, Lu K-J, Zanella M, and Zeeman SC**. Homologs of PROTEIN TARGETING TO STARCH control starch granule initiation in Arabidopsis leaves. Plant Cell. 2017:**29**(7):1657–1677.

**Seung D, Schreier TB, Bürgy L, Eicke S, and Zeeman SC**. Two plastidial coiled-coil proteins are essential for normal starch granule initiation in Arabidopsis. Plant Cell. 2018:**30**(7):1523–1542.

**Seung D, Soyk S, Coiro M, Maier BA, Eicke S, and Zeeman SC**. PROTEIN TARGETING TO STARCH is required for localising GRANULE-BOUND STARCH SYNTHASE to starch granules and for normal amylose synthesis in Arabidopsis. PLoS Biol. 2015:**13**(2):e1002080.

**Shimamura D, Yamano T, Niikawa Y, Hu D, and Fukuzawa H**. A pyrenoid-localized protein SAGA1 is necessary for Ca2+-binding protein CAS-dependent expression of nuclear genes encoding inorganic carbon transporters in Chlamydomonas reinhardtii. Photosynth Res. 2023. https://doi.org/10.1007/s11120-022-00996-7

**Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, et al.** Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011:**7**(1). https://doi.org/10.1038/msb.2011.75

**Sorimachi K, Gal-Coëffet M-FL, Williamson G, Archer DB, and Williamson MP**. Solution structure of the granular starch binding domain of Aspergillus niger glucoamylase bound to β-cyclodextrin. Structure. 1997:**5**(5):647–661.

**Stein A, Weber G, Wahl MC, and Jahn R**. Helical extension of the neuronal SNARE complex into the membrane. Nature. 2009:**460**(7254):525–528.

**Strenkert D, Schmollinger S, Gallaher SD, Salomé PA, Purvine SO, Nicora CD, Mettler-Altmann T, Soubeyrand E, Weber APM, Lipton MS, et al.** Multiomics resolution of molecular events during a day in the life of Chlamydomonas. Proc Natl Acad Sci U S A. 2019:**116**(6):2374–2383.

**Toyokawa C, Yamano T, and Fukuzawa H**. Pyrenoid Starch Sheath Is Required for LCIB Localization and the CO2-Concentrating Mechanism in Green Algae. Plant Physiol. 2020:**182**(4):1883–1893.

**Villarejo A, Martinez F, del Pino Plumed M, and Ramazanov Z**. The induction of the $CO_2$ concentrating mechanism in a starch-less mutant of *Chlamydomonas reinhardtii*. Physiol Plant. 1996:**98**(4):798–802.

**Wang L, Patena W, Van Baalen KA, Xie Y, Singer ER, Gavrilenko S, Warren-Williams M, Han L, Harrigan HR, Hartz LD, et al.** A chloroplast protein atlas reveals punctate

structures and spatial organization of biosynthetic pathways. Cell. 2023:**186**(16):3499-3518.e14.

**Wang L, Yamano T, Takane S, Niikawa Y, Toyokawa C, Ozawa S-I, Tokutsu R, Takahashi Y, Minagawa J, Kanesaki Y, et al.** Chloroplast-mediated regulation of $CO_2$ -concentrating mechanism by $Ca^{2+}$ -binding protein CAS in the green alga *Chlamydomonas reinhardtii*. Proc Natl Acad Sci U S A. 2016:**113**(44):12586–12591.

**Yamano T and Fukuzawa H**. Transformation of the Model Microalga Chlamydomonas reinhardtii Without Cell-Wall Removal. Methods Mol Biol. 2020:**2050**:155–161.

**Yamano T, Miura K, and Fukuzawa H**. Expression analysis of genes associated with the induction of the carbon-concentrating mechanism in *Chlamydomonas reinhardtii*. Plant Physiol. 2008:**147**(1):340–354.

**Yamano T, Toyokawa C, Shimamura D, Matsuoka T, and Fukuzawa H**. CO2-dependent migration and relocation of LCIB, a pyrenoid-peripheral protein in *Chlamydomonas reinhardtii*. Plant Physiol. 2022:**188**(2):1081–1094.

**Yamano T, Tsujikawa T, Hatano K, Ozawa S-I, Takahashi Y, and Fukuzawa H**. Light and low-CO2-dependent LCIB-LCIC complex localization in the chloroplast supports the carbon-concentrating mechanism in Chlamydomonas reinhardtii. Plant Cell Physiol. 2010:**51**(9):1453–1468.

**Yu J, Baek K, Jin E, and Bae S**. DNA-free genome editing of Chlamydomonas reinhardtii using CRISPR and subsequent mutant analysis. Bio Protoc. 2017:**7**(11). https://doi.org/10.21769/bioprotoc.2352

**Zhong Y, Sagnelli D, Topbjerg HB, Hasler-Sheetal H, Andrzejczak OA, Hooshmand K, Gislum R, Jiang D, Møller IM, Blennow A, et al.** Expression of starch-binding factor CBM20 in barley plastids controls the number of starch granules and the level of CO2 fixation. J Exp Bot. 2020:**71**(1):234–246.

**Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, and Alva V**. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol. 2018:**430**(15):2237–2243.

# Chapter 5 – General Discussion and future perspectives

The work included within this thesis represents an effort to elucidate the pyrenoid organelle formation process through the application of a novel proteomics technique – proximity labelling (Chapter 2 and 3), combined with the targeted characterisation of the starch-related gene PINS2 (Chapter 4). Chapter 2 was inspired by the recently developed proximity labelling technique. This technique yields a proximity proteome via the conjugation of a biotin adduct on proximal proteins through the short-lived biotin radicals generated from the biotin-ligase enzyme tag (TurboID; Branon et al. 2018). The speed and spatial resolution of this *in vivo* labelling makes this an increasingly popular technique in the investigation of protein-protein interactions in diverse biological systems (Guo et al. 2023). Further, biotin group addition makes the requirement of maintaining native protein state and interactions dispensable, a pre-requisite that complicates experimental setup during APMS workflows. I therefore applied the TurboID-based proximity labelling to determine the proteomic composition of the dynamic $CO_2$-fixing pyrenoid. By tagging the two main constituent of the pyrenoid matrix, RBCS2 and EPYC1, the biotin labelling was directed towards the pyrenoid matrix. Two chloroplastically-targeted TurboID-tagged bait proteins (RPE1 and PRK1) were used as controls to compare with the pyrenoid targeted baits. This comparison yielded a high-confidence enrichment dataset, supported by 11 out of 30 of the enriched proteins being previously determined pyrenoid proteins. Further, fluorescent protein localisation of 7 proteins selected from the pilot experiment identified 6 novel proteins that were pyrenoid localised. Together they affirm the strength of the proximity labelling approach in producing a spatially specific proteome in *Chlamydomonas reinhardtii*. The analysis of the pyrenoid proxiome also revealed some interesting insights into the pyrenoid in addition to its central role in the CCM. Particularly, many RNA-related proteins (Cre10.g440050, Cre10.g435800, Cre09.g393358 and Cre13.g578650) were enriched in the broader pyrenoid proxiome, which supports a role of the pyrenoid in RNA metabolism. Supporting this, previous work has proposed that the pyrenoid might possess alternative functions in RNA metabolism and possibly chloroplast protein translation (Zhan et al. 2018). The most direct evidence of this RNA-related function lies on the discovery of chloroplastic stress granules (cpSG) within the pyrenoid matrix during high light stress condition, with the pyrenoid enrichment of oxidised RNA visualised by 8-oxoG, and the pyrenoid localisation of nuclease-domain-containing protein PNU1 (Uniacke and Zerges 2008; Zhan et al. 2015; Wang et al. 2023). Regarding the possible role in translation for the pyrenoid, nascent RbcL translation activity has previously been detected within biochemically purified pyrenoids, and that translation-related proteins such as the ABCF6 (localised in this thesis) or protein folding-related proteins CYN20, CYN7 and DEG8 have been localised within the pyrenoid (Wang et al. 2023).

Additionally, a specialised thylakoid membrane region surrounding the pyrenoid called the translational zone (T-zone) is the primary location for chloroplast translation of photosystem components (Sun et al. 2019). Together they suggest that the pyrenoid matrix might function to regulate RNA metabolism and chloroplast translation. However, as a pyrenoid-less strain generated from the replacement of Rubisco small subunit from spinach shows a photosynthetic defect that can be complemented by the addition of bicarbonate (Caspari et al. 2017). It was suggested that the absence of a pyrenoid does not impact the integrity of photosynthetic machinery (Caspari et al. 2017), and in extension the translation of photosystem components. Future work should focus on characterising the behaviour of these RNA-related proteins and stress granules under different conditions to elucidate their physiological relevance. If established, these observations will suggest that the pyrenoid matrix forms a potential multifunctional hub of the chloroplast.

Chapter 3 follows the development of the labelling pipeline by developing a modified cloning strategy in order to target long and complex genes using the recombineering cloning method (Emrich-Mills et al. 2021). I focused on targeting the two structurally critical interfaces of the pyrenoid organelle: the starch sheath gaps and thylakoid-matrix interface. The mass spectrometry results of RBMP2-TurboID shows a high level of overlap with the pyrenoid matrix proxiome determined in Chapter 2, which we attribute to the influence of RBMP2's flexible RBM-containing C-terminus that likely places the TurboID-tag within the pyrenoid matrix, leading to the labelling of matrix components. However, it was surprising that only a few proteins within the RBMP2 proxiome contain transmembrane domains despite RBMP2's membrane localisation, it is possible that the high abundance of Rubisco and surrounding matrix proteins has consumed most of the biotin radicals generated from TurboID, leading to the suboptimal labelling of membrane components. The function of pyrenoid tubules is still largely unresolved. Here we identified RNA-related Bifunctional nuclease (Cre16.g688302) in the RBMP2-proxiome which gives support that the pyrenoid tubules potentially carries a role in RNA-metabolism similar to that postulated of the pyrenoid matrix, although its physiological relevance remains to be validated. While for LCI9 (PINS1), the obtained PINS1 proxiome contains many starch-related proteins of both the starch synthesis and degradation pathways, which again affirms PINS1's association with the starch sheath. However, the enrichment of both pathways precludes us from ascribing a general synthesis or degradation role for PINS1. While one could argue that there is a stronger association with starch synthesis related proteins such as SBE3 due to the higher fold-change detected, the difference in protein abundance or the amount of surface-exposed lysine makes interpreting fold-change value as the direct readout of association strength rather inaccurate. Regardless, the resultant PINS1 proxiome implicates its function in starch sheath

biogenesis. We also saw that multiple coiled-coil domain proteins are enriched within the PINS1 proxiome. The similarity in domain structure, i.e., having both a starch-binding and coiled-coil domain, of these proteins to PINS1 lends an interesting parallel to the Arabidopsis PTST protein family that either direct starch synthases towards existing starch granules to promote amylose production or recruits short oligosaccharides to starch synthases to facilitate starch granule initiation. While Chlamydomonas already has a PTST1 homologue (Cre09.g387100), it is possible that PINS1 might function as an alternative PTST, such as the Arabidopsis atPTST2 and atPTST3, which show poor sequence conservation to atPTST1. Further work on validating PINS1's interaction with starch-related enzymes, and more importantly, how this association alters the starch-related enzyme's catalytic activity, will be crucial to gain a better understanding of how PINS1 influence starch sheath biogenesis.

**Lessons and outlook of the proximity labelling experiments**

Chapter 2 and 3 have provided a few lessons regarding the application of proximity labelling in identifying proximal proteins. While the use of well-designed controls expressed in the same organelle compartment (RPE1/PRK1-TurboID) improved the accuracy of the pyrenoid enrichment, filtering the pyrenoid proteomic composition by RBCS2-TurboID vs RPE1-TurboID or PRK1-TurboID alone still yields a proxiome that contains many cytosolic or known non-pyrenoid localised proteins. This suggests that the extensive and harsh buffer washing steps applied during the streptavidin affinity purification of biotinylated proteins is not sufficient to completely remove non-specific proteins binding to bead material or streptavidin via non-biotin driven interactions, or that nascently translated protein were able to drive biotin labelling during the targeting process. Here, I propose two modifications to the experimental approach that might help to improve the overall accuracy of the labelling pipeline. The first modification is to impose a requirement for biotinylation in at least one of the identified peptide fragments during the MS experiment. This will preclude the protein contamination that arises from non-specific binding to streptavidin or bead material, therefore leaving only proteins which are biotinylated either by TurboID or endogenous mechanisms to be identified in the subsequent MS analysis. However, this modification likely comes at a loss of true positive protein identification, as the proportion of proteins that are biotinylated by TurboID varies, and that trypsin-digested biotinylated peptides might not be easily detected in the MS experiment compared to their non-biotinylated counterpart. Indeed, a recent study coupled biotinylation site identification with peptide-level biotin affinity enrichment in their proximity labelling workflow (Shin et al. 2024). They showed that this method effectively minimises the discovery of false-positive proteins, while retaining most of the true positive proteins, thereby improving the over reliability of the obtained proteome

dataset when compared to the conventional approach. Secondly, as the labelling of true proximal proteins correlates with the increase in biotin incubation time, comparison of protein enrichment following increased incubation time will be an effective strategy to identify true proximal proteins that are labelled by the bait-TurboID fusion protein. In Chapter 3, it is apparent that expression level of TurboID tag significantly impact the validation and quality of the generated proxiome. While the use of native promoters likely reflects a more “natural” context of protein interaction and avoids potentially perturbing endogenous cellular process, the drastic difference in protein expression between baits and control lines makes comparison of protein enrichment skewed towards favouring false negatives. It is therefore beneficial to normalise the biotinylation level of different TurboID-expressing strains. One way to achieve a relative comparable biotinylation level is through the split-tag system where TurboID is split between two low-affinity halves (Cho et al. 2020) that contains a rapamycin-inducible interacting protein pairs FRB and FKBP (Han et al. 2019). A background strain is first generated by expressing one half of the split-TurboID with either FRB or FKBP at the desired cellular compartment driven by a medium-strength promoter. A subsequent transformation of this background strain with a plasmid encoding the bait protein tagged with the complementary split-TurboID will then ensure a stable amount of biotinylation provided that bait-TurboID fusion protein expression level is above that of the background strain. Together, this will allow different bait-TurboID construct to drive a similar level of biotinylation, and therefore facilitates a fair comparison of protein enrichment between baits and controls.

Lastly, a previous report that combines both APMS and BioID-based proximity labelling together (Liu et al. 2018) to generate the molecular context of subcellular compartments shows that both APMS and BioID yields protein enrichment with a sizeable overlap. However, there remain proteins that are only captured exclusively via AP-MS or BioID. It is therefore clear that a complementary strategy, which uses both techniques will be necessary to gain a holistic view of the local protein environment.

**PINS protein family involvement in starch sheath formation**

In Chapter 4, I identified and characterised PINS2 and showed that it forms a three-gene family with PINS1 and PINS3, all which share a similar domain arrangement. The dual localisation of PINS1 and PINS2 demonstrates that the two proteins partially co-localise but not completely, with localisation differences along the starch sheath gap. This is likely due to the different property of their coiled-coil domain, in which PINS2 possesses an additional ~100 aa N-terminal extension of the coiled-coil domain. In addition, Cre09.g415600 similarly shows a CBM20 domain and a short helical domain, and instead is localised homogenously

surrounding the pyrenoid (Wang et al. 2023)**,** showing that the coiled-coil domain in the PINS proteins likely drives their gap-localisation. Through the analysis of TEM and starch-marker confocal microscopy using the starch marker STA2, we discovered that the *pins2* mutant lines exhibit a high-$CO_2$ requiring growth phenotype and a starch-sheath morphology defect, where granules are smaller and often overlap with each other. The $HCO_3^-$ transport into the pyrenoid is supported via BST1 – 3 situated on the thylakoid membrane peripheral to the pyrenoid (Mukherjee et al. 2019). The observed overlapping starch sheaths in *pins2* likely preclude pyrenoid thylakoid entry which limits the $HCO_3^-$ supply to pyrenoid. This can subsequently impede the CAH3-driven $CO_2$ release at the pyrenoid resulting in a defective CCM.

The *pins2* CCM efficiency is also likely impaired by the failure to correctly re-localise LCIB. While LCIB in wildtype cells surrounds the pyrenoid periphery when grown under very low $CO_2$ conditions, LCIB in *pins2* instead aggregated at the basal region of the chloroplast. This phenotype is shared with the mutants of two starch-related genes: *isa1* (4-d1) and *saga1* (Toyokawa et al. 2020; Shimamura et al. 2023), where the former mutant fails to accumulate starch and in turn possesses a thinned and fragmented starch sheath that sparsely decorate the pyrenoid (Posewitz et al. 2004; Toyokawa et al. 2020); and the latter results in formation of multiple smaller pyrenoids, each encased with an elongated starch sheath (Itakura et al. 2019). Interestingly, while *isa1* similarly shows LCIB aggregation at the basal chloroplast region, LCIB in *saga1* fails to re-localise altogether. Since the calcium-signalling related CAS1 protein is also mis-localised in the *saga1* mutant (Shimamura et al. 2023), the failure of LCIB re-localisation could also stem from a $CO_2$ signalling defect rather than that of the starch sheath. While the precise molecular mechanism of LCIB re-localisation is still unclear, these observations together argue that the starch sheath integrity is a critical component of LCIB's dynamic localisation. PINS2 possibly impacts LCIB re-localisation through two separate mechanisms. Firstly, PINS2 could either bind directly to LCIB or indirectly through an intermediate partner. This in turn recruits LCIB to the pyrenoid periphery due to PINS2's pyrenoid localisation. Alternatively, PINS2 might modify the starch surface at the starch sheath gap, thereby generating a specialised surface where LCIB is recruited. It is also possible that the defective CCM stemming from *PINS2* mutation has a knock-on effect. Thus, perturbing the specific signalling event (redox, $CO_2$, or pH level etc.) required for LCIB re-localisation resulting in their mis-localisation to the chloroplast basal region. Further work to characterise the PINS2 function and interactors will therefore be invaluable to elucidating the LCIB re-localisation mechanism.

So how does PINS protein function to remodel the starch sheath during CCM induction? Here we tentatively propose a model of pyrenoid starch sheath assembly during CCM

induction: as cells enter $CO_2$-limiting conditions, stromal starch granules are degraded which occurs likely in parallel to the synthesis of new starch granules surrounding the pyrenoid. This is evidenced by the different starch composition (crystallinity and amylose content) between stromal and pyrenoid starch, which suggests a difference in their synthesis pathway (Izumo et al. 2011; Courseaux et al. 2023). SAGA1 and SAGA2 proteins then function to attach these nascent starch granules to the pyrenoid matrix, and further drives the elongation of starch granules surrounding the pyrenoid matrix (Itakura et al. 2019). PINS2, and possibly in extension PINS1 and 3 are induced under low $CO_2$ condition. Together they coat the pyrenoid-facing side of the starch granule, limiting the starch elongation into the pyrenoid by competitively binding to the starch granule surface and excluding accessible surface for starch synthesis proteins. As the starch granule grows toward each other to encircle the pyrenoid, PINS proteins are enriched on the starch sheath gap surface which prevents the erroneous extension and fusion of starch granules that could sever the pyrenoid tubule. This can be achieved either through the competitive binding mentioned earlier, and also possibly through a physical restraint imposed by the PINS protein oligomer that stretch across the starch sheath gap. The resultant pyrenoid starch sheath is therefore decorated with the PINS protein which localises as a mesh distribution lining the starch sheath gaps.

There remain many outstanding questions regarding the precise molecular function of the PINS protein family. Particularly, whether and how do the coiled-coil domains of PINS proteins drive their recruitment to the starch sheath gap? Do PINS proteins interact with each other? Further explorations into these questions by generating double and triple knockouts, targeted mutations within the coiled-coil domain, as well as the *in vitro* experiments that examine the PINS-starch association will be crucial to better understand how these proteins work to remodel the pyrenoid starch sheath.

## Future perspectives and concluding remarks

As the global population continues to rise, mounting pressure is being put on improving the food production at a speed on par with the growing food demand. Yet the global crop yield increase has begun to show fatigue, adding to the food supply pressure (Ray et al. 2012). Together with the ecological damages in large-scale commercial farming, the imminent global food security issue clearly calls for a more sustainable approach to improve food production in lieu of arable land expansion. The use of genetic engineering to introduce desirable traits such as better drought and disease resistance, and improved photosynthesis is the emergent strategy showing promise to address this growing food security issue (Abdul Aziz et al. 2022). The engineering of an algal CCM is expected to improve the photosynthetic efficiency by up to 60%, making it a highly promising strategy within the

genetic engineering toolbox (Long et al. 2015). The first step of this engineering has been recently demonstrated by the introduction of the EPYC1 linker into *Arabidopsis thaliana* where its endogenous RBCS are replaced by the Chlamydomonas counterparts (Atkinson et al. 2020) resulting in a proto-pyrenoid that is able to retain 50% of the cell's Rubisco. This impressive feat establishes the first step of engineering the pyrenoid-based CCM into vascular plants and suggests that algal CCM components can remain functional in the different biochemical environment of vascular plants. Recent modelling work has outlined a Four-step process to engineer a pyrenoid-based CCM at the lowest energy costs which includes (1) condense Rubisco into a proto-pyrenoid around thylakoids and exclude stromal CA; (2) localise thylakoid lumen CA to the pyrenoid traversing thylakoids; (3) target $HCO_3^-$ channels to thylakoids and; (4) add a starch sheath (Fei et al. 2022). Following the engineering of a proto-pyrenoid (1), (2) and (3) largely concerns the transfer of the previously characterised $CO_2$/$HCO_3^-$ transport components including the Bestrophin-like proteins BST1 – 3 and carbonic anhydrase CAH3 which directs external $CO_2$ into the pyrenoid. However, without the addition of starch sheath as a $CO_2$ barrier (4), cells will suffer a significant cost of ATP per $CO_2$ fixed due to the leaked $CO_2$ from the pyrenoid matrix. A recent report demonstrated that the expression of SAGA1/2 within the proto-pyrenoid containing Arabidopsis line promoted appearance of pyrenoid-peripheral starch granules (Atkinson et al. 2024). The next step will be to finetune starch granule synthesis of these peripheral starch granules into the pyrenoid starch sheath-like structure to create the $CO_2$ diffusion barrier needed. In this thesis, I have established the use of TurboID-based proximity labelling as a viable method to identify possible starch sheath biogenesis proteins in Chlamydomonas. The subsequent characterisation of the novel starch-related protein PINS2 reveals its functions in maintaining starch sheath integrity. Together this work has identified the PINS proteins as potential candidates for finetuning pyrenoid starch sheath biogenesis, adding them to the expanding molecular blueprint of pyrenoid-based CCM engineering.

**Abdul Aziz M, Brini F, Rouached H, and Masmoudi K**. Genetically engineered crops for sustainably enhanced food production systems. Front Plant Sci. 2022:**13**. https://doi.org/10.3389/fpls.2022.1027828

**Atkinson N, Mao Y, Chan KX, and McCormick AJ**. Condensation of Rubisco into a proto-pyrenoid in higher plant chloroplasts. Nat Commun. 2020:**11**(1):6303.

**Atkinson N, Stringer R, Mitchell SR, Seung D, and McCormick AJ**. SAGA1 and SAGA2 promote starch formation around proto-pyrenoids in Arabidopsis chloroplasts. Proc Natl Acad Sci U S A. 2024:**121**(4). https://doi.org/10.1073/pnas.2311013121

**Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, Feldman JL, Perrimon N, and Ting AY**. Efficient proximity labeling in living cells and organisms with TurboID. Nat Biotechnol. 2018:**36**(9):880–887.

**Caspari OD, Meyer MT, Tolleter D, Wittkopp TM, Cunniffe NJ, Lawson T, Grossman AR, and Griffiths H**. Pyrenoid loss in Chlamydomonas reinhardtii causes limitations in CO2 supply, but not thylakoid operating efficiency. J Exp Bot. 2017:**68**(14):3903–3913.

**Cho KF, Branon TC, Udeshi ND, Myers SA, Carr SA, and Ting AY**. Proximity labeling in mammalian cells with TurboID and split-TurboID. Nat Protoc. 2020:**15**(12):3971–3999.

**Courseaux A, George O, Deschamps P, Bompard C, Duchêne T, and Dauvillée D**. BE3 is the major branching enzyme isoform required for amylopectin synthesis in Chlamydomonas reinhardtii. Front Plant Sci. 2023:**14**. https://doi.org/10.3389/fpls.2023.1201386

**Emrich-Mills TZ, Yates G, Barrett J, Girr P, Grouneva I, Lau CS, Walker CE, Kwok TK, Davey JW, Johnson MP, et al.** A recombineering pipeline to clone large and complex genes in Chlamydomonas. Plant Cell. 2021:**33**(4):1161–1181.

**Fei C, Wilson AT, Mangan NM, Wingreen NS, and Jonikas MC**. Modelling the pyrenoid-based CO2-concentrating mechanism provides insights into its operating principles and a roadmap for its engineering into crops. Nat Plants. 2022:**8**(5):583–595.

**Guo J, Guo S, Lu S, Gong J, Wang L, Ding L, Chen Q, Liu W.** The development of proximity labeling technology and its applications in mammals, plants, and microorganisms. Cell Commun Signal. 2023:**21**, 269. https://doi.org/10.1186/s12964-023-01310-1

**Han Y, Branon TC, Martell JD, Boassa D, Shechner D, Ellisman MH, and Ting A**. Directed evolution of split APEX2 peroxidase. ACS Chem Biol. 2019:**14**(4):619–635.

**Itakura AK, Chan KX, Atkinson N, Pallesen L, Wang L, Reeves G, Patena W, Caspari O, Roth R, Goodenough U, et al.** A Rubisco-binding protein is required for normal pyrenoid number and starch sheath morphology in *Chlamydomonas reinhardtii*. Proc Natl Acad Sci U S A. 2019:**116**(37):18445–18454.

**Izumo A, Fujiwara S, Sakurai T, Ball SG, Ishii Y, Ono H, Yoshida M, Fujita N, Nakamura Y, Buléon A, et al.** Effects of granule-bound starch synthase I-defective mutation on the morphology and structure of pyrenoidal starch in Chlamydomonas. Plant Sci. 2011:**180**(2):238–245.

**Liu X, Salokas K, Tamene F, Jiu Y, Weldatsadik RG, Öhman T, and Varjosalo M**. An AP-MS- and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations. Nat Commun. 2018:**9**(1). https://doi.org/10.1038/s41467-018-03523-2

**Long SP, Marshall-Colon A, and Zhu X-G**. Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. Cell. 2015:**161**(1):56–66.

**Mukherjee A, Lau CS, Walker CE, Rai AK, Prejean CI, Yates G, Emrich-Mills T, Lemoine SG, Vinyard DJ, Mackinder LCM, et al.** Thylakoid localized bestrophin-like proteins are essential for the $CO_2$ concentrating mechanism of *Chlamydomonas reinhardtii*. Proc Natl Acad Sci U S A. 2019:**116**(34):16915–16920.

**Posewitz MC, Smolinski SL, Kanakagiri S, Melis A, Seibert M, and Ghirardi ML**. Hydrogen photoproduction is attenuated by disruption of an isoamylase gene in *Chlamydomonas reinhardtii*. Plant Cell. 2004:**16**(8):2151–2163.

**Ray DK, Ramankutty N, Mueller ND, West PC, and Foley JA**. Recent patterns of crop yield growth and stagnation. Nat Commun. 2012:**3**(1):1–7.

**Shimamura D, Yamano T, Niikawa Y, Hu D, and Fukuzawa H**. A pyrenoid-localized protein SAGA1 is necessary for Ca2+-binding protein CAS-dependent expression of nuclear genes encoding inorganic carbon transporters in Chlamydomonas reinhardtii. Photosynth Res. 2023. https://doi.org/10.1007/s11120-022-00996-7

**Shin S, Lee S-Y, Kang M-G, Jang D-G, Kim J, Rhee H-W, and Kim J-S.** Super-resolution proximity labeling with enhanced direct identification of biotinylation sites. Commun Biol. 2024:7(1):554.

**Sun Y, Valente-Paterno M, Bakhtiari S, Law C, Zhan Y, and Zerges W**. Photosystem Biogenesis Is Localized to the Translation Zone in the Chloroplast of Chlamydomonas. Plant Cell. 2019:**31**(12):3057–3072.

**Toyokawa C, Yamano T, and Fukuzawa H**. Pyrenoid Starch Sheath Is Required for LCIB Localization and the CO2-Concentrating Mechanism in Green Algae. Plant Physiol. 2020:**182**(4):1883–1893.

**Uniacke J and Zerges W**. Stress induces the assembly of RNA granules in the chloroplast of Chlamydomonas reinhardtii. J Cell Biol. 2008:**182**(4):641–646.

**Wang L, Patena W, Van Baalen KA, Xie Y, Singer ER, Gavrilenko S, Warren-Williams M, Han L, Harrigan HR, Hartz LD, et al.** A chloroplast protein atlas reveals punctate structures and spatial organization of biosynthetic pathways. Cell. 2023:**186**(16):3499-3518.e14.

**Zhan Y, Dhaliwal JS, Adjibade P, Uniacke J, Mazroui R, and Zerges W**. Localized control of oxidized RNA. J Cell Sci. 2015:**128**(22):4210–4219.

**Zhan Y, Marchand CH, Maes A, Mauries A, Sun Y, Dhaliwal JS, Uniacke J, Arragain S, Jiang H, Gold ND, et al.** Pyrenoid functions revealed by proteomics in Chlamydomonas reinhardtii. PLoS One. 2018:**13**(2):e0185039.