nature plants

Article

https://doi.org/10.1038/s41477-024-01871-0

Hornworts reveal a spatial model for pyrenoid-based CO_2-concentrating mechanisms in land plants

Received: 19 June 2024

Accepted: 31 October 2024

Published online: 3 January 2025

Check for updates

Tanner A. Robison[1,2], Zhen Guo Oh[2], Declan Lafferty[1], Xia Xu[1], Juan Carlos A. Villarreal[3], Laura H. Gunn[2] ✉ & Fay-Wei Li[1,2] ✉

Pyrenoid-based CO_2-concentrating mechanisms (pCCMs) turbocharge photosynthesis by saturating CO_2 around Rubisco. Hornworts are the only land plants with a pCCM. Owing to their closer relationship to crops, hornworts could offer greater translational potential than the green alga *Chlamydomonas*, the traditional model for studying pCCMs. Here we report a thorough investigation of a hornwort pCCM using the emerging model *Anthoceros agrestis*. The pyrenoids in *A. agrestis* exhibit liquid-like properties similar to those in *Chlamydomonas*, but they differ by lacking starch sheaths and being enclosed by multiple thylakoids. We found that the core pCCM components in *Chlamydomonas*, including BST, LCIB and CAH3, are conserved in *A. agrestis* and probably have similar functions on the basis of their subcellular localizations. The underlying chassis for concentrating CO_2 might therefore be shared between hornworts and *Chlamydomonas*, and ancestral to land plants. Our study presents a spatial model for a pCCM in a land plant, paving the way for future biochemical and genetic investigations.

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), the carbon-fixing enzyme, is the gatekeeper for virtually all biologically available carbon. Despite its central importance in global primary productivity, Rubisco is considered to have two major limitations: a low rate of activity and poor specificity for carbon dioxide (CO_2)[1]. Photosynthetic organisms have largely overcome the first of these limitations by simply making more of the enzyme, so much so that Rubisco is considered to be the most abundant enzyme in the biosphere[2]. Rubisco's poor specificity means that it can also react with oxygen (O_2), resulting in photorespiration, which costs energy and leads to the net loss of fixed CO_2[3,4].

To reduce photorespiration, some plants have evolved systems to concentrate inorganic carbon (Ci) around Rubisco (called CO_2-concentrating mechanisms (CCMs)), through either biophysical or biochemical means[5]. Biochemical CCMs (C_4 or CAM photosynthesis) use enzymatic pathways to concentrate carbon, either in vacuoles or in bundle sheath cells. Biophysical CCMs, seen in cyanobacteria, algae and hornworts, concentrate Ci in subcellular compartments (pyrenoids or carboxysomes) where Rubisco is highly abundant[6].

The model organism for studying pyrenoid-based CCMs (pCCMs) is the green alga *Chlamydomonas reinhardtii*. *Chlamydomonas* pyrenoids are a liquid-like proteinaceous compartment whose phase separation is mediated by the interaction of Rubisco and the linker protein EPYC1 (refs. 7–9). In the algal pCCM, a series of Ci channels, such as LCI1, LCIA and bestrophin (BST) channels, facilitate Ci transport from the extracellular space into specialized thylakoid tubules[10]. These tubules traverse pyrenoids and contain a specific carbonic anhydrase (CA), CAH3. This CA catalyses the conversion of bicarbonate (HCO_3^-) into CO_2, which can freely diffuse out of the thylakoid tubules and into the pyrenoid. A starch sheath surrounds pyrenoids and might function as a CO_2 diffusion barrier to enhance the efficiency of the CCM[11,12]. Another CA, LCIB, localizes around the gaps of the starch sheath to recapture leaked CO_2 by converting it back to HCO_3^-. Recent modelling work suggests that a functional pCCM requires the joint operation of BST,

[1]Boyce Thompson Institute, Ithaca, NY, USA. [2]Plant Biology Section, Cornell University, Ithaca, NY, USA. [3]Department of Biology, Laval University, Quebec City, Quebec, Canada. ✉e-mail: lhg42@cornell.edu; fl329@cornell.edu

https://doi.org/10.1038/s41477-024-01871-0

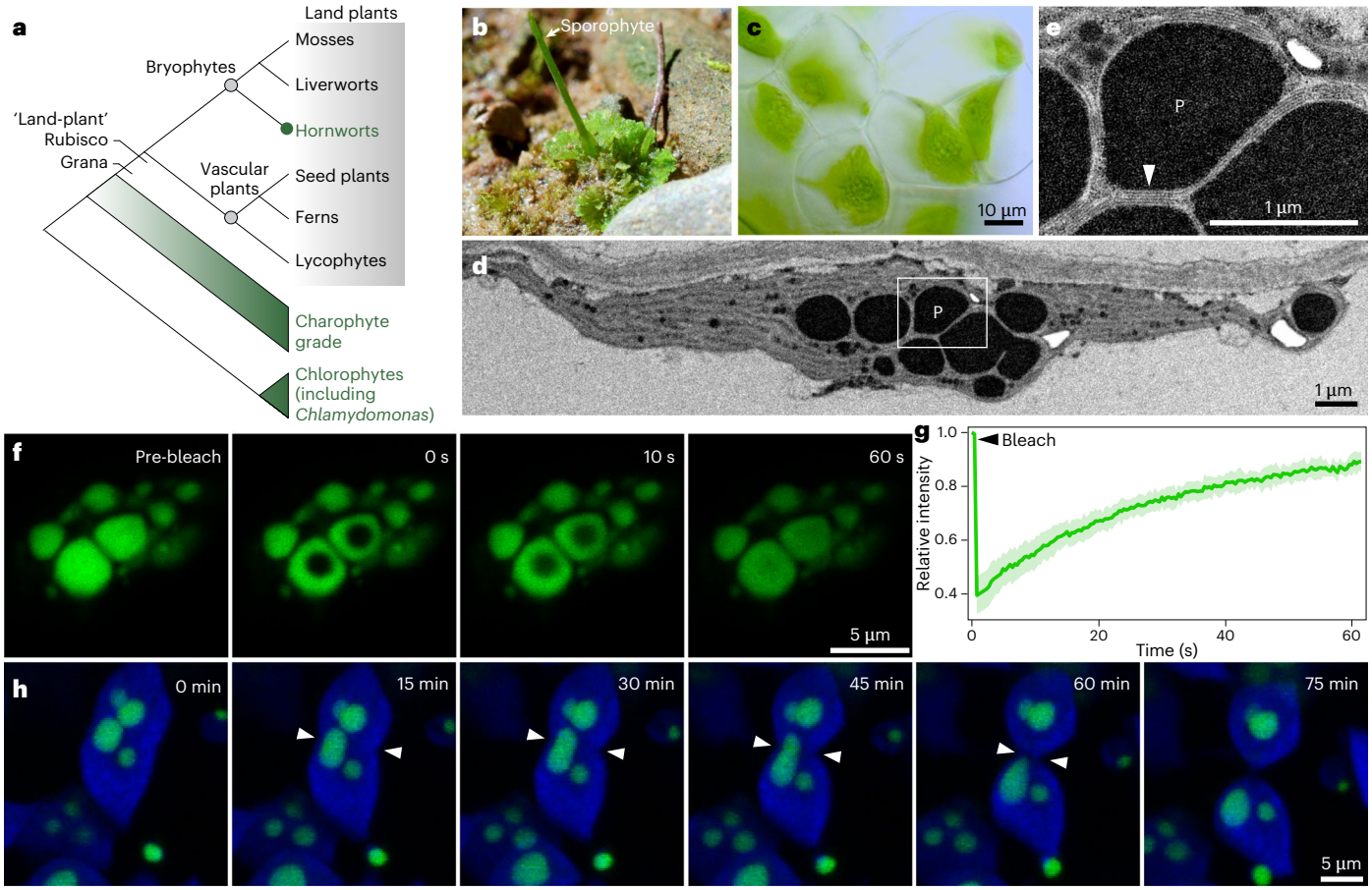

Fig. 1 | Morphology and physical properties of pyrenoids in *A. agrestis*. **a**, Hornworts are the only land plants that have a pCCM (pyrenoid-containing lineages are highlighted in green). Compared with *Chlamydomonas*, hornworts and crop plants share more common features in their chloroplast and Rubisco structures. **b**, The model hornwort *A. agrestis* in native habitat. **c**, *A. agrestis*, like many other hornwort species, has a single chloroplast per cell. **d**, Example TEM image of an *A. agrestis* chloroplast with multiple pyrenoids (P) ($n = 6$). **e**, A close-up view of a pyrenoid from **d**. The arrowhead points to a stack of thylakoid membranes encasing the pyrenoid. **f**, FRAP of hornwort pyrenoids labelled with RCA–mVenus. Photobleaching was targeted at the centre of two selected pyrenoids. **g**, FRAP recovery curve. The shaded area represents the s.e.m. ($n = 6$). **h**, Time-lapse images during cell division, showing signals of chlorophyll autofluorescence (blue) and RCA–mVenus fluorescence (green) over 75 minutes. The cell division plane is demarcated with white arrowheads. The full time series can be found in Supplementary Video 3.

LCIB and CAH3 (or the equivalent thereof), in addition to a pyrenoid enclosed by diffusion barriers[13].

Installing a pCCM into crop plants may boost CO_2 fixation by as much as 60%[4,14,15]. However, transplanting an algal CCM into land plants is complicated by the fact that around one billion years of evolution separate these lineages[16], over which time considerable differences have accumulated in chloroplast structure and protein sequences. Hornworts, a group of bryophytes (Fig. 1a,b), are the only known land plants with a pyrenoid or biophysical CCM of any kind[17–19]. Hornwort pyrenoids are functionally analogous to algal pyrenoids, acting as a locus of Rubisco accumulation and the focal point of the pCCM[17,18,20]. Characterizing the land plant pCCM may provide significant translational advantages. However, virtually nothing is known about the functional components enabling a pCCM in hornworts.

In this study, we characterized the morphology and physical properties of hornwort pyrenoids, as well as proteins involved in the pCCM, using the emerging model hornwort *Anthoceros agrestis*[21–23]. The diffusion properties of fluorescently tagged proteins and live-cell imaging indicate that *A. agrestis* pyrenoids are liquid-like, similar to *Chlamydomonas* pyrenoids. Putative CCM components for protein subcellular localization studies were selected by leveraging both Rubisco co-immunoprecipitation (co-IP) and 41 genomes spanning plant and algal diversity, 11 of which were newly sequenced from hornworts[24]. We found that hornworts possess orthologues of several core pCCM components from *Chlamydomonas*, and we provide evidence for shared functional roles. We thus infer that the chassis for the pCCM was present in the last common ancestor of land plants. Comparative genomics and co-IP did not reveal any clear EPYC1 homologue or analogue, implying that hornworts might have adopted a different strategy for Rubisco condensation. On the basis of our findings, we propose a spatial model for a land plant pCCM, which is consistent with reaction–diffusion modelling for a functional pCCM, and set the stage for future biochemical and genetic investigation.

Results

A. agrestis has multiple pyrenoids enclosed by thylakoids

We first characterized the morphology of *A. agrestis* pyrenoids using transmission electron microscopy (TEM). Compared with those of other hornwort species[17,20], *A. agrestis* pyrenoids are slightly larger, with a diameter ranging from 500 nm to 5 µm (Fig. 1c–e and Supplementary Fig. 1). *A. agrestis* (and hornworts in general) have multiple pyrenoids per chloroplast, though the number appears to be highly variable (Fig. 1c,d and Supplementary Fig. 1). Many of the pyrenoids do not have obvious connections to one another (indicated by arrows in Supplementary Videos 1 and 2), which differs from the singular pyrenoid found in most algal species, including *Chlamydomonas*[25].

https://doi.org/10.1038/s41477-024-01871-0

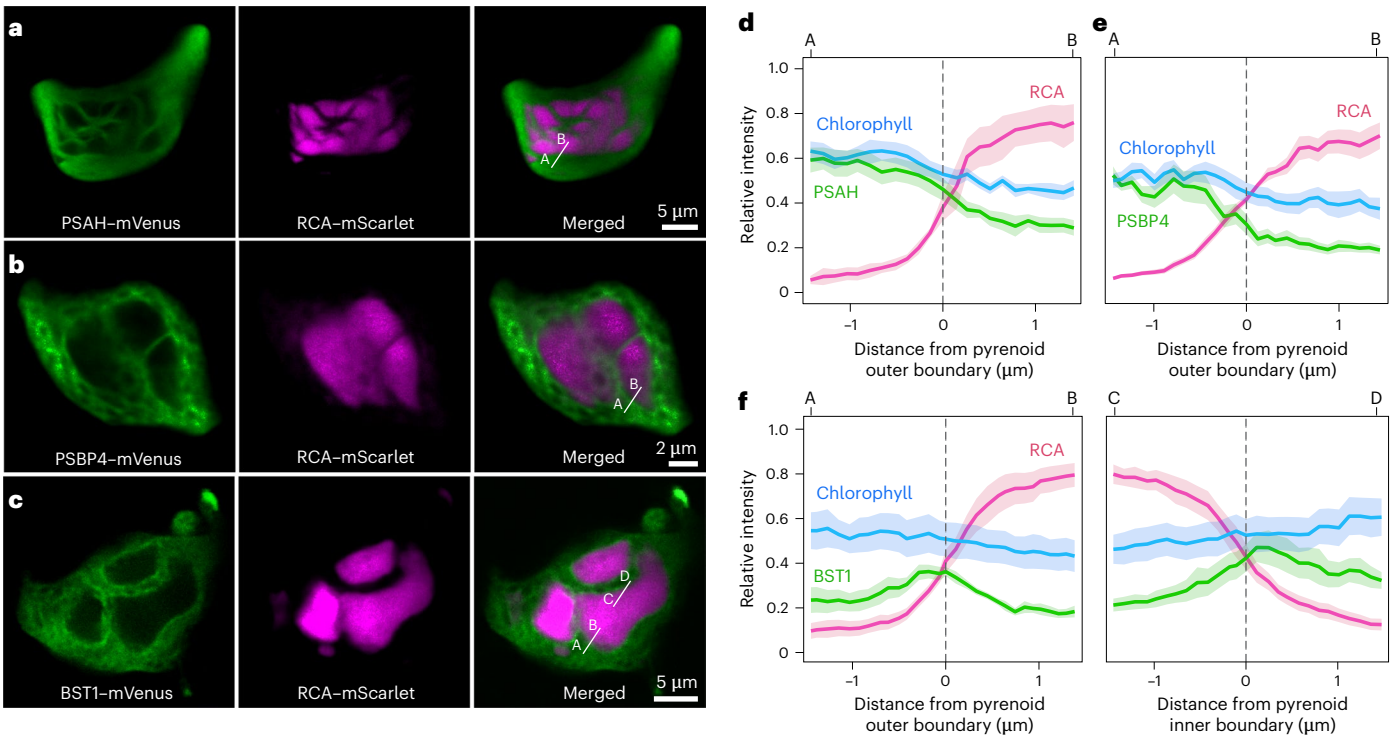

Fig. 2 | Distribution of photosystems and BST channels on thylakoids of *A. agrestis*. **a**,**b**, PSI (marked by PSAH–mVenus) (**a**) and PSII (marked by PSBP4–mVenus) (**b**) are distributed throughout the thylakoid network. **c**, BST1–mVenus has elevated fluorescence signals on the thylakoids immediately adjacent to pyrenoids. **d**,**e**, The fluorescence intensities of PSAH–mVenus (**d**) and PSBP4–mVenus (**e**) decline along the transects going into pyrenoids (white lines in **a** and **b**). **f**, Quantification of elevated BST1–mVenus fluorescence signals around pyrenoids (white lines in **c**). RCA–mScarlet was used as a pyrenoid marker. The shaded areas represent the s.e.m. on each curve (n = 10, 11 and 7 for BST, PSBP4 and PSAH, respectively).

Furthermore, we found a clear lack of any kind of starch sheath surrounding the pyrenoids of *A. agrestis* (Fig. 1d,e), whereas in *Chlamydomonas*, this sheath is believed to act as a critical diffusion barrier to prevent CO_2 leakage[11,12]. While no starch sheath is present, *A. agrestis* has many layers of stacked thylakoids wrapping around the pyrenoids (Fig. 1e, white arrowhead). Such stacked thylakoids might be sufficient to prevent CO_2 leakage, as suggested by recent modelling of the *Chlamydomonas* pCCM[13].

A. agrestis pyrenoids are liquid-like

While we have previously demonstrated that the pyrenoids of *A. agrestis* are the sites of Rubisco accumulation[23], the physical properties of these organelles have not been investigated. *Chlamydomonas* pyrenoids are known to be liquid-like, whereas the pyrenoids of the diatom *Phaeodactylum tricornutum* exhibit much less internal mixing[26]. To infer the physical properties of *A. agrestis* pyrenoids, we applied the fluorescence recovery after photobleaching (FRAP) technique on our previously generated stable transgenic line expressing Rubisco activase (RCA) tagged with mVenus[23]. RCA was shown to colocalize with Rubisco in *A. agrestis*[23] and can thus be used as a pyrenoid marker. We found that the mVenus signal exhibited a rapid recovery ($t_{1/2}$ = 16.2 s) post photobleaching (Fig. 1f,g). The unbleached region equilibrates with the bleached region of the same pyrenoid within 60 s (Fig. 1f,g), suggesting high levels of internal mixing, at least for RCA. Furthermore, live-cell imaging of dividing chloroplasts in the same RCA–mVenus line demonstrated elongation of a pyrenoid as it is being pulled to one of the daughter cells, again suggesting that *A. agrestis* pyrenoids are liquid-like (Fig. 1h and Supplementary Video 3).

Photosystems I and II localize close to pyrenoids

Given the tight association between thylakoids and pyrenoids in hornworts (Fig. 1d,e), we next examined how thylakoids are distributed across the chloroplasts of *A. agrestis*, especially whether there was differential localization of Photosystem I (PSI) and Photosystem II (PSII) relative to the pyrenoids. It is thought that in hornworts PSII localizes mainly to the grana, while PSI has a more ubiquitous distribution, including in the channel thylakoids (which are thought to be analogous to stroma lamellae in vascular plants)[20,27]. The evidence supporting this claim is, however, limited. To visualize PSI and PSII distribution, we transiently expressed fluorescently tagged PSI reaction centre subunit VI (PSAH) and PSII subunit P4 (PSBP4), respectively, in *A. agrestis* using a biolistics transformation approach[23].

In agreement with TEM images (Fig. 1d,e), confocal imaging of tagged PSAH and PSBP4 showed a tight association between thylakoids and pyrenoids (Fig. 2a,b and Supplementary Fig. 2). Pyrenoids are almost entirely enveloped by thylakoids (Fig. 2a,b), and both PSI and PSII localize in the membranes directly adjacent to the pyrenoids (Fig. 2a,b,d,e). Consistent with previous reports[20], PSI is uniformly distributed across the thylakoid network in hornworts, so much so that the signal is absent only in pyrenoid bodies (Fig. 2a and Supplementary Fig. 2). The localization of PSII appears to be less uniform than that of PSI, with gaps as well as smaller regions of much higher intensity, probably grana, scattered throughout the chloroplast (Fig. 2b and Supplementary Fig. 2). Neither PSAH nor PSBP4 in *A. agrestis* displays distinctive localization relative to pyrenoids, whereas in *Chlamydomonas* PSAH concentrates in the thylakoid tubules and PSBP4 forms puncta at the periphery of the pyrenoid[28]. The distributions of PSI and PSII shown here provide a useful point of comparison to other thylakoid-localized proteins in *A. agrestis*.

Thylakoids adjacent to pyrenoids are enriched in BST

A biophysical CCM in eukaryotes requires the presence of thylakoid membrane channels to allow HCO_3^- to enter the thylakoid lumen. In *Chlamydomonas*, a number of BST channels fill this role[29]. There are

https://doi.org/10.1038/s41477-024-01871-0

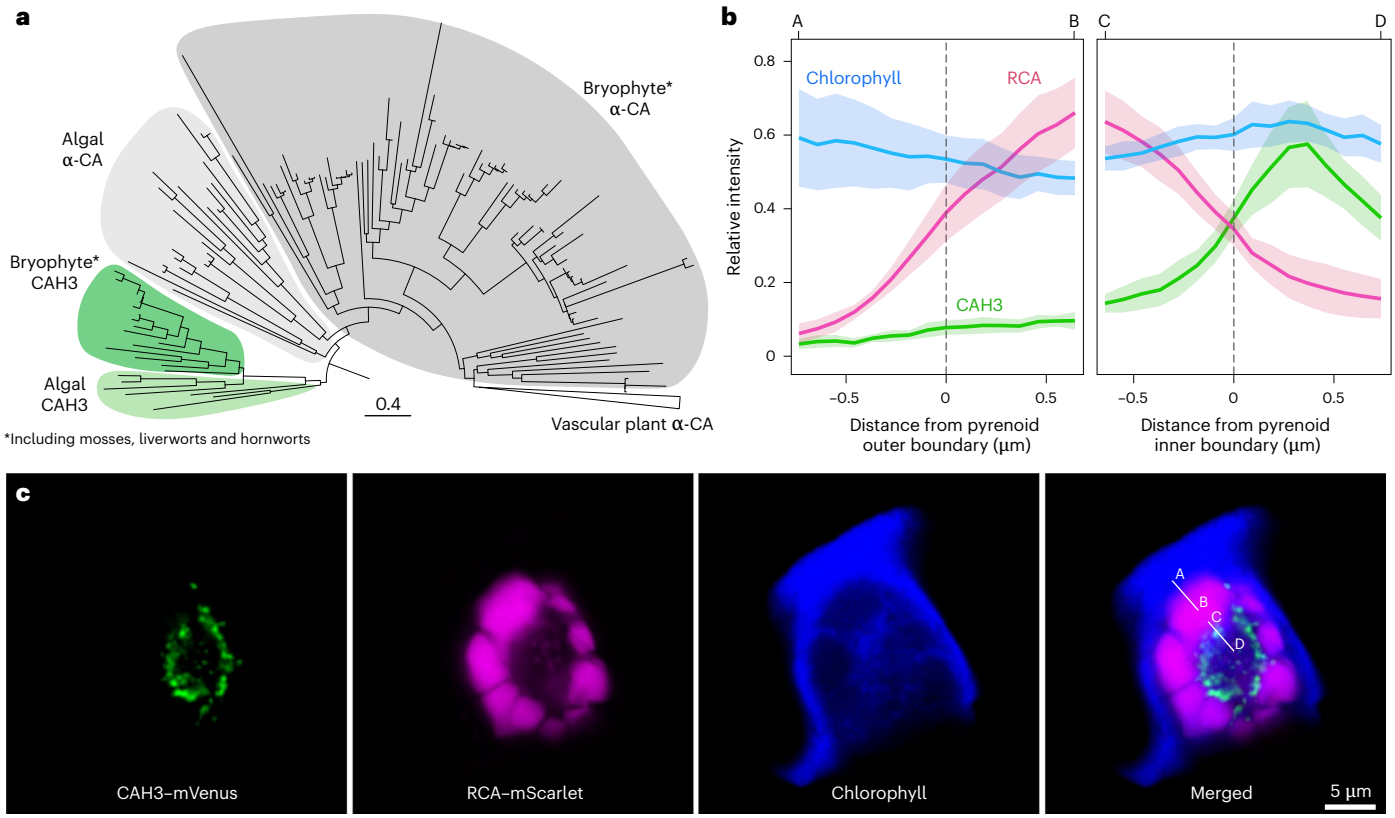

Fig. 3 | Pyrenoids are organized around CAH3 in *A. agrestis*. **a**, Phylogeny of α-CAs, including CAH3. Orthologues of *Chlamydomonas* CAH3 can be found in hornworts and other bryophytes (shaded in dark green) but were lost in vascular plants. **b**, CAH3–mVenus fluorescence intensity spikes at the interior side of pyrenoids (white lines in **c**). The shaded areas represent the s.e.m. on each curve ($n = 10$). **c**, Example images of a cell co-transformed with CAH3–mVenus (green) and RCA–mScarlet (magenta). Chlorophyll autofluorescence is shown in blue. The *Z*-stack series can be found in Supplementary Video 4.

four *Chlamydomonas* BST orthologues in the *A. agrestis* genome, but only one, BST1, is expressed abundantly (Supplementary Table 1). Fluorescent protein tagging indicates that BST1 in *A. agrestis* is distributed throughout the thylakoid network but exhibits elevated fluorescence intensity immediately adjacent to the pyrenoids (Fig. 2c,f and Supplementary Fig. 3). This pattern is in contrast to that of PSAH and PSBP4, where the fluorescence intensities continuously decline going into pyrenoids (Fig. 2d,e). The localization of *A. agrestis* BST1 is similar to what was reported in *Chlamydomonas*[29] and implies that BST1 probably has the same function in hornworts, transporting HCO_3^- to the thylakoid lumen.

β-CA1 localizes to the cytoplasm

We next examined the localization of several CAs in *A. agrestis*. There are three known CA families in land plants (α, β and γ), which catalyse the reversible hydration of CO_2 to HCO_3^- (ref. 30). We first examined β-CA1 because it is the second most abundant protein in plants after Rubisco and is typically located in the chloroplast stroma[31]. However, the presence of a CA in the stroma, and specifically the pyrenoid matrix, in hornwort chloroplasts would probably convert CO_2 into HCO_3^-, thus decreasing the availability of CO_2 for Rubisco and short-circuiting the pCCM. β-CA1 in *A. agrestis*, unlike its orthologues in C_3 plants, lacks a predicted chloroplast transit peptide[32]. In fact, no CA in *A. agrestis*, except for CAH3 and LCIB (mentioned below), possesses any chloroplast transit peptides. Fluorescently tagged *A. agrestis* β-CA1 confirmed the cytoplasmic localization of this protein (Supplementary Fig. 4). This apparent absence of a diffuse stromal β-CA in *A. agrestis* is consistent with models of a pCCM[13] and reinforces the idea that the relocalization of stromal CAs to the cytoplasm is a key step towards engineering CCMs in crop plants[33].

CAH3 localizes to the space between pyrenoids

While a stroma-localized CA would prohibit pCCM function, treatment with a CA inhibitor (ethoxzolamide) greatly decreased the rate of photosynthesis and significantly increased the CO_2 compensation point of hornworts, but not liverworts[34], demonstrating that a CA is required for hornwort pCCM. In *Chlamydomonas*, CAH3 is located in thylakoid tubules, which form a knot in the centre of the pyrenoid[35,36]. *Chlamydomonas* CAH3 converts lumenal HCO_3^- to CO_2, allowing for diffusion of CO_2 across the thylakoids and into the pyrenoid space. We discovered that the genomes of hornworts and some other bryophytes possess orthologues of the *Chlamydomonas* CAH3 gene, while vascular plants appear to have lost it (Fig. 3a and Supplementary Fig. 5). *A. agrestis* CAH3 is abundantly expressed and has retained not only the three critical histidine residues at the active site[37] but also the lumenal transit peptide (Supplementary Fig. 6 and Supplementary Table 1). Importantly, we found that fluorescently tagged *A. agrestis* CAH3 formed puncta localized to the periphery of pyrenoids, concentrated adjacent (<500 nm) to the most interior side of the pyrenoids (Fig. 3b,c and Supplementary Fig. 6; see Supplementary Video 4 for the *Z*-stack). CAH3 thus appears to be at the centre of the chloroplast and in the space between pyrenoids. This localization pattern supports CAH3's role in supplying CO_2 to the surrounding pyrenoids in the hornwort pCCM.

LCIB localizes to the chloroplast membrane

Another CA crucial to *Chlamydomonas* pCCM is LCIB[38,39], which localizes to the periphery of *Chlamydomonas* pyrenoids under low-CO_2 conditions to recapture leaked CO_2 from pyrenoids[40]. LCIB orthologues are present in hornworts and have conserved residues for the coordination of Zn^{2+} ion binding and catalysis[21]. Apart from hornworts, no LCIB homologue has been identified in other land plant genomes[21].

https://doi.org/10.1038/s41477-024-01871-0

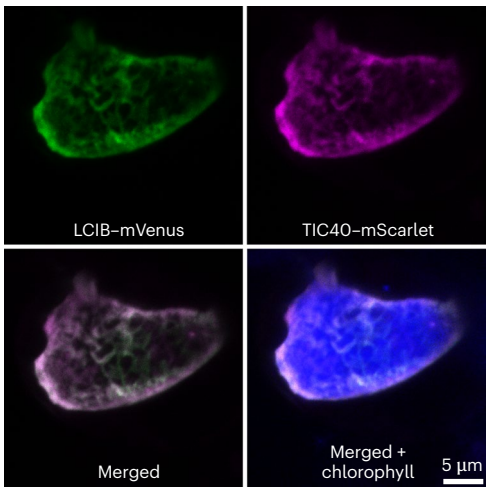

Fig. 4 | *A. agrestis* LCIB localizes to the chloroplast membrane. Example images of maximum-intensity projection of *A. agrestis* co-transformed with LCIB–mVenus (green) and TIC40–mScarlet (magenta). Chlorophyll autofluorescence is shown in blue. $n = 12$.

This exclusive presence of LCIB in pyrenoid-bearing algae and hornworts suggests that LCIB may have a role in hornwort pCCMs. We discovered that, unlike LCIB in *Chlamydomonas*, fluorescently tagged *A. agrestis* LCIB did not localize to the immediate pyrenoid periphery but instead localized to the edge of the chloroplast (Supplementary Fig. 7). To interrogate this pattern further, we colocalized LCIB–mVenus with mScarlet-tagged TIC40, which is a known component of the inner chloroplast membrane[41]. The fluorescence of LCIB–mVenus and TIC40–mScarlet clearly overlapped (Fig. 4), thus supporting the idea that LCIB is indeed membrane-bound.

There are several reasons why the membrane-localized LCIB in *A. agrestis* is logical. In *Chlamydomonas*, LCIB localization to the starch plates is probably mediated by its homologue LCIC[40]. Hornworts have neither starch plates nor LCIC; thus, we would not expect to see a similar localization of LCIB. Furthermore, if LCIB were to surround each pyrenoid in hornworts, it would probably rehydrate any CO_2 that had just been released by CAH3 (which is placed outside of pyrenoids; Fig. 3c), thus short-circuiting the pCCM. This scenario would be impossible in *Chlamydomonas* because CAH3 is localized in the thylakoid tubules embedded within pyrenoids. Indeed, previous reaction–diffusion modelling of a pCCM showed that when thylakoid stacks are used as a diffusion barrier (as is the case in hornworts), LCIB localization to the chloroplast envelope is optimal. This is because localizing LCIB as far from the pyrenoids as possible minimizes 'stealing' of CO_2 from the pyrenoid matrix and can still serve as the last line of defence for CO_2 leakage[13]. Furthermore, for any passively diffused CO_2 coming into the chloroplast, membrane-localized LCIB would quickly convert it into HCO_3^- before it could escape back to the cytoplasm.

No evidence of a canonical Rubisco linker in *A. agrestis*

Pyrenoids, which house Rubisco, are central to pCCMs. Rubisco linker proteins that mediate pyrenoid formation have been identified in the green algae *Chlamydomonas* (EPYC1)[7] and *Chlorella sorokiniana* (CsLinker)[42], as well as in the diatom *Phaeodactylum tricornutum* (PYCO1)[26]. A putative linker has also been proposed in the chlorarachniophyte alga *Amorphochlora amoebiformis*[43]. While there is little sequence similarity between these linker proteins, they share certain physicochemical properties, and importantly, they can all be pulled down using co-IP with an anti-Rubisco antibody. To identify the possible Rubisco linker in *A. agrestis*, we optimized the lysis of *A. agrestis* tissue and conducted co-IP using a custom anti-Rubisco antibody with soluble proteins clarified from hornwort lysate (Supplementary Fig. 8). From our proteomic analysis, we found that a total of 375 proteins were significantly enriched by IP relative to the control where no antibodies were applied. Both Rubisco large and small subunits as well as RCA were among the highest enriched proteins (Supplementary Fig. 9). Gene Ontology terms relating to chloroplasts, photosynthesis and thylakoid membranes were significantly overrepresented in the immunoprecipitated samples, probably reflecting the tight association between pyrenoids and thylakoids (Supplementary Fig. 10). Together, these data suggest that our co-IP approach was successful in enriching the bait (that is, Rubisco) and its prey proteins (that is, potential interactors).

Of the proteins significantly enriched by Rubisco co-IP, we searched for ones with similar physicochemical properties to other known Rubisco linker proteins. Namely, the candidate protein needs to (1) be highly expressed, (2) be rich in repeat motif sequences, (3) be highly disordered and (4) have a high isoelectric point (see Methods for the details). Only one protein was found (AnagrOXF.S1G403800.t1) possessing promising characteristics as a linker protein, which we named Putative Pyrenoid Protein 1 (PPP1). While it lacks any identifiable repeats, PPP1 displays an oscillating disorder profile, a chloroplast transit peptide and high enrichment in the co-IP fraction. PPP1 also contains Structural Maintenance of Chromosomes (SMC) domains, which have been found in other pyrenoid-associated proteins in *Chlamydomonas*[12] (Supplementary Fig. 11). We found that fluorescently tagged PPP1 did not localize to the pyrenoid but instead localized to the stroma (Supplementary Fig. 11). Interestingly, in some instances its localization became more concentrated in a region central to the chloroplast (Supplementary Fig. 11), reminiscent of the CAH3 localization pattern (Fig. 3). In any case, it is clear that PPP1 is not a Rubisco linker protein.

While our co-IP experiment found no evidence for a canonical linker in *A. agrestis*, we cannot rule out the possibility that one exists but has evaded our notice. We therefore broadened our search to include the entire predicted proteome of *A. agrestis*, using the same screening criteria as before. Only one candidate protein, PPP2 (AnagrOXF.S4G428300.t2), was found meeting these criteria. Compared with the known Rubisco linkers (EPYC1, CsLinker and PYCO1), the repeats of PPP2 were much more irregular and spaced erratically across the protein sequence (Supplementary Fig. 12). Using fluorescent protein tagging, we showed that PPP2 localized not to the pyrenoid but instead to the spaces between pyrenoids (Supplementary Fig. 12). Similar to PPP1, no evidence supports PPP2 being a linker protein for *A. agrestis* Rubisco. Our results thus point to a possibility that *A. agrestis* employs a different mode of pyrenoid condensation from what has been reported in algae.

Additional protein components in *A. agrestis* pyrenoids

While the Rubisco linker in *A. agrestis* remains elusive, we found other proteins of interest that were enriched by IP. For example, Rubisco assembly factors (Raf1, Cpn60α and Cpn60β) showed significant enrichment by co-IP (Supplementary Fig. 9). We tested the localization of Cpn60β, a chaperonin required for Rubisco folding[44,45], by fluorescent protein tagging and found Cpn60β–mVenus to be clearly pyrenoid-localized (Fig. 5a and Supplementary Fig 13). To further explore whether Rubisco biogenesis might occur within pyrenoids, we tagged another Rubisco assembly factor, RbcX2 (ref. 46). While RbcX2 was not co-immunoprecipitated, it is also localized to the pyrenoids (Fig. 5b and Supplementary Fig. 13).

Another protein enriched by co-IP was glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Chloroplastic GAPDH catalyses the reductive step of the Calvin–Benson–Bassham (CBB) cycle and is composed of A and B subunits, both of which were significantly enriched in the immunoprecipitated fraction (Supplementary Fig. 9). To examine whether GAPDH is a pyrenoid component, we fluorescently tagged the B subunit (GapB) and confirmed its pyrenoid localization (Fig. 5c and Supplementary Fig. 14). It is possible that the CBB cycle (or part

https://doi.org/10.1038/s41477-024-01871-0

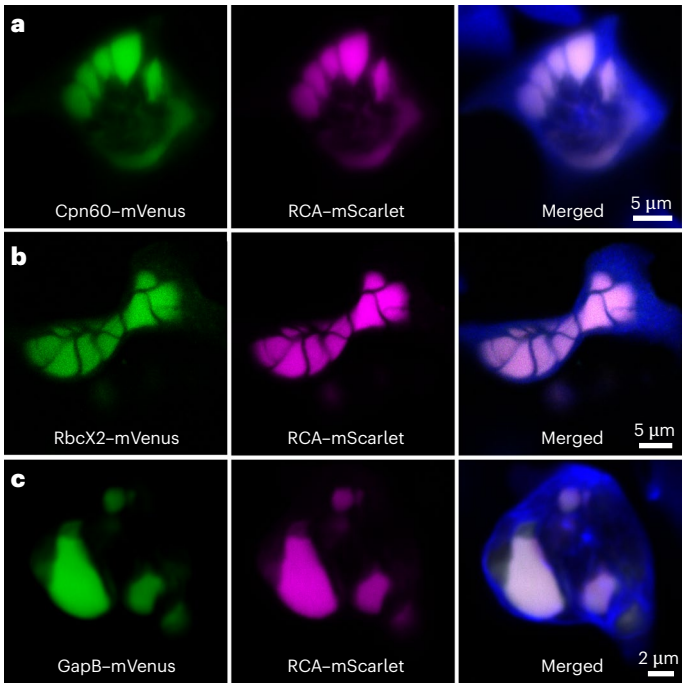

Fig. 5 | *A. agrestis* Rubisco assembly and CBB cycle proteins localize to pyrenoids. a–c, Example images of cells expressing Cpn60 ($n = 16$) (**a**), RbcX2 ($n = 12$) (**b**) or GapB ($n = 5$) (**c**) tagged with mVenus (green). RCA–mScarlet was co-transformed to mark pyrenoids (magenta). The merged images also include chlorophyll autofluorescence (blue).

of it) takes place in pyrenoids to limit the diffusion time of CBB cycle intermediates. In contrast, CBB enzymes in *Chlamydomonas* are localized to the periphery of the pyrenoid[47].

On the basis of our results, it is plausible that *A. agrestis* pyrenoids are an inclusive compartment, where multiple cellular processes revolving around Rubisco take place. Future studies systematically tagging Rubisco assembly and CBB cycle proteins are needed to further test this hypothesis.

Discussions

A working model for a hornwort pCCM

Our current evidence suggests a model (Fig. 6) for the hornwort pCCM that bears similarity to the algal pCCM but has some key differences. First, since hornworts largely do not exist in an aqueous environment (where CO_2 diffusion is severely limited), we suspect that periplasmic CAs and active HCO_3^- transporters such as LCI1/HLA are not essential. Likewise, despite retaining the LCIB–BST–CAH3 chassis, hornworts do not have an orthologue to LCIA, the hypothesized HCO_3^- channel in *Chlamydomonas*. Given this, we hypothesize that the hornwort CCM probably operates under a 'passive', LCIB-dependent mode as described by Fei et al.[13]. Once CO_2 passively diffuses across the chloroplast membrane, it is rapidly converted to HCO_3^- by membrane-localized LCIB, trapping it in the stroma. In the stroma, HCO_3^- can enter the thylakoid space via BST1, where a concentration gradient drives it towards the pyrenoid-adjacent thylakoids enriched in CAH3. There, CAH3 catalyses the dehydration of HCO_3^-, creating a CO_2-enriched environment for the nearby pyrenoids. Leakage of CO_2 from the pyrenoid matrix is hindered by the highly reticulated thylakoid network, and any escaped CO_2 can be recaptured by LCIB at the chloroplast membrane. While we cannot rule out the presence of active Ci transport, the condition to drive an efficient pCCM is met on the basis of the localizations of LCIB, BST and CAH3, thylakoid-enclosed pyrenoids, and the reaction–diffusion model by Fei et al.[13].

The spatial model we present here provides a framework for future hypothesis-driven research. Although protein localization alone does not provide the full proof of a protein's function, no reverse genetics method has yet been successfully developed in *A. agrestis*. Future work investigating the carbon assimilation efficiency of selective gene knockdown or knockout mutants, as well as biochemical in vitro characterizations, will be crucial to confirm or revise our model.

The first land plants probably had the chassis for a CCM

Operating an effective biophysical CCM requires the strategic placement of Ci channels and CAs. It is striking that some of these key components (LCIB, BST and CAH3) characterized in *Chlamydomonas* have orthologues in hornworts and probably play similar roles in their respective pCCMs. It is difficult, if not impossible, to determine whether hornwort pCCMs represent an ancestral trait or an independent origin. In either case, however, the presence of the LCIB–BST–CAH3 chassis in hornworts suggests that the underlying infrastructure to concentrate CO_2 was present in the last common ancestor of land plants. Considering that various algae possess biophysical CCMs even in the absence of pyrenoids[48–51], it is possible that the earliest land plants could have operated a functional CCM.

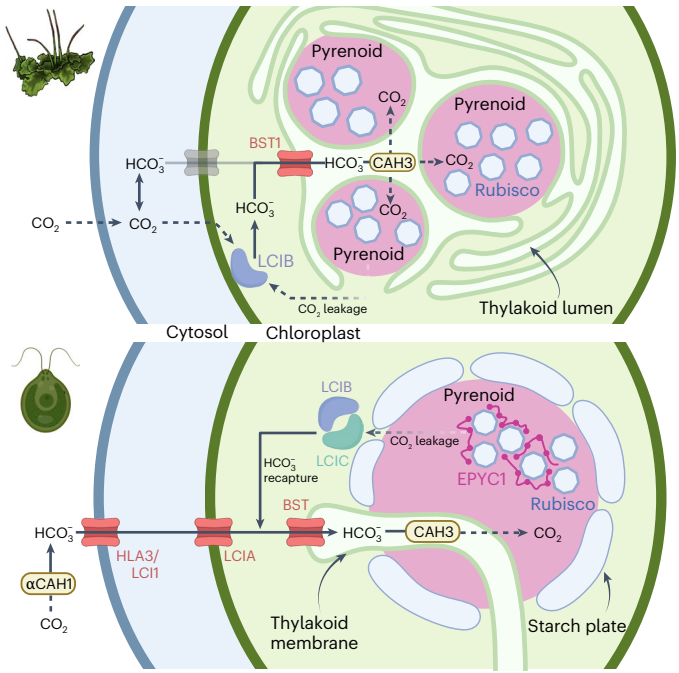

Fig. 6 | The spatial model of the *A. agrestis* pCCM and comparison with that of *Chlamydomonas*. We propose that the hornwort pCCM (top) operates in a similar manner to that of *Chlamydomonas* (bottom) but with a few key differences. First, the primary species of Ci entering the cytosol, and subsequently to the chloroplast, is probably CO_2 rather than HCO_3^-. The route of HCO_3^- import in *A. agrestis* is thus displayed in grey. Second, LCIB is localized to the chloroplast envelope and might serve two functions: (1) capturing CO_2 as it enters the chloroplast, thus trapping it in the stroma, and (2) recapturing CO_2 leaking from the pyrenoid space. Third, CAH3 is enriched in specialized thylakoids found in the centre of the chloroplast, around which the pyrenoids are organized. This pattern is different from that in *Chlamydomonas*, where CAH3 is concentrated in the thylakoid tubules that penetrate into the pyrenoid matrix. Fourth, while the pyrenoid matrix in *A. agrestis* has similar biophysical properties to that of the *Chlamydomonas* pyrenoids, we could not find evidence of a disordered linker protein like EPYC1. Lastly, the pyrenoids of hornworts are enclosed by thylakoids, which might be functionally equivalent to the starch plates in *Chlamydomonas*. Dashed and solid arrows represent the movement of gaseous CO_2 and bicarbonate ions, respectively. Figure created with BioRender.com. Credit: hornwort illustration, Rebecca Key.

We propose two scenarios that could result in hornworts being the only land plants having a pCCM. In one, pyrenoids were de novo evolved in hornworts on top of the LCIB–BST–CAH3 chassis already in place. The loss of LCIB and CAH3 in other plant lineages rendered the evolution of pyrenoids less advantageous. The supposed absence of a canonical Rubisco linker for pyrenoid formation in *A. agrestis* lends support for the separate origins of pyrenoids. Alternatively, pyrenoids could be ancestral to land plants but were subsequently lost once in setaphytes (mosses and liverworts) and another time in vascular plants. In this scenario, hornworts did not independently evolve pyrenoids but rather retained them as a plesiomorphic trait. Future studies comprehensively comparing shared CCM-related orthologues between hornworts and *Chlamydomonas* might provide additional insights into the evolution of CCMs in land plants.

Implications for the repeated evolution of pyrenoids in hornworts

Within hornworts, pyrenoids have been lost and gained multiple times across the phylogeny[52]. Interestingly, LCIB, BST and CAH3 are present in all the sequenced hornwort species, even those lacking pyrenoids[24]. Considering that pyrenoid-absent species have CO_2 compensation points that are intermediate between those of liverworts and pyrenoid-containing hornworts[53], a baseline biophysical CCM probably exists among most hornworts, and the strength of this CCM could be accentuated by the presence of a pyrenoid. The universal presence of this LCIB–BST–CAH3 chassis could explain the repeated gains of pyrenoids, as it might have predisposed the evolution of pyrenoids. The specific localization of CAs and Ci channels is also probably the first step to the evolution of a fully functional pCCM, rather than pyrenoid formation itself.

A blueprint for engineering a biophysical CCM in land plants

Significant strides have been made in transplanting parts of the algal CCM to *Arabidopsis* in recent years. For example, by introducing EPYC1, SAGA1 and SAGA2, proto-pyrenoids with starch plates have been successfully built in *Arabidopsis*[54,55]. However, the entire algal CCM module has yet to be integrated into flowering plants. Our investigation into the hornwort system demonstrates that some components of the algal pCCM may not be required in land plants. First, a starch sheath around pyrenoids may not be necessary as thylakoid stacks could provide a sufficient CO_2 diffusion barrier (Fig. 1). Second, if hornworts indeed operate a passive pCCM, then our work provides empirical evidence for its feasibility[13]. In a passive pCCM, the proper localization of a minimal set of CAs (for example, LCIB and CAH3) and thylakoid-localized Ci channels (for example, BST) may be sufficient to concentrate CO_2 without the need for active Ci transport to the chloroplast[13]. Importantly, recent modelling suggested that introducing active HCO_3^- transporters to build pCCMs in land plants is viable only if chloroplast membrane permeability to CO_2 is low[56]. The passive pCCM of hornworts would not encounter this challenge and could be simpler to engineer. Taken together, hornworts provide an alternative blueprint to design a future pCCM.

Methods

A. agrestis cultures

A. agrestis axenic cultures were maintained on *A. agrestis* gametophyte growth medium (AG medium[57]) supplemented with 0.2% sucrose following the protocol described in Lafferty et al.[23]. The growth conditions included a temperature of 22 °C, a 16 h/8 h light/dark cycle and a light intensity ranging from 6 to 25 µmol m^{-2} s^{-1}. Warm white light and cool white light were emitted by Ecolux XL Starcoat F32T8 XL SP30 ECO fluorescent bulbs and F32T8 XL SP41 ECO bulbs, respectively (General Electric). A total of 3–6 g of fresh thallus tissue was homogenized and prepared for transformation[23]. The homogenate was filtered using a 100 µM cell strainer (MTC Bio), and the thallus tissue was washed with sterile deionized water and plated on AG medium supplemented with 2% sucrose.

TEM

Plant materials were processed for TEM following the published protocol[58], with some modifications. Pure cultures were fixed in 3% glutaraldehyde, 1% formaldehyde and 0.75% tannic acid in 0.05 M Na-cacodylate buffer, pH 7, for 3–4 h at room temperature. After several rinses in Sorensen's 0.1 M buffer, the samples were post-fixed in 2% osmium tetroxide, dehydrated in an ethanol series and embedded in Spurr's resin for 24 h. Thin sections were cut with a diamond knife, stained with methanolic uranyl acetate for 15 min followed by Reynolds' lead citrate for 10 min and observed with a Hitachi H-7100 transmission electron microscope at the Imaging-Microscopy Platform of the IBIS, Université Laval.

Construct design and cloning

Constructs were designed and built using the OpenPlant toolkit[59]. All genes of interest were driven by the same constitutive promoter (*A. agrestis* native *Elongation factor 1 alpha* promoter) and the *Nos* terminator (OP-053). Coding sequences of the target genes were synthesized by Twist Biosciences with synonymous substitutions made where necessary to remove internal BsaI or SapI Type II restriction cut sites or to reduce repetitive nucleotide sequences if required for synthesis. Assembly reactions were performed using a one-pot mixture that consisted of Type II restriction enzymes with T4 DNA ligase (NEB)[59]. Heat shock transformation with NEB 5-alpha Competent *Escherichia coli* was used to regenerate ligated products on the respective Luria broth agar plates with either kanamycin or streptomycin. Colony PCR was performed with KOD One PCR Master Mix (Toyobo) to screen colonies before recovering plasmids with a 5 ml overnight culture and miniprep (QIAgen). Whole-plasmid sequencing (Eurofin Genomics) was performed as a final validation. To prepare plasmids for biolistic mediated transformation, a total of 50 ml of *E. coli* culture was grown overnight at 37 °C. The plasmid DNA was extracted using the PureYield Plasmid Midiprep system (Promega), followed by vacuum concentration in a SpeedVac Concentrator (Thermo Fisher Scientific) to obtain DNA concentrations of 1 µg $µl^{-1}$.

Transient expression of fluorescently tagged proteins

To obtain transient expression of fluorescently tagged proteins in *A. agrestis*, biolistic transformation was performed following the protocol of Lafferty et al.[23]. A total of 20 µl of 1 µg $µl^{-1}$ DNA was combined with 50 µl of 50 mg ml^{-1} gold microparticles (1 µm in diameter, suspended in 50% sterile glycerol), 100 µl of 2.5 M $CaCl_2$ and 40 µl of 0.1 M spermidine. The mixture was vortexed for 1 min and then centrifuged for 10 s. The supernatant was removed, and the gold pellet was washed with 180 µl of 70% ethanol before being vortexed again for 1 min, followed by 10 s of centrifugation. The supernatant was removed, and an additional wash with 200 µl of 100% ethanol was followed by a final centrifugation step, after which 135 µl of the supernatant was removed. Resuspended particles were loaded onto the PDS-1000/He system (Bio-Rad). *A. agrestis* tissues were bombarded with a target distance of 14 cm and a burst pressure of 450 psi, under a vacuum of 28 inHg. The tissue was left to recover for three to ten days, under standard culturing conditions as described above, before imaging.

Confocal imaging

Imaging was performed using a Zeiss LSM710 confocal microscope, a Leica TCS SP5 Laser Scanning Confocal Microscope or a Leica Stellaris 5 confocal microscope. Proteins tagged with mVenus were excited with a 514 nm laser line with an emission band of 524/551 nm. Chlorophyll autofluorescence and proteins tagged with mScarlet were excited with a 561 nm laser line, with emission bands of 658/699 and 579/609 nm, respectively. Each FRAP experiment started with four initial scans

before the region of interest was bleached. Circular regions of interest with a 1.3 µm diameter were exposed eight times at 20% intensity of the 514 nm laser, before recovery was allowed for 160 cycles (66 seconds) with laser power attenuated at 0.5% intensity. Replicates were collected by photobleaching individual pyrenoids within a cell. Subsequent analysis was conducted using the FRAP wizard in Leica Application Suite Advanced Fluorescence (v.4.6.1.27508), and the data were plotted in R (v.4.3.1)[60]. Intensity profiles for all tagged proteins were collected by using the line tool to transect regions of interest (for example pyrenoid boundaries, chloroplast boundaries) and then plotting the intensity profile in Fiji (v.2.14.0)[61]. These intensity profiles were then exported for subsequent analyses and plot generation in R.

Homology-based approach to identify candidate CCM proteins

To identify candidate CCM genes in hornworts, we used Orthofinder (v.2.5.4)[62] on a broad sampling of plant and algal genomes (Supplementary Table 2), including 11 species of hornworts[24], to generate orthogroups. We then selected orthogroups containing key *Chlamydomonas* CCM genes (Supplementary Table 1). Because orthogroups can sometimes represent large gene families, to infer the direct orthologues in hornworts, we performed phylogenetic reconstruction using IQ-TREE v.2.2.0 (ref. 63) to identify the hornwort sequences in the same subfamily as the *Chlamydomonas* genes of interest. From these, we selected *A. agrestis* genes that showed high expression levels using the RNA-seq data from Li et al.[21]. In cases where there were multiple orthologues present in *A. agrestis*, we selected the gene with the highest expression. In addition to CCM genes, we chose PSAH and PSBP4 as PSI and PSII markers, respectively, to compare the distribution of photosystems and to visualize thylakoids.

Targeted bioinformatic search for a Rubisco linker protein

A targeted search for potential linker proteins was conducted in a manner similar to but less stringent than that described in Mackinder et al.[7]. Genomic sequences of two *A. agrestis* strains[21] were first screened for tandem repeats using Xstream v.1.73 (ref. 64). The default parameters were used with the following exceptions: minimum period, 20; maximum period, 100; minimum copy number, 2; and minimum tandem repeat domain, 40. Proteins that passed this screen were then analysed for a high isoelectric point (>8) using Expasy[65] and a high proportion of disordered sequence (>50%) using the PONDR server[66]. From this screen, lowly expressed genes were filtered on the basis of their corresponding RNA transcript abundance (transcripts per million, <10)[24].

Optimizing lysis of *A. agrestis* thallus

To derive soluble proteins from *A. agrestis* thallus, we experimented with the lysis of *A. agrestis* by testing four different buffers: buffer A (20 mM HEPES (pH 7.0), 50 mM NaCl, 12.5% v/v glycerol), buffer B (buffer A with 2% v/v Triton X-100), buffer C (20 mM CAPS (pH 11.0), 50 mM NaCl, 12.5% v/v glycerol) and buffer D (buffer C supplemented with 2% v/v Triton X-100). For each lysis, 5 g of *A. agrestis* tissue was harvested from liquid cultures in the AG medium[57]. The biomass was frozen with liquid N_2 and crushed to fine powder with a mortar and pestle. A total of 5 ml of test buffer supplemented with 1 mM phenylmethylsulfonyl fluoride and one protease inhibitor tablet (Roche) was added to resuspend the fine powder. The suspension was lysed with a French pressure cell press (American Instrument Company) at 1,000 psi. Polyvinylpolypyrrolidone (2%) was added to the French-pressed lysate, and the mixture was sieved through four layers of Miracloth, pre-wetted with the respective test buffer. The filtrate was centrifuged (21,000 *g* for 45 min at 4 °C) to separate the pellet and supernatant fractions. The supernatant fraction was further filtered through a 0.22 µm syringe filter. Both fractions were analysed via SDS–PAGE and western blot using 8–16% Mini-PROTEAN TGX Precast Protein Gels (Bio-Rad). For immunoblotting of Rubisco, a polyclonal antibody was raised in rabbits against the *A. agrestis* Rubisco large subunit carboxy-terminus peptide EVWKEIKFVFETIDTL and affinity purified (Life Technologies).

Co-IP of Rubisco

A total of 100 µl of Protein A resin (Dynabeads, Invitrogen) was aliquoted into 1.5 ml microcentrifuge tubes and washed thrice with buffer C (20 mM CAPS (pH 11.0), 50 mM NaCl, 12.5% v/v glycerol) with a magnetic rack. Next, the resin was incubated with 7.5 µg ml^{-1} anti-Rubisco on a rotator (2 h/4 °C) and washed twice with buffer C. Controls were performed by not priming the resin with anti-Rubisco antibodies. A total of 750 µl of soluble lysate was then added to anti-Rubisco-bound resin, incubated on a rotator overnight (~18 h/4 °C) and washed thrice with buffer C. Protein elution was carried out by adding 80 µl of 2.5× SDS loading buffer, without 2-mercaptoethanol. Eluted proteins were separated from the Dynabeads, and 2-mercaptoethanol was added to a final concentration of 100 mM; the proteins were then boiled (95 °C/5 min) and shipped for liquid chromatography–tandem mass spectrometry analysis. The co-IP and control experiments each had five technical replicates.

Proteomic analysis

Liquid chromatography–tandem mass spectrometry analysis was performed at the Environmental Molecular Sciences Laboratory at the Pacific Northwest National Laboratory. The samples were processed using Filter Aided Sample Preparation[67] by adding 400 µl of 8 M urea to 30 K molecular weight cut-off Filter Aided Sample Preparation spin columns along with 40 µl of the sample in SDS BME buffer and centrifuged at 14,000 *g* for 20 min. Urea washes were repeated three additional times followed by the addition 400 µl of 50 mM ammonium bicarbonate, pH 8.0, and two repeated centrifugations for 20 min. The columns were then placed into clean and labelled collection tubes. The digestion solution was made by dissolving 5 µg of trypsin in 75 µl of 50 mM ammonium bicarbonate solution, which was added to each sample. The samples were then incubated for 3 h at 37 °C with 600 rpm shaking on a thermomixer with a ThermoTop (Eppendorf) to reduce condensation into the caps of the collection tubes. The resultant peptides were then centrifuged through the filter and into the collection tube by centrifuging at 14,000 *g* for 15 min. The peptides were concentrated to ~30 µl using a vacuum concentrator. Final peptide concentrations were determined using a bicinchoninic acid assay (Thermo Scientific), and each sample was prepared at 0.1 µg $µl^{-1}$ for MS analysis

Digested protein samples were analysed using an Orbitrap Eclipse Tribrid MS (Thermo Scientific) outfitted with a high-field asymmetric-waveform ion mobility spectrometry (FAIMS) interface, using data-dependent acquisition mode. Peptides were ionized using a voltage of 2.4 kV and with an ion transfer tube temperature of 300 °C. The data acquisition time was 2 h following a 20 min delay to avoid dead time between the injection and elution of peptides. A proprietary method for transferring identification on the basis of FAIMS filtering was used to fractionate ionized peptides via the FAIMSpro interface using a 3-compensation voltage; −45, −60, −75 V method. Fractionated ions with a mass range of 400–1,800 m/z were scanned with Orbitrap at 120,000 resolution with an injection time of 50 ms and an automatic gain control target of 4×10^5. Cycle times of 1.0 s were used for the 3-compensation voltage method. Precursor ions with intensities $>1 \times 10^4$ were fragmented with an isolation window of 0.7 by 30% higher-energy collisional dissociation energy and scanned with an automatic gain control target of 1×10^4 and an injection time of 35 ms.

The raw data files were referenced to *A. agrestis* nuclear encoded and chloroplast encoded proteins[21], and peptide abundances were extracted from the raw spectra using MASIC v.3.2.8152 (ref. 68) and $\log_2$-transformed to remove skewness in the distribution of measured abundances. The transformed abundance values were then normalized using the mean central tendency method implemented in

InfernoR v.1.1.7995 (ref. 69). The normalized peptide abundances were de-logged, summed, transformed (log_2) and normalized again in InfernoR to produce normalized abundances for the protein-level roll-up. Proteins that were missing in more than one replicate of the conditions (control or anti-RbcL) were filtered from the final analysis to limit the imputation of too many missing values. Left-censored missing values were imputed using the minimum probability method with the default parameters. Differential enrichment analysis was conducted using the DEP Bioconductor package v.1.27.0 (ref. 70) with a *P*-value cut-off of 0.05. The plots were generated using ggplot2 v.3.5.1 (ref. 71) and further polished using Adobe Illustrator v.27.1.1. Gene Ontology enrichment analysis was carried out using the GO Enrichment module of TBtools v.2.096 (ref. 72) with goslim_plant selected. The background file was set as the entire *A. agrestis* proteome, and proteins found to be significantly enriched by co-IP were chosen as the selection set. The resulting table was then used to generate an enrichment bar plot.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Newly generated proteomics data have been deposited in MassIVE, under accession no. MSV000095322. The gene expression data can be found in the NCBI Sequence Read Archive under BioProject no. PRJNA996135. The protein structure used for aiding Rubisco antibody design can be found in the Protein Data Bank (accession no. 2V63).

References

1. Bar-On, Y. M. & Milo, R. The global mass and average rate of Rubisco. *Proc. Natl Acad. Sci. USA* **116**, 4738–4743 (2019).
2. Ellis, R. J. The most abundant protein in the world. *Trends Biochem. Sci.* **4**, 241–244 (1979).
3. Leegood, R. C. A welcome diversion from photorespiration. *Nat. Biotechnol.* **25**, 539–540 (2007).
4. Long, S. P., Zhu, X.-G., Naidu, S. L. & Ort, D. R. Can improvement in photosynthesis increase crop yields? *Plant Cell Environ.* **29**, 315–330 (2006).
5. Griffiths, H., Meyer, M. T. & Rickaby, R. E. M. Overcoming adversity through diversity: aquatic carbon concentrating mechanisms. *J. Exp. Bot.* **68**, 3689–3695 (2017).
6. Badger, M. R. & Price, G. D. CO_2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *J. Exp. Bot.* **54**, 609–622 (2003).
7. Mackinder, L. C. M. et al. A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. *Proc. Natl Acad. Sci. USA* **113**, 5958–5963 (2016).
8. Freeman Rosenzweig, E. S. et al. The eukaryotic CO_2-concentrating organelle is liquid-like and exhibits dynamic reorganization. *Cell* **171**, 148–162.e19 (2017).
9. Wunder, T., Cheng, S. L. H., Lai, S.-K., Li, H.-Y. & Mueller-Cajar, O. The phase separation underlying the pyrenoid-based microalgal Rubisco supercharger. *Nat. Commun.* **9**, 5076 (2018).
10. He, S., Crans, V. L. & Jonikas, M. C. The pyrenoid: the eukaryotic CO_2-concentrating organelle. *Plant Cell* **35**, 3236–3259 (2023).
11. Ramazanov, Z. et al. The induction of the CO–-concentrating mechanism is correlated with the formation of the starch sheath around the pyrenoid of *Chlamydomonas reinhardtii*. *Planta* **195**, 210–216 (1994).
12. Itakura, A. K. et al. A Rubisco-binding protein is required for normal pyrenoid number and starch sheath morphology in *Chlamydomonas reinhardtii*. *Proc. Natl Acad. Sci. USA* **116**, 18445–18454 (2019).
13. Fei, C., Wilson, A. T., Mangan, N. M., Wingreen, N. S. & Jonikas, M. C. Modelling the pyrenoid-based CO_2-concentrating mechanism provides insights into its operating principles and a roadmap for its engineering into crops. *Nat. Plants* **8**, 583–595 (2022).
14. McGrath, J. M. & Long, S. P. Can the cyanobacterial carbon-concentrating mechanism increase photosynthesis in crop species? A theoretical analysis. *Plant Physiol.* **164**, 2247–2261 (2014).
15. Yin, X. & Struik, P. C. Can increased leaf photosynthesis be converted into higher crop mass production? A simulation study for rice using the crop model GECROS. *J. Exp. Bot.* **68**, 2345–2360 (2017).
16. Irisarri, I. et al. Unexpected cryptic species among streptophyte algae most distant to land plants. *Proc. R. Soc. B* **288**, 20212168 (2021).
17. Vaughn, K. C., Campbell, E. O., Hasegawa, J., Owen, H. A. & Renzaglia, K. S. The pyrenoid is the site of ribulose 1,5-bisphosphate carboxylase/oxygenase accumulation in the hornwort (Bryophyta: Anthocerotae) chloroplast. *Protoplasma* **156**, 117–129 (1990).
18. Smith, E. & Griffiths, H. A pyrenoid-based carbon-concentrating mechanism is present in terrestrial bryophytes of the class Anthocerotae. *Planta* **200**, 203–212 (1996).
19. Li, F.-W., Villarreal Aguilar, J. C. & Szövényi, P. Hornworts: an overlooked window into carbon-concentrating mechanisms. *Trends Plant Sci.* **22**, 275–277 (2017).
20. Vaughn, K. C. et al. The anthocerote chloroplast: a review. *N. Phytol.* **120**, 169–190 (1992).
21. Li, F.-W. et al. *Anthoceros* genomes illuminate the origin of land plants and the unique biology of hornworts. *Nat. Plants* **6**, 259–272 (2020).
22. Szövényi, P. et al. Establishment of *Anthoceros agrestis* as a model species for studying the biology of hornworts. *BMC Plant Biol.* **15**, 98 (2015).
23. Lafferty, D. J. et al. Biolistics-mediated transformation of hornworts and its application to study pyrenoid protein localization. *J. Exp. Bot.* **75**, 4760–4771 (2024). erae243.
24. Schafran, P. et al. Pan-phylum genomes of hornworts reveal conserved autosomes but dynamic accessory and sex chromosomes. *Nat. Plants* (in the press).
25. Barrett, J., Girr, P. & Mackinder, L. C. M. Pyrenoids: CO_2-fixing phase separated liquid organelles. *Biochim. Biophys. Acta Mol. Cell. Res.* **1868**, 118949 (2021).
26. Oh, Z. G. et al. A linker protein from a red-type pyrenoid phase separates with Rubisco via oligomerizing sticker motifs. *Proc. Natl Acad. Sci. USA* **120**, e2304833120 (2023).
27. Burr, F. A. Phylogenetic transitions in the chloroplasts of the Anthocerotales. I. The number and ultrastructure of the mature plastids. *Am. J. Bot.* **57**, 97–110 (1970).
28. Mackinder, L. C. M. et al. A spatial interactome reveals the protein organization of the algal CO_2-concentrating mechanism. *Cell* **171**, 133–147.e14 (2017).
29. Mukherjee, A. et al. Thylakoid localized bestrophin-like proteins are essential for the CO_2 concentrating mechanism of *Chlamydomonas reinhardtii*. *Proc. Natl Acad. Sci. USA* **116**, 16915–16920 (2019).
30. Ignatova, L., Rudenko, N., Zhurikova, E., Borisova-Mubarakshina, M. & Ivanov, B. Carbonic anhydrases in photosynthesizing cells of C_3 higher plants. *Metabolites* **9**, 73 (2019).
31. Badger, M. R. & Price, G. D. The role of carbonic anhydrase in photosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **45**, 369–392 (1994).
32. Almagro Armenteros, J. J. et al. Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance* **2**, e201900429 (2019).

33. Jethva, J. et al. Realisation of a key step in the evolution of C_4 photosynthesis in rice by genome editing. Preprint at *bioRxiv* https://doi.org/10.1101/2024.05.21.595093 (2024).
34. Smith, E. C. & Griffiths, H. The role of carbonic anhydrase in photosynthesis and the activity of the carbon-concentrating-mechanism in bryophytes of the class Anthocerotae. *N. Phytol.* **145**, 29–37 (2000).
35. Sinetova, M. A., Kupriyanova, E. V., Markelova, A. G., Allakhverdiev, S. I. & Pronina, N. A. Identification and functional role of the carbonic anhydrase Cah3 in thylakoid membranes of pyrenoid of *Chlamydomonas reinhardtii*. *Biochim. Biophys. Acta* **1817**, 1248–1255 (2012).
36. Mitra, M. et al. The carbonic anhydrase gene families of *Chlamydomonas reinhardtii*. *Can. J. Bot.* **83**, 780–795 (2005).
37. Terentyev, V. V. & Shukshina, A. K. CAH3 from *Chlamydomonas reinhardtii*: unique carbonic anhydrase of the thylakoid lumen. *Cells* **13**, 109 (2024).
38. Wang, Y. & Spalding, M. H. LCIB in the *Chlamydomonas* CO_2-concentrating mechanism. *Photosynth. Res.* **121**, 185–192 (2014).
39. Jin, S. et al. Structural insights into the LCIB protein family reveals a new group of β-carbonic anhydrases. *Proc. Natl Acad. Sci. USA* **113**, 14716–14721 (2016).
40. Yamano, T., Toyokawa, C., Shimamura, D., Matsuoka, T. & Fukuzawa, H. CO_2-dependent migration and relocation of LCIB, a pyrenoid-peripheral protein in *Chlamydomonas reinhardtii*. *Plant Physiol.* **188**, 1081–1094 (2021).
41. Chou, M.-L. et al. Tic40, a membrane-anchored co-chaperone homolog in the chloroplast protein translocon. *EMBO J.* **22**, 2970–2980 (2003).
42. Barrett, J. et al. A promiscuous mechanism to phase separate eukaryotic carbon fixation in the green lineage. Preprint *bioRxiv* https://doi.org/10.1101/2024.04.09.588658 (2024).
43. Moromizato, R. et al. Pyrenoid proteomics reveals independent evolution of the CO_2-concentrating organelle in chlorarachniophytes. *Proc. Natl Acad. Sci. USA* **121**, e2318542121 (2024).
44. Aigner, H. et al. Plant RuBisCo assembly in *E. coli* with five chloroplast chaperones including BSD2. *Science* **358**, 1272–1278 (2017).
45. Feiz, L. et al. Ribulose-1,5-bis-phosphate carboxylase/oxygenase accumulation factor1 is required for holoenzyme assembly in maize. *Plant Cell* **24**, 3435–3446 (2012).
46. Saschenbrecker, S. et al. Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell* **129**, 1189–1200 (2007).
47. Wang, L. et al. A chloroplast protein atlas reveals punctate structures and spatial organization of biosynthetic pathways. *Cell* **186**, 3499–3518.e14 (2023).
48. Badger, M. R. et al. The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO_2-concentrating mechanisms in algae. *Can. J. Bot.* **76**, 1052–1071 (1998).
49. Kevekordes, K. et al. Inorganic carbon acquisition by eight species of *Caulerpa* (Caulerpaceae, Chlorophyta). *Phycologia* **45**, 442–449 (2006).
50. Gee, C. W. & Niyogi, K. K. The carbonic anhydrase CAH1 is an essential component of the carbon-concentrating mechanism in *Nannochloropsis oceanica*. *Proc. Natl Acad. Sci. USA* **114**, 4537–4542 (2017).
51. Steensma, A. K., Shachar-Hill, Y. & Walker, B. J. The carbon-concentrating mechanism of the extremophilic red microalga *Cyanidioschyzon merolae*. *Photosynth. Res.* **158**, 203 (2023).
52. Villarreal, J. C. & Renner, S. S. Hornwort pyrenoids, carbon-concentrating structures, evolved and were lost at least five times during the last 100 million years. *Proc. Natl Acad. Sci. USA* **109**, 18873–18878 (2012).
53. Hanson, D., John Andrews, T. & Badger, M. R. Variability of the pyrenoid-based CO_2 concentrating mechanism in hornworts (Anthocerotophyta). *Funct. Plant Biol.* **29**, 407–416 (2002).
54. Atkinson, N., Mao, Y., Chan, K. X. & McCormick, A. J. Condensation of Rubisco into a proto-pyrenoid in higher plant chloroplasts. *Nat. Commun.* **11**, 6303 (2020).
55. Atkinson, N., Stringer, R., Mitchell, S. R., Seung, D. & McCormick, A. J. SAGA1 and SAGA2 promote starch formation around proto-pyrenoids in *Arabidopsis* chloroplasts. *Proc. Natl Acad. Sci. USA* **121**, e2311013121 (2024).
56. Kaste, J. A. M., Walker, B. J. & Shachar-Hill, Y. Reaction–diffusion modeling provides insights into biophysical carbon concentrating mechanisms in land plants. *Plant Physiol.* **196**, 1374–1390 (2024).
57. Gunadi, A., Li, F.-W. & Van Eck, J. Accelerating gametophytic growth in the model hornwort *Anthoceros agrestis*. *Appl. Plant Sci.* **10**, e11460 (2022).
58. Villarreal, J. C., Duckett, J. G. & Pressel, S. Morphology, ultrastructure and phylogenetic affinities of the single-island endemic *Anthoceros cristatus* Steph. (Ascension Island). *J. Bryol.* **39**, 226–234 (2017).
59. Sauret-Gueto, S. et al. Systematic tools for reprogramming plant gene expression in a simple model, *Marchantia polymorpha*. *ACS Synth. Biol.* **9**, 864–882 (2020).
60. R Core Team. R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, 2023).
61. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
62. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
63. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
64. Newman, A. M. & Cooper, J. B. XSTREAM: a practical algorithm for identification and architecture modeling of tandem repeats in protein sequences. *BMC Bioinform.* **8**, 382 (2007).
65. Gasteiger, E. et al. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–3788 (2003).
66. Romero, P. et al. Sequence complexity of disordered protein. *Proteins* **42**, 38–48 (2001).
67. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).
68. Monroe, M. E., Shaw, J. L., Daly, D. S., Adkins, J. N. & Smith, R. D. MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features. *Comput. Biol. Chem.* **32**, 215–217 (2008).
69. Polpitiya, A. D. et al. DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* **24**, 1556–1558 (2008).
70. Zhang, X. et al. Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nat. Protoc.* **13**, 530–550 (2018).
71. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. (Springer-Verlag, 2016).
72. Chen, C. et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **13**, 1194–1202 (2020).

Acknowledgements

This work is supported by National Science Foundation grant no. MCB-2213841 to F.-W.L. and grant no. MCB-2213840 to L.H.G., an Environmental Molecular Sciences Laboratory User Grant to F.-W.L., a Triad Foundation Grant to F.-W.L. and a Schmittau-Novak Graduate Student Grant to T.A.R. We thank A. Skirycz, K. Eshenour, A. Hotto and

D. Stern of Boyce Thompson Institute for providing access to tools and reagents to establish preliminary lysis and co-IP experiments. We thank R. Key at the University of Florida for the illustration in Fig. 6. We also thank C. Nicora and N. Tolic from the Environmental Molecular Sciences Laboratory at the Pacific Northwest National Laboratory for technical assistance with mass spectrometry processing, M. Srivastava at the Boyce Thompson Institute Plant Cell Imaging Center for technical assistance with confocal imaging using the Leica TCS SP5 Laser Scanning Confocal Microscope and A. Roeder at Cornell University for providing access to the Zeiss LSM710 and Leica Stellaris 5 confocal microscopes. Finally, we thank the York Physics of Pyrenoids research community for feedback and Li and Gunn lab members for discussions.

Author contributions

T.A.R., L.H.G. and F.-W.L. conceived the project. T.A.R. made the gene constructs and carried out the confocal imaging. T.A.R., D.L. and X.X. performed the hornwort transformation. Z.G.O. designed the Rubisco antibody and optimized the lysis of hornwort thallus. T.A.R. and Z.G.O. performed the co-IP experiments. J.C.A.V. performed TEM. T.A.R., Z.G.O., L.H.G. and F.-W.L. analysed the data. T.A.R. and F.-W.L. wrote the manuscript with contributions and comments from all authors. L.H.G. and F.-W.L. secured the funding and supervised the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41477-024-01871-0.

Correspondence and requests for materials should be addressed to Laura H. Gunn or Fay-Wei Li.

Peer review information *Nature Plants* thanks Nicky Atkinson, Manon Demulder and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Fay-Wei Li and Laura Gunn

Last updated by author(s): Oct 30, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We used the LEICA Application Suite (LAS) X v4.6.1.27508 to acquire the confocal images, and the built-in LAS FRAP Wizard to collect the FRAP data. To extract fluorescence intensity data, we used FIJI v2.14.0.
Data analysis	We used Orthofinder v2.5. 4 to generate orthogroups (see supplementary table 2 for complete list of genomes). We used MASIC v3.2.8152, InfernoR vl.l.7995, DEP vl.27.0, and TBtools v2.096 for proteomic analysis. We used XSTREAM vl.73, Expasy web server, and PONDR web server to search for disordered proteins. We used ggplot2 v3.5.l to produce graphs, which were further polished using Adobe Illustrator v27.l.l. We used Iqtree v2.2.0 to reconstruct phylogenetic trees. No custom algorithm or pipeline was developed in this study.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Newly generated proteomics data were deposited in MassIVE, with under the accession MSV000095322. Gene expression data can be found in the NCBI Sequence Read Archive under the BioProject PRJNA996135. The protein structure used for aiding Rubisco antibody design can be found in the Protein Data Bank (accession: 2V63).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The co-IP and control experiments each had five technical replicates. For fluorescent protein tagging and FRAP experiments, we aimed for at least 10 biological replicates. Overall, our goal was to obtain as largest sample size as possible given our budgetary and time constraints.
Data exclusions	No data were excluded
Replication	All the experiments (including fluorescent protein tagging, Co-IP, and FRAP) were replicated. Additional protein localization images, collected from separate transformation events, were provided in the supplementary data.
Randomization	For the Co-IP experiment, the plant materials used for the control and antibody treatment came from the same batch of culture, and were randomly divided into two groups. For fluorescent protein tagging and FRAP, transformants were generated randomly by biolistics, and chosen for further analyses based on the ease of obtaining high quality images (i.e. cell orientation and depth in the tissue).
Blinding	Blinding cannot be done because the experiments and data analyses were done by the same investigators.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☐	☒ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used: The Rubisco Antibody used in this study was a customized rabbit polyclonal antibody raised against a synthetic peptide of Anthoceros agrestis Rbcl (-terminus (EVWKEIKFVFETIDTL). Affinity purification of the antibody was also performed before shipment (Life Technologies, USA).

Validation: Purified antibody was titered by indirect ELISA against the synthetic peptide bound to the solid-phase to measure the reactivity of the antibodies (Life Technologies, USA). We validated the antibody against Anthoceros agrestis Rubsico lysate prior to performing coimmunoprecipitation
experiments.

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes
☒	☐ Public health
☒	☐ National security
☒	☐ Crops and/or livestock
☒	☐ Ecosystems
☒	☐ Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes
☒	☐ Demonstrate how to render a vaccine ineffective
☒	☐ Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐ Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐ Increase transmissibility of a pathogen
☒	☐ Alter the host range of a pathogen
☒	☐ Enable evasion of diagnostic/detection modalities
☒	☐ Enable the weaponization of a biological agent or toxin
☒	☐ Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	The wild type Anthoceros agrestis Oxford strain was originally established and described in https://doi.org/10.1186/s12870-015-0481-x.
Novel plant genotypes	To examine protein subcellular localization, biolistics-mediated transformation was done to either transiently or stably express fluorescently tagged proteins.
Authentication	Successful transformation was determined primarily based on fluorescent protein signals. In the case of stable transformation, we also relied on antibiotics selection and PCR verification.