

Rapid plastid isolation reveals the chloroplast proteome and structures of the chlororibosome large subunit and RuBisCO in *Marchantia polymorpha*

Parth K. Raval^{1†}, Cristopher Mitchell^{2,3†}, Maria Lozano-Quiles¹, Sarah O’Keefe^{2,3}, Tuula A. Nyman⁴,
Brendan Battersby², Sarah J. Butcher^{2,3}, Sven B. Gould^{1,5}

¹Institute for Molecular Evolution, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany

²HiLIFE-Institute of Biotechnology, University of Helsinki, Helsinki, Finland

³Department of Molecular and Integrative Biosciences, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

⁴Department of Immunology, Institute of Clinical Medicine, University of Oslo and Oslo University Hospital, Oslo, Norway

⁵Institute for Population Genetics, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany

† Equal contribution

*Corresponding author: gould@hhu.de

Abstract

Plastids house the biology of eukaryotic photosynthesis. The majority of a plastid’s proteome is imported after cytosolic translation, but a few dozen proteins on average remain organelle-encoded, translated by the plastid’s own ribosomes. While 1000s of plastid genomes have been sequenced, the availability of less than ten proteomes and only two species with full 70S plastid ribosomal structures limit our understanding of land plant evolution. To address this, we optimized a protocol for the rapid isolation of *Marchantia polymorpha* chloroplasts that provides a highly enriched and intact organelle fraction from gradient volumes as little as 2 mL. Our approach was successfully applied to six other species, including *Chlamydomonas reinhardtii* and *Nicotiana tabacum*. Focusing on *M. polymorpha*, we determined the proteome of the chloroplast fraction, identifying 1337 nuclear-encoded proteins with a high confidence, where 83% belong to orthologs shared with angiosperms. We further isolated large protein complexes by RNA affinity purification using poly-lysine and provide the high-resolution structures of the 50S subunit of the chloroplast ribosome and RuBisCO from this bryophyte using cryogenic EM and image reconstruction to 2.23 and 2.12 Å resolution, respectively, highlighting the structural conservation of both complexes. For chloroplasts, our data show that the genome reduction event experienced by the common ancestor of bryophytes has had little impact on the organelle’s complexity and that they underscore a high level of structural conservation of core components of plastid biology. Our data provide novel resources and methods to explore the functional evolution of plastid proteomes and major macromolecular complexes of cyanobacterial origin.

Keywords: bryophytes; chloroplast; proteome; 50S ribosome; RuBisCO; cryo-EM

Introduction

The transition from a cyanobacterial endosymbiont to the different types of plastids of algae and plants entails the loss and transfer of most of the endosymbiont's genetic information¹⁻⁵. Crop plants such as pea or maize hence encode more than a thousand proteins in their nuclear genome that need to be imported by their plastids after cytosolic translation^{6,7}. Plastid proteomes are complemented by a few proteins that are kept organelle-encoded, maybe due to the need for the co-localisation of genes of the electron transport chain and the bioenergetic membrane itself for functional redox regulation⁸⁻¹⁰, and protein hydrophobicity impairing membrane translocation¹¹⁻¹⁴. Subunits required during the early steps of 70S ribosome assembly are likely retained due to constraints associated with securing proper ribosome maturation¹⁵. Typically, as is the case for the 70S ribosome, plastid-encoded proteins interact with nuclear-encoded ones to form macromolecular complexes such as photosystem I and II, the NADH dehydrogenase complex, the ribulose-1,5-bisphosphate-carboxylase/-oxygenase (RuBisCO) enzyme, and ATP synthase.

A significant event in plastid evolution was the transition from water to land^{16,17}. Plant terrestrialization and the origin of the Embryophyta is characterised by the emergence of the cuticle, stomata, spores and seeds, multiplastidy, the adaptation of stress response networks already present in the streptophyte algal ancestor, a step-wise expansion of genomes, and a general increase in mitochondrial and plastid proteome size¹⁷⁻²³. An early-branching and ecologically important group of land plants are the bryophytes, uniting liverworts, hornworts, and mosses (Fig. 1). After genome expansions at the origin of the Chloroplastida and again at the origin of Streptophyta^{24,25}, the bryophyte ancestor experienced genome reduction that drove the lineage's phenotypic evolution¹⁹. Now, after more than 500 million years, bryophytes are characterized by highly derived genomes that in sum encode a larger variety of gene families than their vascular relatives²⁶.

The positioning of bryophytes on the phylogenetic tree provides an ideal opportunity to explore the trajectory of plastid biology after genome reduction in comparison to vascular plant plastids. Experimentally curated plastid proteomes and high-resolution structures of macromolecular complexes outside of a few angiosperm model systems, however, are scarce (Fig. 1). To address this, we optimized a chloroplast isolation protocol that requires minimal resources and time. It yields intact chloroplasts from a diverse set of species and eases their study, enabling chloroplast biology exploration across a broad variety of taxa. Focusing on the liverwort *Marchantia polymorpha*, we obtained a highly enriched chloroplast proteome, which corroborates previously predicted evolutionary trajectories of land plant plastid proteomes²⁰. With the aim of isolating chlororibosomes, we subjected the plastid fraction to an RNA affinity purification using poly-lysine²⁷. This resulted in the enrichment of multiple macromolecular complexes from the isolated chloroplasts. Using cryogenic electron microscopy (cryo-EM), we resolved the structures of the 50S ribosomal subunit and RuBisCO at respective resolutions of 2.23 and 2.12 Å. Our data provide new resources to study plastid biology and allow us to hypothesize that the chloroplast proteome and ribosome of the bryophyte were not significantly affected by genome reformatting at the origin of the bryophytes.

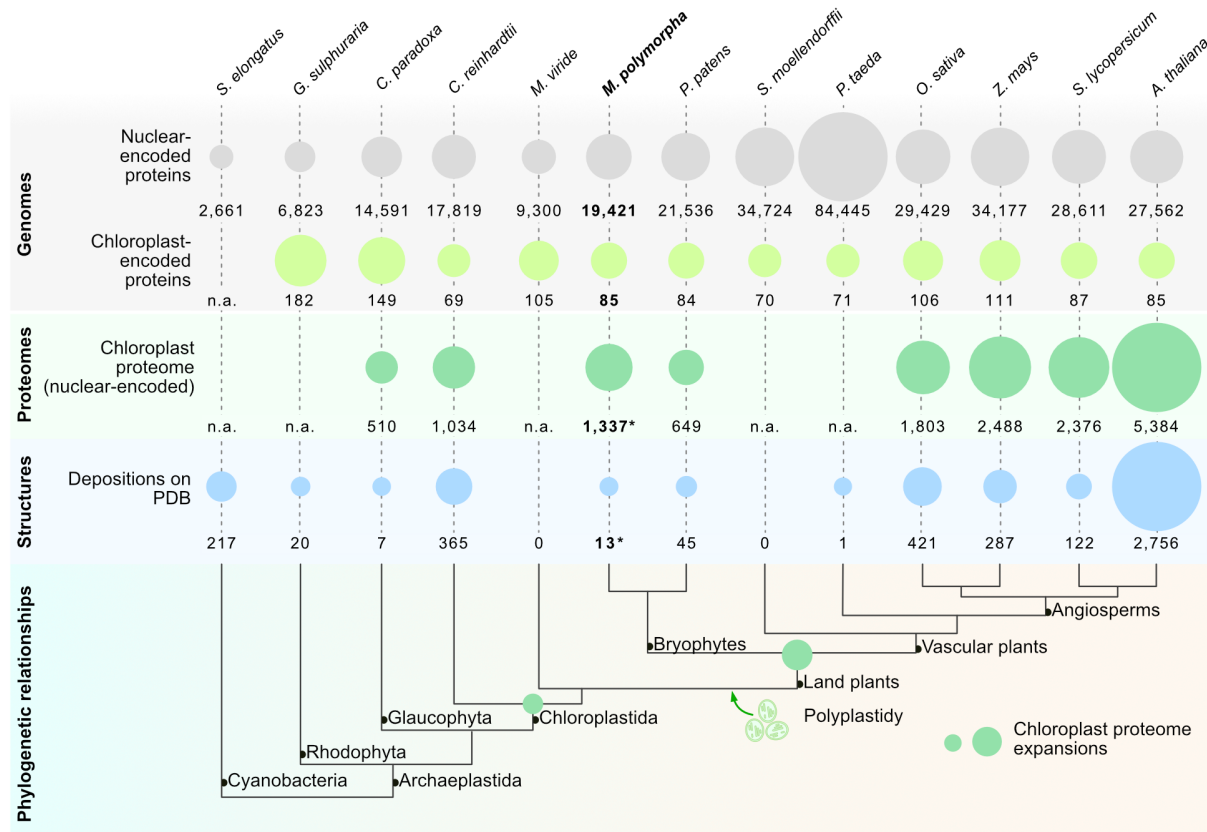


Figure 1: Phylogenetic relationships and key numbers for exemplary species. A cladogram of key species representing the diversity of Archaeplastida lineage, with the number of nuclear- and chloroplast-encoded proteins and the number of proteins in the chloroplast proteome and the number of entries in the World-Wide Protein Data Bank (last accessed on 1st May 2026). Asterisks indicate numbers that also include the current study. Key changes in plastid biology over the course of plant evolution are indicated on the cladogram^{18,20,28,29}. The number of nuclear encoded proteins as per the whole genome models³⁰ and the number of plastid localized proteins from previous studies: *Cyanophora paradoxa*³¹, *C. reinhardtii*^{32,33}, *Physcomitrium patens*³⁴, *Oryza sativa*³⁵, *Zea mays*³⁵, *Solanum lycopersicum*³⁵, and *A. thaliana*³⁶ (N.a. = not available; species from the left: *Synechococcus elongatus*, *Galdieria sulphuraria*, *Cyanophora paradoxa*, *Chlamydomonas reinhardtii*, *Micromonas viride*, *Marchantia polymorpha*, *Physcomitrium patens*, *Selaginella moellendorffii*, *Pinus taeda*, *Oryza sativa*, *Zea mays*, *Solanum lycopersicum*, *Arabidopsis thaliana*).

Results

A rapid mini-prep protocol for the isolation of plastids from diverse species

Plastid isolation protocols are available for very few species and they can be complex and resource intensive. A widely applicable and efficient purification protocol has the potential to accelerate and expand plastid research across diverse plant lineages. To this end, we systematically scaled down a Percoll[®] gradient-based isolation protocol and tested it first on the bryophyte *Marchantia polymorpha*. From as little

as 500 mg of starting tissue material (i.e. only a few, two-week old thalli), we observed the formation of a characteristic dark green band at the Percoll® interface, which was confirmed as clean and intact chloroplasts by microscopy (Fig. 2a). We tested our protocol on standard model systems such as *Arabidopsis thaliana* and *Nicotiana tabacum*, as well as *Moricandia arvensis* (a model for C3-C4 intermediate photosynthesis), *Diploaxis viminea* (a model for stress physiology), and *Erucastrum gallicum* (a model in plant ecology). Our protocol performed equally well across each model system, taking under one hour to complete, from less than 1 g of leaf tissue for each gradient (Fig. 2a). For *Chlamydomonas reinhardtii*, we scaled down a previously reported protocol requiring a three-, rather than a two-, step Percoll gradient³⁷ and found it to be similarly effective for the algal model system (Fig. 2a).

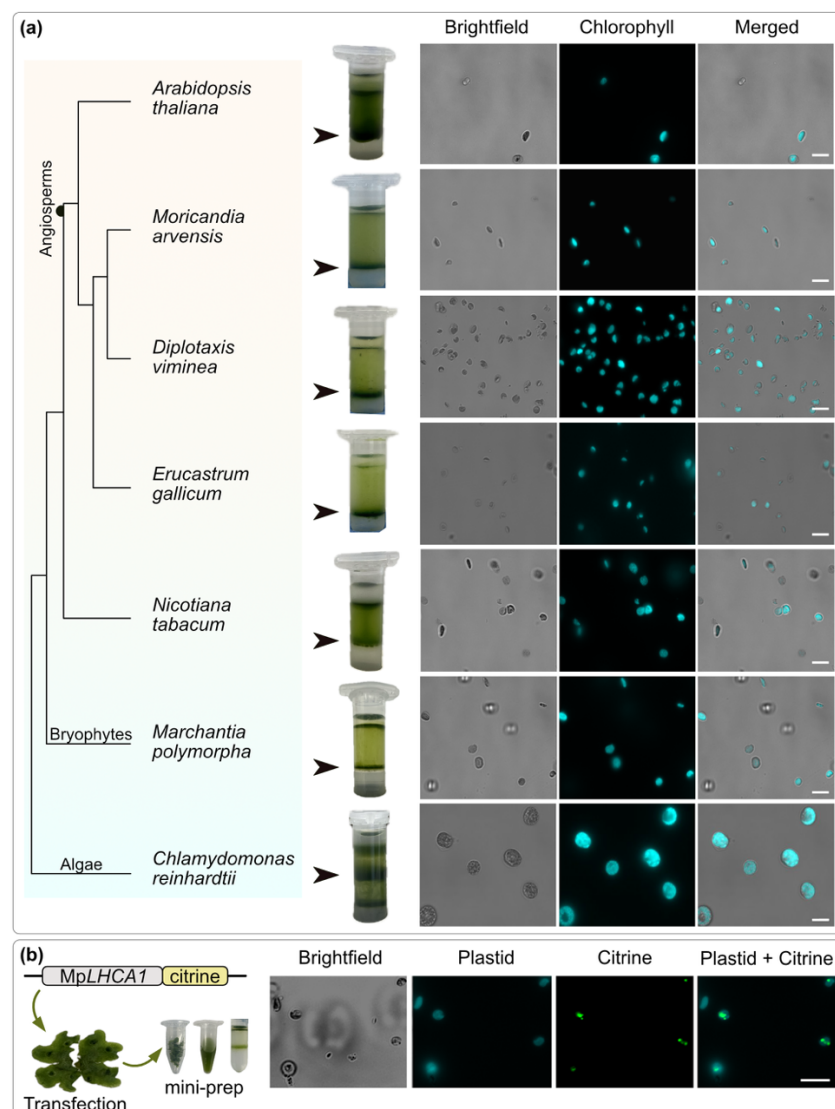


Figure 2: Small-scale plastid isolation from diverse species. (a) Plastids isolated from diverse species using the mini-prep protocol and visualized through their chlorophyll autofluorescence. Scalebars 10µm. **(b)** Chloroplasts isolated from a *M. polymorpha* transfectant line expressing LHCA1:citrine shows colocalization of chloroplasts and the citrine signal.

Subcellular fractionation is the first step in characterizing the proteome of a compartment. After the isolation of the *M. polymorpha* chloroplasts, we assessed the fraction microscopically alongside a transfectant line expressing the citrine-tagged light-harvesting complex A1 protein (LHCA1; Mp2g15420). This showed a clear colocalization of the reporter with autofluorescence and the shape of the purified organelles (Fig. 2b). We further confirmed the purity of the isolated chloroplast fraction by Western blotting

using antibodies specific for established marker proteins of other compartments, such as BiP1 for the endoplasmic reticulum and SHMT for mitochondria (Fig. S1). From biological triplicates of chloroplasts isolated from individually grown plants, tandem mass spectrometry (MS) identified 1337 nuclear- and 43 plastid genome-encoded (Table S1). 61.7% of the nuclear encoded proteins identify a homologue in at least one of the 56 diverse cyanobacteria that were part of our protein database for screening (Fig. S2), thus underscoring the high degree of proteome conservation of this photosynthetic organelle of eukaryotes over a span of more than 1.5 billion years³⁸.

A functional annotation of the chloroplast proteome identified here shows that most proteins play a direct role in photosynthesis, with functions related to chlorophylls, protein synthesis, carbon and amino acid metabolism and chloroplast-specific secondary metabolism pathways such as cofactor and pigment biosynthesis (Fig. 3a). These categories, expected of a chloroplast proteome, speak for the purity of the isolated fraction. The largest single fraction of proteins (184 proteins) returned no clear annotated function in the MarpolBase³⁹ (accessed 1st August 2025). 125 of these 184 proteins had homologues in *A. thaliana* (Table S2A), of which 31 are involved in thylakoid and ribosome biogenesis, RNA processing and localized in thylakoid membranes, but the remaining 94 homologues had no clear functional annotation for *Arabidopsis* either. 38 of these 184 proteins lack homologs in angiosperms and three appear unique to *M. polymorpha*. A screen for known domains identified 13 proteins encoded tetratricopeptide repeats (TPR), two encoded pentatricopeptide repeats (PPR) and 20 encoded RNA related domains, making TPR/PPR and RNA related proteins the single largest fraction among the 184 proteins with an uncertain function (Table S2B). For 13 proteins (Table S2C), however, no functional domains could be predicted but we found a match for two of them by predicting the structures with AlphaFold3 and using Dali^{40,41} to search for structural homologues (Mp3g01370 aligned with PsbQ and Mp5g18410 with an AFMP4P virulence factor), leaving eleven plastid proteins that require experimental investigation.

Lineage-specific gene family expansions contribute to the *M. polymorpha* plastid proteome

Comparative proteomics is an important component in understanding organelle evolution and the contribution of the compartment to the ecophysiology of a plant. Ancestral state reconstruction of mitochondrial and plastid proteomes across a billion years of chloroplastal evolution recently underscored the expansion of organelle proteomes from algae to angiosperms²⁰. It predicted an intermediate chloroplast proteome size for *M. polymorpha* and we find that 85.6% of the experimentally-curated nuclear encoded chloroplast proteins were predicted based on orthology with previously experimentally characterized proteins (Fig. 3b). Apart from corroborating the identity of individual chloroplast proteins, it also shows that an orthology-based approach to predict an organelle proteome has merit for plotting broader evolutionary patterns in the absence of an experimentally-curated proteome. Hence, we clustered the genomes of 57 diverse chloroplastal species (Table S3) into orthogroups following the approach used previously²⁰ and focused on 4,797 plastid orthogroups (POGs). The comparisons of the POGs and of the experimental proteome recapitulates the expansion of the chloroplast proteome from algae to angiosperm²⁰, with that of *M. polymorpha* – by sheer number of POGs – ranging in the middle (Fig. 1, Fig. S3).

For 2857 POGs, a representative protein has been experimentally localized to the plastid of at least *M. polymorpha*, *A. thaliana*, or *C. reinhardtii* (Fig. 3c). 897 POGs are shared by both *A. thaliana* and *M. polymorpha*, but no protein from the *C. reinhardtii* genome was present in these POGs (Fig. 3c), suggesting they were likely gained or assigned to the chloroplasts of the Embryophyta ancestor (e.g. after

153 duplications) and retained by *Marchantia*. The latter would have happened in the background of a global
154 genome reduction event postulated for the bryophyte ancestor¹⁹. 48.9% of these POGs contain fewer gene
155 members in *M. polymorpha* than in *A. thaliana*, as expected from ancestral genome reduction. 41.2% of
156 POGs, however, contain the same number of genes in *A. thaliana* and *M. polymorpha* and 9.8% (88 POGs;
157 largely associated with secondary metabolism and redox balance) contain more genes in *M. polymorpha*
158 (Fig. 3c), the latter hinting at lineage-specific expansions. Comparisons of POGs across 57 species also
159 identifies 341 POGs that are absent in more than 90% of vascular plants and present only in algae or
160 bryophytes (Table S3). 91% of these POGs remain uncharacterized and they likely contain lineage-specific
161 expansions that helped bryophytes to adapt to a terrestrial lifestyle in ways that differ from tracheophytes.

162 Even for those categories that are not uniquely expanded, such as those of translation, redox balance or
163 detoxification (Fig. 3a), their evolution is informative. For instance, in angiosperms the NDH super
164 complex is stabilized by PSI linkers (i.e. LHCA5 and LHCA6). This NDH-PSI interaction is crucial for
165 photoprotection and cyclic electron flow under diverse light conditions encountered by land plants⁴²⁻⁴⁵.
166 Based on the consistent distribution of the linker subunits across angiosperms, as compared to their sporadic
167 presence outside of angiosperms, a stepwise evolutionary acquisition of these linkers has been proposed⁴⁶⁻
168 ⁴⁹. We identified peptides for the baseline subunits of the NDH complex such as subcomplex A and E, and
169 the core LHCA1-4 subunits, while no linkers were detected (Table S1). As such, our observation supports
170 the view that a monomeric NDH complex preceded the NDH-PSI super complex. Future structural studies
171 on the liverwort's NDH complex and an interpretation in light of secondary genome reduction in bryophytes
172 is hence key in judging the credibility of a monomer first model.

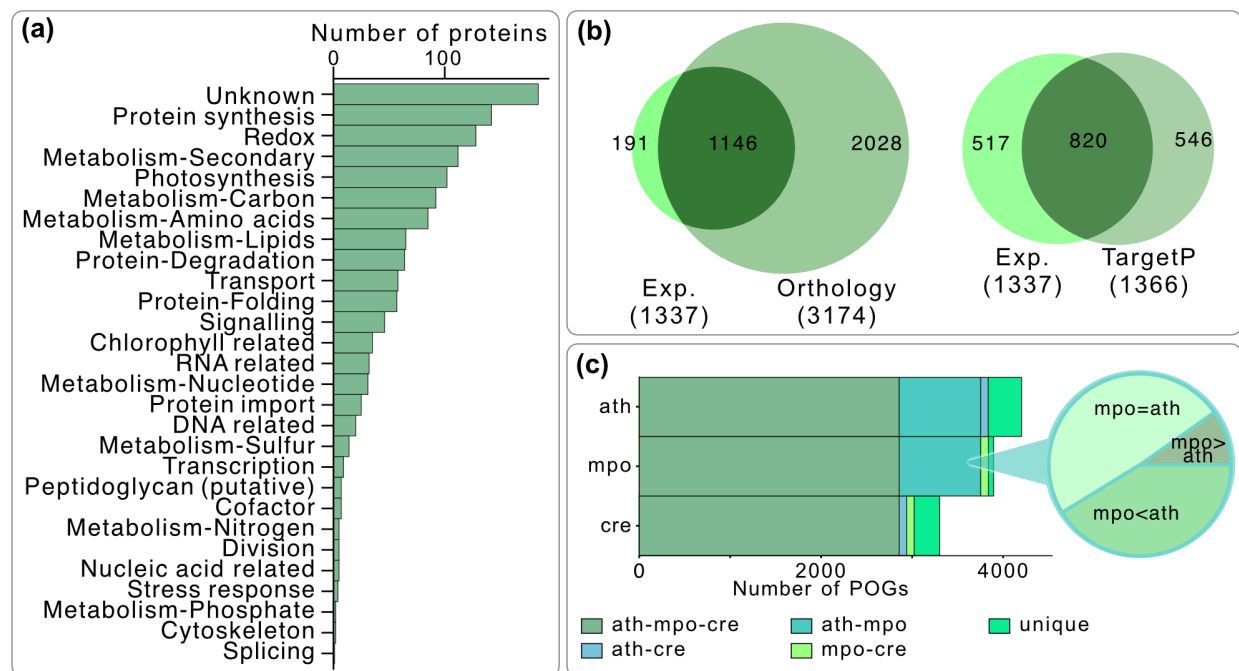


Figure 3: The *M. polymorpha* plastid proteome gained almost as many proteins since genome remodeling at the origin of bryophytes as it had lost. (a) Distribution of the identified plastid proteins across functional categories. (b) Comparison of the experimentally-curated *Marchantia* plastid proteome with bioinformatically predicted orthology groups⁵⁰, and proteins predicted to be plastid targeted by TargetP v2.0 (c) Comparisons of POGs of *A. thaliana* (ath), *M. polymorpha* (mpo) and *C. reinhardtii* (cre), with each stacked region representing the number of shared POGs. Of the 897 POGs shared between *A. thaliana* and *M. polymorpha*, half are represented by a similar number of proteins

per group, while about 40% are reduced and 10% are expanded in the number of genes in *M. polymorpha* in comparison to *A. thaliana* (pie chart on the right).

The structures of the 50S ribosomal subunit and RuBisCO demonstrate a high level of macromolecular complex conservation

With the aim of determining a chlororibosome structure for *M. polymorpha*, we subjected the purified organelles to RNA affinity purification by poly-lysine (RAPPL)²⁷. The eluate of this purification was used for mass spectrometry, which showed an enrichment of all chlororibosome subunits, the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) enzyme, and proteins of PSII, PSI, and the ATP synthase (Table S4). The eluate was used for cryo-EM grid preparation and subsequent single-particle reconstruction. 2D classification revealed several large protein complexes including recognizable views of the chlororibosomal large 50S subunit (cLSU), RuBisCO, the soluble F1-ATPase complex, and a chaperone (Fig. S4). Of these, there were sufficient varied views of RuBisCO and the 50S for three-dimensional reconstruction. No particle classes containing the full 70S or the small 30S ribosomal subunit were identified.

We reconstructed high-resolution density maps of the cLSU and RuBisCO to respective resolutions of 2.23 and 2.12 Å, reported at the 0.143 Fourier shell correlation cut-off⁵¹ (Fig. 4a & 5a, S4-5). These maps were used for atomic modeling using previously published sequences⁵²⁻⁵⁸. The ribosomal protein (RP) structures were first predicted using AlphaFold3⁴¹ and then fitted and refined into the experimental map. The rRNA was mutated from a *Spinacia oleracea* model (PDB ID 5MLC⁵⁹). The quality of the density map allowed for modeling 6 rRNA modifications in the cLSU 23S rRNA chain (Fig. S6). Out of the 6 modifications, 5 were identified previously in *Chlamydomonas* and *Escherichia coli* structures, whereas the remaining modification, U579 to 5-methyluridine, appears to be novel to *Marchantia* (Fig. S6)⁶⁰⁻⁶². Alignment of the spinach 50S to the *Marchantia* 50S revealed only 5 subunits out of 31 with a root mean square deviation (RMSD) above 3.0 Å. These were L5, L13, L21, L33, and PSRP6, with RMSDs of 3.7, 8.1, 3.1, 3.2, and 4.5 Å respectively. The values reported for L5 and L33 are due to variation in the position of a single small loop region in each. The main difference between PSRP6's across the structures is that 9 more C-terminal residues were modeled in the *Marchantia* model, however these residues are also present in the spinach PSRP6 sequence, so this represents a difference in flexibility or map quality. The main noticeable difference between the *Marchantia* and spinach models was found at the surface near the exit tunnel, where RPs L13, L21, and L22 are present (Fig. 4b). In *S. oleracea*, these RPs have regions at their N- and C-termini that stick out from the surface and they were hypothesized to form an interface with the thylakoid membrane⁵⁹.

Comparisons of structure and sequence suggest that the C-terminal region was extended (e.g. L21) or gained (e.g. in L22) by the ancestor of eudicots, since they are absent or shorter in other chloroplastidial species, including *M. polymorpha* (Fig. 4b). The N-terminal sequence of L13 is present in the annotation of the *Marchantia* protein but does not appear in the density map (Fig. 4b). Overall, while the amino acid sequences for chloroplastic RPs between the *S. oleracea* and *M. polymorpha* 50S share on average 68% identity, they exhibit highly similar structures through conserved folds (Table S5).

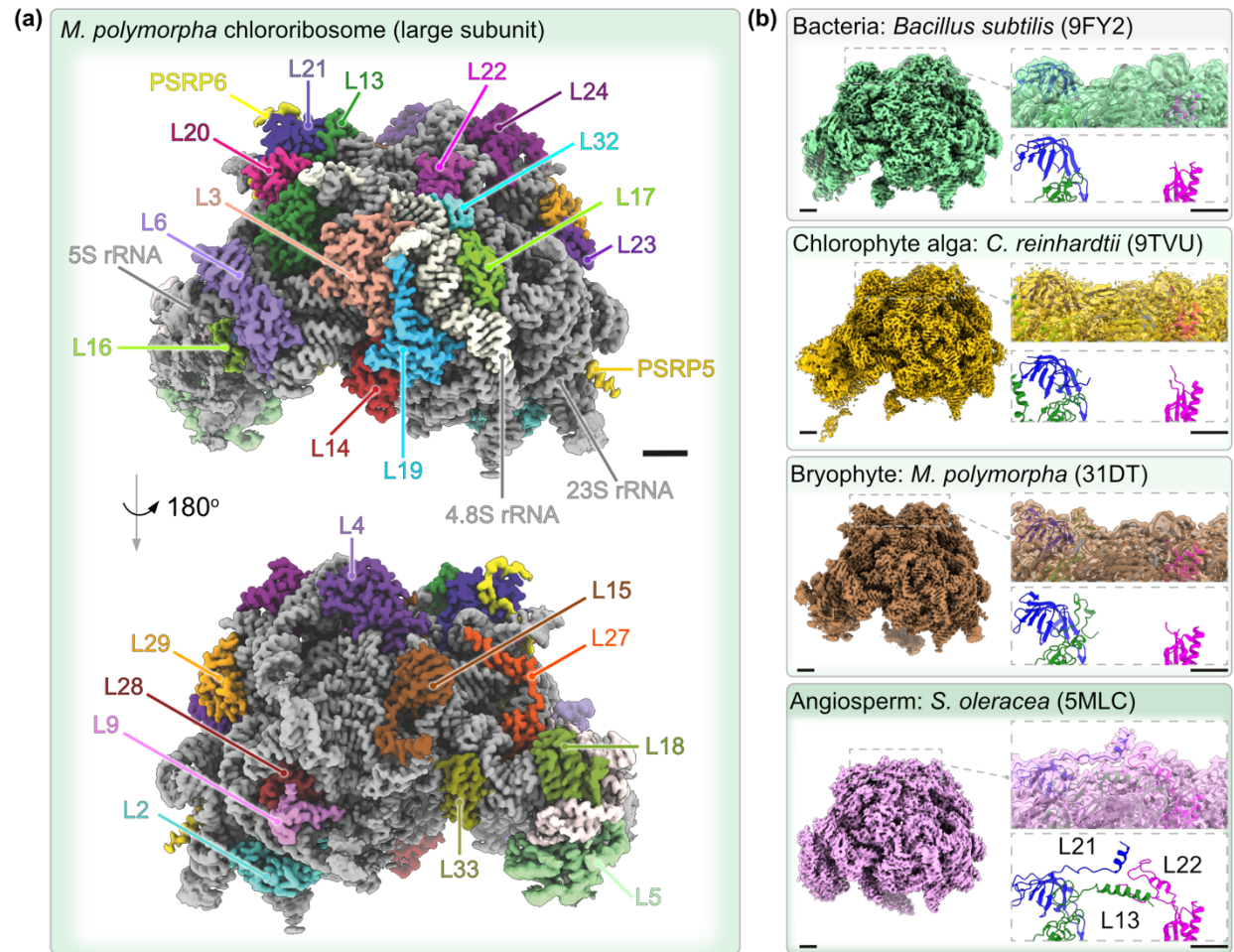


Figure 4: The *M. polymorpha* chlororibosome 50S subunit differs from that of spinach in a region proposed to interact with the thylakoid membrane. (a) Sharpened map of Marchantia cLSU, zoned around model and colored according to individual RPs and rRNAs shown as a surface at 1 σ above the mean. (b) Comparative density maps (depicted as surfaces) of the LSU's from the bacterium *Bacillus subtilis* (PDB ID 9FY2, zoned around LSU model, 3 σ above the mean), chlororibosome from the alga *C. reinhardtii* (PDB ID 9TVU, zoned around LSU model, 6 σ above the mean), the angiosperm *S. oleracea* (PDB ID 5MLC, 2 σ above the mean), and *M. polymorpha* chlororibosome (PDB ID 31DT, 1 σ above the mean). Images on the right show atomic models in cartoon representation focused on proteins L13 (green), L21 (blue), and L22 (magenta) with density maps depicted as transparent surfaces at thresholds of 3, 4, 2, and 1 σ above the mean (respective to order depicted, top to bottom) and surrounding RPs and rRNA colored gray. Scale bars 20 Å.

RuBisCO is well known for its high level of structural conservation⁶³ and the structure of the *M. polymorpha* enzyme makes no exception. It aligns well with structures of different species spanning more than a 1.5 billion years of diversification that we compared it to (PDB ID 9FWV, 1IWA, 1GK8, 9N37, 4RUB, 9CQ5, 4HHH^{64–71}). This is evident in the visual comparison (Fig. 5), and the z-scores calculated in Dali⁴⁰ where a z-score above 8 indicates likely homology and a score above 20 indicates definite homology. The z-score is influenced by matching secondary structural elements so, the *M. polymorpha* large and small

subunit chains were separately compared in Dali. For the chains of the large subunit, which include the enzyme's active site, the z-scores ranged from 53.4 for *T. antarctica* to 57.5 for *N. tabacum* with a maximum possible score of 66.2 (Fig. S7). For the small subunit, which shows some variance in a single loop region (D176-T198), the z-scores ranged from 15.1 for *G. partita* to 19.9 for *S. oleracea* in comparison to the maximum possible score of 24.5 (Fig. S7).

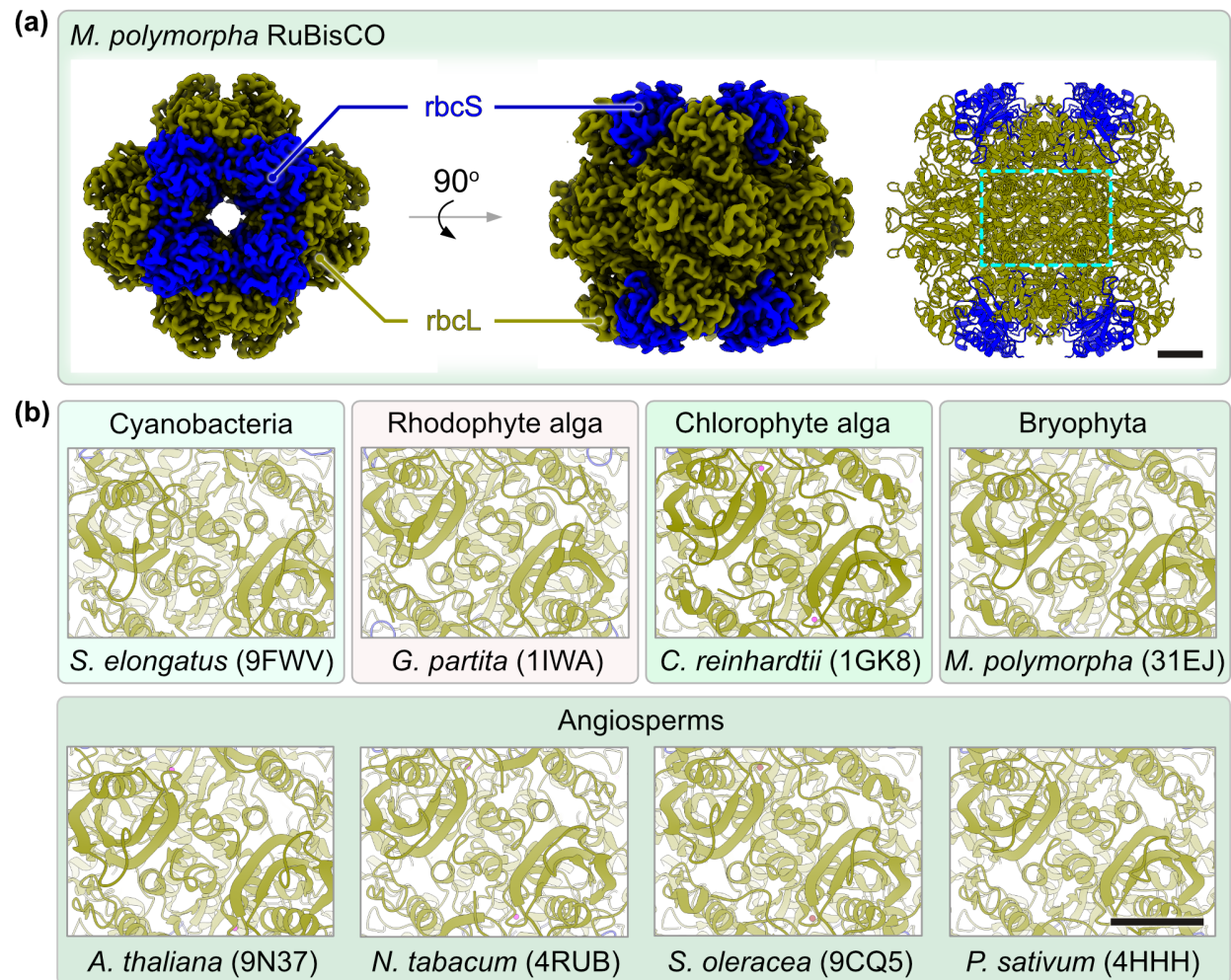


Figure 5: The RuBisCO structure of *M. polymorpha* underscores the enzyme's high level of conservation since the origin of the hetero-oligomeric form I in cyanobacteria. (a) 2.12 Å density map and atomic model of *M. polymorpha* RuBisCO with the small subunits shown in blue and the large subunits shown in green, depicted as a surface, colored by the model at a threshold of 3 σ above the mean; cartoon model on the right with a region containing 2 active sites in adjacent large chains highlighted by a dashed rectangle. (b) An assembly of the same region as highlighted in (a) of eight resolved RuBisCO structures from diverse photosynthetic eukaryotes and a cyanobacterium to highlight the level of conservation spanning >1.8 billion years of evolution. PDB IDs are provided next to the species name. Ions are present in the active sites in 4 of the structures (Mg^{2+} colored pink and Mn^{2+} colored red). Scale bars 20 Å.

Discussion

Here, we have developed a rapid and efficient plastid purification protocol that is effective across six tested plant lineages. By combining our plastid isolation approach with a RAPPL protocol to enrich chloroplast protein complexes, we provide high-resolution structures of the large subunit of the plastid ribosome and RuBisCO from *M. polymorpha*. Together with the plastid proteome, our high-resolution structures of two key protein complexes present a unique window into land plant evolution, as bryophytes are characterized by an ancient genome reduction event and a subsequent gene family expansion independent from that of vascular plants^{19,26,72}.

The oldest fossil record of a plastid is that of the filamentous red alga *Rafatazmia chitrakootensis*. Their rhodoplasts were captured in 1.6-billion-year-old rock formations of the Tirohan Dolomites from central India³⁸. At that point in time, these red algae (rhodophyte) plastids had already evolutionarily diverged from those of their common ancestor of Archaeplastida, from which the Glaucophyta and Chloroplastida (the green lineage) also evolved⁷³ (Fig. 1). Ever since, the photosynthetic organelles of algae and plants have experienced constant adaptations, with the transition to a terrestrial habitat unequivocally representing a groundbreaking step for the evolution of life.

Bryophytes occupy an evaluative position in the evolution of land plants. They are the sister-clade to all vascular land plants^{72,74}, have retained ancestral features such as flagellated sperms or a dominant, photosynthetic haploid gametophyte^{75–77}, and their genomes have independently diverged from those of vascular plants after having been reduced in their common ancestor^{19,26}. We find the *Marchantia* chloroplast proteome to comprise just over 1300 proteins, of which more than 80% are represented by orthologs in angiosperms. The large number of orthologs speaks for a trade-off between ancient nuclear genome reduction and the need to retain – or even expand – gene families needed for the complexity of plastid biology that had to adapt to a terrestrial habitat. We predict that additional organelle proteomes from bryophytes will support this pattern.

Annotating the *M. polymorpha* plastid proteome finds almost 200 proteins of unknown function, with only sporadic domains being identifiable in some cases. TPR and PPR proteins were particularly enriched. TPR domains are key for protein-protein interactions and are involved in chloroplast chlorophyll biogenesis, photosystem assembly and intracellular chloroplast distribution^{78–81}. PPR domains are crucial for plastid RNA editing and organelle-targeted members of this protein family expanded with the origin and diversification of land plants, probably to compensate for an increase in mutations caused by higher levels of UV radiation^{20,82–85}. Theory has it that complex thalloid liverworts such as *M. polymorpha* lack mRNA-editing possibly due to secondary loss^{86,87}, but recent results on *Cyathodium cavernum* revealed a minimum of 129 canonical C-to-U mRNA editing sites, suggesting that this is not a universal trait of the taxon⁸⁸. Studies thus far describe on average 100 proteins with PPR (or related RNA processing) domains encoded by liverwort genomes^{89,90}, though their intracellular localization was not clear. We find 35 proteins (including 22 with unclear functions) with such domains among the chloroplast proteome of *M. polymorpha* – a species that evidently does not edit its plastid-encoded genes in the same way that *C. cavernum* does. This serves as a reminder that studying diversity often challenges generalizations and it raises the question why dozens of PPR domain-containing proteins are then imported. Considering the versatile functions of these proteins ranging from RNA editing, translation activation, splicing, to even acting as subunits of the

ribosomal complex^{91–97}, a detailed investigation of each of these proteins will likely reveal novel aspects of chloroplast biology in the liverwort.

In the World Wide Protein Data Bank, 49 protein structures are deposited for the entire bryophyte clade (as of May 2026), juxtaposed to 2684 structures for *A. thaliana* alone. Similarly, while the plastid ribosome synthesizes key components of photosynthesis, chlororibosome structures are available only for *S. oleracea*⁹⁸ and *C. reinhardtii*⁹⁹. This constrains modern evolutionary studies that, in addition to primary sequences work with protein structures^{100,101}, are used for the design and optimization of semi-synthetic enzymes^{102–104}. Our study deposits structures for 30 bryophyte proteins and 3 rRNAs that constitute two key super complexes of cyanobacterial origin, the large subunit of the chlororibosome and the hetero-oligomeric form I of RuBisCO with its eight small and large subunits. The RAPPL proteome does contain proteins of the small subunit of the chlororibosome (Table S4) and, in combination with the absence of additional assembly factors on the cLSU structure, it suggests the 70S chlororibosome disintegrated during preparation.

Several ribosomal proteins such as L2 to L6, L14, and L16 are among the very few reliable phylogenetic markers across the diversity of life¹⁰⁵. The main reasons are two-fold: their conserved core structure over 4 billion years of evolution and an almost strict vertical inheritance even in prokaryotes^{63,106–108}. The bryophyte's chlororibosome shares this overall level of structural conservation, but there are noticeable differences concerning the absence of terminal loops in L13, L22 and L23 when compared to the spinach chlororibosome (Fig. 4). The extended loops in L22 and L23 appear eudicot-specific, however, and are hypothesized to form an interface with the thylakoid membrane⁵⁹. Their functional relevance in eudicots remains to be demonstrated, along with proteins that might functionally replace these regions in the liverwort. The absence of the L13 N-terminal extension in the density map – while being present in the annotated gene – implies either a high degree of flexibility, disorder or potentially a post-translational cleavage event. Whatever the reason, this region of the bryophyte's chlororibosome seems to better resemble its ancestral roots than that of its eudicot relatives on land.

RuBisCO takes a special place in evolution. It is likely the most abundant protein structure on Earth and the vast majority of biomass owes its origin to it, yet it remains one of the least efficient enzymes we know^{109–111}. It also appears resistant to change. Its macromolecular structure has not been noticeably modified across the vast diversity of photosynthetic eukaryotes since its origin as a hetero-oligomeric complex in cyanobacteria some 3 billion years ago¹⁰⁶ (Fig. 5). The evolution of RuBisCO witnessed the rise of O₂, a molecule that lowers its efficiency by several orders of magnitude. Despite a strong selection pressure for its active center to exclude oxygen, RuBisCO remained unaffected and, instead, distinct carbon capture mechanisms evolved¹¹⁰. When introduced into *E. coli*, not a single nucleotide mutation was observed after 150 chemostat generations for the two genes encoding RuBisCO, despite the prokaryote's biochemistry mutationally adapting to use CO₂ as its sole source for synthesizing hexose, pentose and triose sugars instead¹¹². Thus, RuBisCO presents us with an extreme case of an enzyme that is part of a functional module, in which proteins with many interaction partners or those at the center of critical metabolic fluxes, change in sequence significantly slower than their partners^{113,114}. The enzyme's resilience to environmental changes is telling and might explain why its biotechnological optimization remains kind of a Sisyphean task.

The combination of our plastid isolation and the RAPPL protocol, together with cryogenic EM offers a powerful toolset for exploring the structural proteome of large, soluble protein complexes of the plastid stroma. The application of this workflow in a single species illuminates how plastid proteomes have evolved

following ancestral genome reduction in the liverwort and structure determination of the ribosome hints at modifications in ribosomal proteins that are eudicot-specific and likely aid in the co-translational insertion of plastid-encoded thylakoid proteins. We conclude that our data and development of a widely applicable and efficient plastid purification protocol can accelerate and expand our understanding of plastid biology across diverse plant lineages.

Material and methods

Plant growth conditions

We used ecotype Takaragaire-1 (Tak-1) of the liverwort *Marchantia polymorpha* throughout the study. The wild type gemmae were grown under continuous light (70 $\mu\text{mol m}^{-2} \text{s}^{-1}$) on half-strength Gamborg B5 vitamin agar-plates (GVA).

Plastid isolation

The plant tissues were lysed using mortar and pestle, in the presence of 5-10 mL isolation buffer (IB; 50 mM HEPES KOH pH 7.5, 0.33 M sorbitol, 1 mM MgCl_2 , 1 mM MnCl_2 , 2 mM EDTA pH 8, 0.5% w/v BSA, 0.5% w/v PVP40, and freshly added 5 mM ascorbate). The lysate was filtered through a pre-soaked (in IB) Mira cloth and centrifuged at 1000 g, 10 min. The pellet was resuspended in 200 μL IB and loaded onto a gradient of 70% Percoll[®] (400 μL at the bottom of the tube) and 30% Percoll[®] (1 mL overlayed). Density gradient centrifugation was conducted at 4300 g, 30 min (in a swing bucket rotor). Broken plastids were gently removed first from the top of the 30% Percoll[®] layer. Intact plastids were harvested from the 70-30 layer and washed with 10 mL wash buffer (WB; IB without BSA and PVP40) by centrifugation at 3000 g, 10 min. The plastid pellet was washed again with 10 mL WB by centrifugation at 2000 g, 10 min. The pellet was washed one final time with 1 mL WB by centrifugation at 1000 g, 10 min. From the plastid pellets, crude lysate was formed by pulverizing them in presence of liquid nitrogen. Protein isolation buffer (Invitrogen[™], Catalogue number FNN0011) was added to the powder and was incubated at 50 °C for 15 min for resuspension. The debris were pelleted at 10,000 g, 10 min and the supernatant was used for western blotting.

Mass spectrometry analysis

The crude plastid extract (obtained as per above) was separated by SDS-PAGE and isolated gel pieces were reduced with 50 μL of 10 mM DTT, alkylated with 50 μL of 50 mM iodoacetamide, and digested with 6 μL (200 ng) trypsin in 100 mM ammonium bicarbonate. The peptides were resolved in 15 μL 0.1% trifluoroacetic acid and used for liquid chromatography. LC-MS data was analyzed using QExactive Plus (Thermo Scientific) connected to an Ultimate 3000 Rapid Separation liquid chromatography system (Thermo Scientific) and equipped with an Aurora Ultimate column (75 μm inner diameter, 25 cm length, 1.7 μm particle size, IonOpticks) as the analytical separation column and an Acclaim PepMap 100 C18 column (75 μm inner diameter, 2 cm length, 3 μm particle size, Thermo Fisher Scientific) as the trap column and with a gradient length of 180 min. The mass spectrometer, operated in positive mode, was coupled to a nano-electrospray ionization source at capillary temperature of 250 °C, source voltage of 1.7 kV. Survey scans were acquired over a mass range of 350–2000 m/z and at a resolution of 140,000. The 10 most intense

peptide ions were isolated and fragmented by high-energy collision dissociation (HCD). The peptides were matched against uniprot entries for *Marchantia polymorpha* and proteins represented by at least three peptides (of which two unique) and q-value below 0.05 in at least two biological replicates were included in the plastid proteome.

The whole enriched RAPPL sample and 25% of the flow through were precipitated onto magnetic beads (MagReSynAmine, Resyn Biosciences) with 70% acetonitrile. The proteins were washed on the beads with 100% acetonitrile, 70% ethanol and then resuspended in 100 μ L of 50 mM NH_4HCO_3 . The proteins were reduced by the addition of 0.5 M DTT to a final concentration of 10 mM and incubation at 37 °C for 45 min. Then, to alkylate proteins, 2.7 μ L of 550 mM IAA was added to a final concentration of 15 mM and samples were incubated at RT in the dark for 30 min. 0.5 μ g trypsin was added to each sample for overnight on-bead protein digestion at 37 °C. The resulting peptides were cleaned for mass spectrometry by the STAGE-TIP method using a C18 resin disk (Affinisep).

Samples were then analysed by a nanoElute UHPLC coupled to a timsTOF fleX mass spectrometer (Bruker Daltonics, Bremen, Germany) via a CaptiveSpray ion source. Peptides were separated on a 25 cm reversed-phase C18 column (1.6 μ m bead size, 120 Å pore size, 75 μ m inner diameter, Ion Optics) with a flow rate of 0.3 μ L/min and a solvent gradient from 0-35% B in 60 min. Solvent B was 100% acetonitrile in 0.1% formic acid and solvent A 0.1% formic acid in water. The timsTOF fleX mass spectrometer was operated in DIA-PASEF mode. Mass spectra for MS were recorded between m/z 100 and 1700. Ion mobility resolution was set to 0.85–1.30 V·s/cm over a ramp time of 100 ms. The MS/MS mass range was limited to m/z 475-1000 and ion mobility resolution to 0.85-1.30 V·s/cm to exclude singly charged ions. The estimated cycle time was 0.95 s with 8 cycles using DIA windows of 25 Da. Collisional energy was ramped from 20 eV at 0.60 V·s/cm to 59 eV at 1.60 V·s/cm.

Raw data files from LC-MS/MS analyses were submitted to DIA-NN (version 2.0.2) for protein identification and label-free quantification using *in silico* generated libraries from *Marchantia polymorpha* (trEMBLand Swiss-Prot 2026) and a common laboratory contaminants database from MaxQuant. Carbamidomethyl (C) was set as a fixed modification. Trypsin without proline restriction cleavage option was used, with one allowed miscleavage and peptide length range was set to 7-30 amino acids. MS1 and MS2 mass accuracy was set to 15 ppm and the precursor false discovery rate (FDR) allowed was 0.01 (1%). Match between runs was activated and also RT-dependent cross-run normalization.

Subcellular localization

Marchantia proteins were localized as previously described²⁸. Briefly, full length sequences were amplified from cDNA and fused with citrine on pMpGWB106¹¹⁵, through the *Gateway* (Invitrogen) entry vector pDONR221TM, by BP and LR reactions set as per the manufacturer's protocol. The final vectors were transfected via *Agrobacterium*¹¹⁶ and isogenic plant lines were selected on GVA by Hygromycin (10 μ g/mL) and Cefotaxime (100 μ g/mL). Localization experiments were conducted in protoplasts isolated from 4–7-day old thalli of isogenic transfectant lines. For protoplast isolation, the thalli were incubated with 10 mL of 8% mannitol, 6.3 g/L Gamborg B5 Vitamins (MGV) with 15 mg/mL cellulase and 5 mg/mL Macroenzyme R10. Protoplasts were dislodged and filtered with a 50 μ m strainer. After washing twice with MGV (at 300 g, 3min) and gently resuspending, the protoplasts were used for mitochondrial staining by MitoTrackerTM Red CMXRos according to the manufacturer's instructions. Protoplasts were imaged using Nikon eclipse Ti imaging platform and collages were prepared in ImageJ (Fiji).

420

421 *Comparative organelle proteomics*

422 We clustered 56 Chloroplastida genomes and sorted them into organelle orthogroups as previously
 423 described²⁰. Briefly, we downloaded whole genome models for diverse Chloroplastida species, clustered
 424 them into orthologous protein families (or orthogroups) using OrthoFinder version 2.5.4¹¹⁷ and annotated a
 425 subset of these orthogroups as plastid orthogroups (POGs) based on the presence of experimentally
 426 validated plastid proteins from species where available: *A. thaliana*³⁶, maize³⁵, rice³⁵,
 427 *Physcomitrium patens*³⁴ and *C. reinhardtii*^{32,33}, as well as *M. polymorpha* (current study). Through this
 428 approach, more than 99% of ca. 15k experimentally validated plastid proteins were efficiently clustered
 429 into ca. 5k POGs (Table S3). The high efficiency is likely explained by the added experimental proteomes
 430 and removed non-chloroplastidal species, where homologues are more diverged.

431 *Ribosome isolation from Marchantia chloroplasts*

432 Isolated plastids were lysed at 4 °C for 30 min with rotation in lysis buffer (2% (v/v) Triton X-100, 50 mM
 433 Tris-HCl pH 7.5, 50 mM KCl, 2 mM DTT, 10 mM Mg(OAc)₂, 2 mM spermidine, 40 U/μL RNase inhibitor
 434 murine, 0.1 mM phenylmethylsulfonyl fluoride). Lysate was clarified by centrifugation at 20,800 g at 4 °C
 435 for 45 min. Purification of the clarified lysate was conducted following the RAPPL protocol. 50 μL poly-
 436 L-lysine bead slurry (Molecular Cloning Laboratories) was added to lysate supernatant and rotated at 4 °C
 437 for 30 min. After placement in a magnetic tube rack, the flowthrough was removed and 250 μL wash buffer
 438 (100 mM HEPES pH 7.4, 50 mM KCl, 10 mM Mg(OAc)₂, 1 mM DTT, 40 U/μL RNase inhibitor murine,
 439 2 mM spermidine) was added. After mixing by pipetting and vortexing, the tube was placed into the
 440 magnetic rack and the solution of unbound proteins was removed. The wash was repeated twice (3 in total).
 441 20 μL elution buffer (wash buffer containing 10 mg/mL poly-D-glutamic acid) was added to the beads and
 442 then incubated at 4 °C with rotation for 30 min.

443 *Cryogenic EM and TEM imaging*

444 Samples were vitrified on a Leica GP using ANTCryo R1.2/1.3 Au 300 grids. The grids were imaged with
 445 a 200 kV Talos Arctica TEM equipped with a Falcon 4i direct electron detector (FINStruct cryo-EM unit,
 446 University of Helsinki). Data collection with Thermo Fisher EPU was conducted at 150,000x nominal
 447 magnification with a pixel size of 0.92 Å and total electron dose of 39.69 e/Å². In total, 5,652 movies were
 448 collected in the .eer format (see Table S1).

449 *Single-particle processing*

450 Movies were processed in cryoSPARC v4.7¹¹⁸. The EER up-sampling factor was set to 1 during movie
 451 import. After patch motion correction and patch CTF estimation, pre-processed micrographs were checked
 452 and micrographs with an estimated CTF above 5 Å were rejected, leaving 4,934 accepted for further
 453 processing. 2,964,872 particles were initially selected using a blob picker and, after inspection, 1,280,024
 454 particles were extracted with a box size of 416 pixels. Extracted particles were run in 2D classification
 455 iteratively, using a variable number of 2D classes, 40 online-EM iterations, a batchsize per class of 200,
 456 and minimize over per particle scale after the first iteration. 150 2D classes composed of 379,270 particles
 457 were used for template picking and 1,309,380 inspected template picked particles were extracted with a
 458 box size of 416 pixels. After another round of iterative 2D classification (same parameters as above), 126
 459 classes composed of 364,829 particles were selected. 3D volumes were created using ab-initio
 460 reconstruction, with 3 classes, a 260 initial minibatch size, and 900 final minibatch size. All 3 classes were

used in a single heterogeneous refinement with a batch size of 2,000 per class and a box size of 416 voxels. 264,465 particles comprised the largest refined class (2.89 Å resolution) which was used for subsequent non-uniform refinement with optimize per-particle defocus, optimize per-group CTF params, and positive EWS correction¹¹⁹. The resulting 2.37 Å map was used as the input volume for reference-based motion correction¹²⁰ where movies with wrong frame count were excluded. The resulting 262,814 particles, were used for another non-uniform refinement with the same parameters as above, resulting in a 2.23 Å map (Fig. S4).

Non-ribosomal particles were selected from the blob picked 2D classes for separate analysis using an identical pipeline and similar individual job parameters (Fig. S4). 1,485,133 particles were extracted with a 256 pixel box size. 13 2D classes composed of 471,791 particles were used for ab-initio reconstruction and subsequent heterogeneous refinement using 3 ab-initio volumes. The largest class, composed of 299,945 particles (3.20 Å), was identified as a map of RuBisCO and was used as an input for non-uniform refinement resulting in a 2.52 Å map in C1 symmetry. When D4 symmetry was applied in a separate non-uniform refinement, the result was a 2.21 Å map. After reference-based motion correction and a final non-uniform refinement with D4 symmetry applied, 299,471 particles composed the 2.12 Å map.

Local resolution estimation was conducted for both maps (2.23 Å cLSU and 2.12 Å RuBisCO), with the results analyzed in UCSF ChimeraX v1.9 (Fig. S4,¹²¹). Both maps were sharpened in DeepEMhancer¹²² (within cryoSPARC) with the 3 available models used for both maps and manually compared in UCSF ChimeraX, with the reported maps sharpened using the wideTarget model for RuBisCO and highRes model for the LSU maps respectively.

Modeling and validation

Protein sequences were obtained through BLASTP starting from 5MLC *S. oleracea* 50S chlororibosome structure and 7JN4 *C. reinhardtii* RuBisCO looking for *M. polymorpha* sequences¹²³. The following GenBank accession numbers were used for 50S sequences: BBN04705.1, PTQ33360.1, OAE33939.1, P12142.1, OAE31853.1, OAE19889.1, P06392.3, P12196.3, OAE31933.1, OAE33704.1, PTQ42878.1, PTQ48266.1, BDD77304.1, P06388.1, P06387.1, P06385.3, PTQ47902.1, PTQ36989.1, PTQ46014.1, YP_009479654.1, OAE34979.1, YP_009479653.1, OAE23259.1, PTQ48626.1, OAE20865.1, APT68014.1, PTQ48664.1, P06378.1, and RuBisCO sequences: ALV87782.1, PTQ31246.1^{53,54,56–58}. BLASTP sequences were run through AlphaFold3 on the publicly-accessible web server and then fit to the previously mentioned PDB structures using the Matchmaker tool in UCSF ChimeraX. They were subsequently fit to their respective density maps using the Fit in Map tool in UCSF ChimeraX. Models were fit residue by residue using ISOLDE in ChimeraX and regions of the sequences with no corresponding density were deleted from the models (Table S6)¹²⁴.

Nucleotide sequences were obtained through BLASTN starting from the PDB ID 5MLC spinach chlororibosome structure with the following GenBank accession numbers: NC_042505.1, KT356925.1, KT356925.1^{55,59}. The 5MLC structure was used as a starting point for rRNA structures. Individual point mutations and deletions were made in UCSF ChimeraX using swapnt and delete commands. Insertions were made with the Build Structure tool in UCSF ChimeraX and Coot v0.9.8.96, which was used to re-order and change chain IDs¹²⁵. Models were fit residue by residue using ISOLDE and regions of the sequences with no corresponding density were deleted from the models (Table S6).

Ions were fit to the cLSU structure using the isolde add ligand command where a strong signal with appropriate coordination was present in the sharpened density map. Specific ion assignment was based on

coordination distances and interacting atoms. Coordination was confirmed with water molecules using the un-sharpened map and the isolate add water command. Spermidine was added based on unmodeled density in the sharpened map with the Build Structure tool. rRNA modifications were added based on unmodeled density in the sharpened map with the Build Structure tool. All ligands and modifications were fit with refinement in ISOLDE. Validation was conducted during modeling in ISOLDE and by the wwPDB OneDep validation servers.

Data availability

Structures were deposited to the Worldwide Protein Data Bank through the wwPDB OneDep system. The cLSU structure was given the following accession codes: PDB ID 31DT, EMD-58329. The RuBisCO structure was given the following accession codes: PDB ID 31EJ, EMD-58338. Micrographs are deposited through EMPIAR (deposition ID: 47488363, accession code: EMPIAR-13780) and linked to both Electron Microscopy Data Bank accessions. Supplementary figures are available with this submission. Additional data are available on Zenodo (10.5281/zenodo.21338199), which includes, but is not limited to, figure source data, a supplementary movie for *M. polymorpha* chloroplast ribosome LSU, supplementary tables and in-house scripts.

Author contributions

PKR: Conceptualization, Experimental design, Methodology, Investigation, Data curation, Formal analysis, Validation and Visualization, Writing - original draft, review and editing. CM: Experimental design, Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing - original draft, review and editing. MLQ: Investigation; Data curation and Validation. SOK: Experimental design, Methodology; Funding and Resource acquisition; Writing - review and editing; TAN: Formal analysis; BJB: Supervision; Experimental design; Funding and Resource acquisition; Writing - review and editing; SJB: Supervision; Experimental design; Funding and Resource acquisition; Writing - review and editing. SBG: Conceptualization; Project administration; Supervision; Experimental design; Funding and Resource acquisition; Data visualization; Writing - review and editing.

Funding

SBG acknowledges funding from the DFG (581702633) and the HHU through its strategic research fund. SJB is grateful for support from the University of Helsinki Boost funding and the Sigrid Jusélius Foundation (260021). BJB received funding from the Sigrid Jusélius Foundation Senior Investigator Award, Research Council of Finland (357469), the Magnus Ehrnrooth Foundation, and the University of Helsinki Boost funding. SOK acknowledges the support and the use of resources of Instruct-ERIC through the R&D pilot scheme APPID3490. CM is a fellow of the MACS graduate programme, University of Helsinki.

Acknowledgments

We acknowledge support from the high-performance computing cluster (HILBERT) at the Heinrich Heine University Düsseldorf and thank Daniel Wasim Djamriani, Carolina Garcia Garcia and Margarete Stracke for their help. We also thank Dr. Anja Stefanski and Molecular Proteomics Laboratory at HHU-Düsseldorf for conducting the mass-spectrometry. PKR and MLQ are grateful to Prof. William F. Martin for providing financial support. Mass spectrometry-based proteomic analyses of RAPPL sample preparations were performed by the Proteomics Core Facility, Department of Immunology, University of Oslo/Oslo University Hospital, which is supported by the Core Facilities program of the South-Eastern Norway Regional Health Authority. This core facility is also a member of the National Network of Advanced Proteomics Infrastructure (NAPI), which is funded by the Research Council of Norway INFRASTRUKTUR-program (project number: 295910). We thank Kiran L.L. Ahmad and Tuomas Niemi-Aro (University of Helsinki) for technical assistance in cryo-EM. The facilities and expertise of the HiLIFE cryo-EM unit at the University of Helsinki, a member of Instruct-ERIC Centre Finland, FINStruct, and Biocenter Finland are gratefully acknowledged. We acknowledge computational resources from CSC – IT Center for Science, Finland. Molecular graphics and analyses performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.

561 References

562

- 563 1. Timmis, J.N., Ayliff, M.A., Huang, C.Y., and Martin, W. (2004). Endosymbiotic gene transfer:
564 Organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* 5, 123–135.
565 <https://doi.org/10.1038/nrg1271>.
- 566 2. Kelly, S. (2021). The economics of organellar gene loss and endosymbiotic gene transfer. *Genome*
567 *Biol.* 22. <https://doi.org/10.1186/s13059-021-02567-w>.
- 568 3. Martin, W., and Herrmann, R.G. (1998). Gene Transfer from Organelles to the Nucleus: How
569 Much, What Happens, and Why? *Plant Physiol.* 118, 9–17. <https://doi.org/10.1104/pp.118.1.9>.
- 570 4. Kleine, T., Maier, U.G., and Leister, D. (2009). DNA transfer from organelles to the nucleus: The
571 idiosyncratic genetics of endosymbiosis. *Annu. Rev. Plant Biol.* 60, 115–138.
572 <https://doi.org/10.1146/annurev.arplant.043008.092119>.
- 573 5. Green, B.R. (2011). Chloroplast genomes of photosynthetic eukaryotes. *Plant Journal* 66, 34–44.
574 <https://doi.org/10.1111/j.1365-313X.2011.04541.x>.
- 575 6. Sun, Y., and Jarvis, R.P. (2023). Chloroplast Proteostasis: Import, Sorting, Ubiquitination, and
576 Proteolysis. *Annual Review of Plant Biology* 12, 19. <https://doi.org/10.1146/annurev-arplant-070122>.
- 577 7. Gao, L.L., Hong, Z.H., Wang, Y., and Wu, G.Z. (2023). Chloroplast proteostasis: A story of birth,
578 life, and death. Preprint at Cell Press, <https://doi.org/10.1016/j.xplc.2022.100424>
579 <https://doi.org/10.1016/j.xplc.2022.100424>.
- 580 8. Allen, J.F. (2015). Why chloroplasts and mitochondria retain their own genomes and genetic
581 systems: Colocation for redox regulation of gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 112, 10231–
582 10238. <https://doi.org/10.1073/pnas.1500012112>.
- 583 9. Race, H.L., Herrmann, R.G., and Martin, W. (1999). Why have organelles retained genomes ?
584 *Trends Genet.* 15, 364–370. [https://doi.org/10.1016/s0168-9525\(99\)01766-7](https://doi.org/10.1016/s0168-9525(99)01766-7).
- 585 10. Allen, J.F., and Martin, W.F. (2016). Why have organelles retained genomes? *Cell Syst.* 2, 70–72.
586 <https://doi.org/10.1016/j.cels.2016.02.007>.
- 587 11. Giannakis, K., Arrowsmith, S.J., Richards, L., Gasparini, S., Chustecki, J.M., Røyrvik, E.C., and
588 Johnston, I.G. (2022). Evolutionary inference across eukaryotes identifies universal features shaping
589 organelle gene retention. *Cell Syst.* 13, 874–884.e5. <https://doi.org/10.1016/j.cels.2022.08.007>.
- 590 12. Björkholm, P., Ernst, A.M., Hagström, E., and Andersson, S.G.E. (2017). Why mitochondria need
591 a genome revisited. *FEBS Lett.* 591, 65–75. <https://doi.org/10.1002/1873-3468.12510>.
- 592 13. Björkholm, P., Harish, A., Hagström, E., Ernst, A.M., and Andersson, S.G.E. (2015). Mitochondrial
593 genomes are retained by selective constraints on protein targeting. *Proc. Natl. Acad. Sci. U. S. A.* 112, 10154–
594 10161. <https://doi.org/10.1073/pnas.1421372112>.
- 595 14. von Heijne, G. (1986). Why mitochondria need a genome. *FEBS Lett.* 198, 1–4.
596 [https://doi.org/10.1016/0014-5793\(86\)81172-3](https://doi.org/10.1016/0014-5793(86)81172-3).

- 597 15. Maier, U.G., Zauner, S., Woehle, C., Bolte, K., Hempel, F., Allen, J.F., and Martin, W.F. (2013).
598 Massively convergent evolution for ribosomal protein gene content in plastid and mitochondrial genomes.
599 *Genome Biol. Evol.* 5, 2318–2329. <https://doi.org/10.1093/gbe/evt181>.
- 600 16. Bowman, J.L. (2022). The origin of a land flora. Preprint at Nature Research,
601 <https://doi.org/10.1038/s41477-022-01283-y> <https://doi.org/10.1038/s41477-022-01283-y>.
- 602 17. Schreiber, M., Rensing, S.A., and Gould, S.B. (2022). The greening ashore. Preprint at Elsevier
603 Ltd, <https://doi.org/10.1016/j.tplants.2022.05.005> <https://doi.org/10.1016/j.tplants.2022.05.005>.
- 604 18. de Vries, J., Stanton, A., Archibald, J.M., and Gould, S.B. (2016). Streptophyte Terrestrialization in
605 Light of Plastid Evolution. *Trends Plant Sci.* 21, 467–476. <https://doi.org/10.1016/j.tplants.2016.01.021>.
- 606 19. Harris, B.J., Clark, J.W., Schrempf, D., Szöllösi, G.J., Donoghue, P.C.J., Hetherington, A.M., and
607 Williams, T.A. (2022). Divergent evolutionary trajectories of bryophytes and tracheophytes from a complex
608 common ancestor of land plants. *Nat. Ecol. Evol.* 6, 1634–1643. [https://doi.org/10.1038/s41559-022-01885-](https://doi.org/10.1038/s41559-022-01885-x)
609 x.
- 610 20. Raval, P.K., MacLeod, A.I., and Gould, S.B. (2024). A molecular atlas of plastid and mitochondrial
611 proteins reveals organellar remodeling during plant evolutionary transitions from algae to angiosperms. *PLoS*
612 *Biol.* 22. <https://doi.org/10.1371/journal.pbio.3002608>.
- 613 21. Feng, X., Zheng, J., Irisarri, I., Yu, H., Zheng, B., Ali, Z., de Vries, S., Keller, J., Fürst-Jansen,
614 J.M.R., Dadras, A., et al. (2024). Genomes of multicellular algal sisters to land plants illuminate signaling
615 network evolution. *Nat. Genet.* 56, 1018–1031. <https://doi.org/10.1038/s41588-024-01737-3>.
- 616 22. Donoghue, P.C.J., Harrison, C.J., Paps, J., and Schneider, H. (2021). The evolutionary emergence
617 of land plants. Preprint at Cell Press, <https://doi.org/10.1016/j.cub.2021.07.038>
618 <https://doi.org/10.1016/j.cub.2021.07.038>.
- 619 23. Dadras, A., Fürst-Jansen, J.M.R., Darienko, T., Krone, D., Scholz, P., Sun, S., Herrfurth, C.,
620 Rieseberg, T.P., Irisarri, I., Steinkamp, R., et al. (2023). Environmental gradients reveal stress hubs pre-dating
621 plant terrestrialization. *Nat. Plants* 9, 1419–1438. <https://doi.org/10.1038/s41477-023-01491-0>.
- 622 24. Leebens-Mack, J.H., Barker, M.S., Carpenter, E.J., Deyholos, M.K., Gitzendanner, M.A., Graham,
623 S.W., Grosse, I., Li, Z., Melkonian, M., Mirarab, S., et al. (2019). One thousand plant transcriptomes and
624 the phylogenomics of green plants. *Nature* 574, 679–685. <https://doi.org/10.1038/s41586-019-1693-2>.
- 625 25. Wang, S., Li, L., Li, H., Sahu, S.K., Wang, H., Xu, Y., Xian, W., Song, B., Liang, H., Cheng, S., et
626 al. (2020). Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nat. Plants* 6,
627 95–106. <https://doi.org/10.1038/s41477-019-0560-3>.
- 628 26. Dong, S., Wang, S., Li, L., Yu, J., Zhang, Y., Xue, J.-Y., Chen, H., Ma, J., Zeng, Y., Cai, Y., et al.
629 (2025). Bryophytes hold a larger gene family space than vascular plants. *Nat. Genet.*
630 <https://doi.org/10.1038/s41588-025-02325-9>.
- 631 27. Erath, J., Kemper, D., Mugo, E., Jacoby, A., Valenzuela, E., Jungers, C.F., Beatty, W.L., Hashem,
632 Y., Jovanovic, M., Djuranovic, S., et al. (2025). A rapid, simple, and economical method for the isolation of

633 ribosomes and translational machinery for structural and functional studies. *Nat. Commun.* *16*, 7185.
634 <https://doi.org/10.1038/s41467-025-62314-8>.

635 28. Raval, P.K., García, C.G., Sanchez, M.-D.S., and Gould, S.B. (2025). Evolutionary refinement of
636 mitochondrial and plastid targeting sequences coincides with the late diversification of land plants. *Mol. Biol.*
637 *Evol.* <https://doi.org/10.1093/molbev/msaf240>.

638 29. de Vries, J., and Gould, S.B. (2018). The monoplastidic bottleneck in algae and plant evolution. *J.*
639 *Cell Sci.* *131*, 1–14. <https://doi.org/10.1242/jcs.203414>.

640 30. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., and Ishiguro-Watanabe, M. (2023). KEGG
641 for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* *51*, D587–D592.
642 <https://doi.org/10.1093/nar/gkac963>.

643 31. Facchinelli, F., Pribil, M., Oster, U., Ebert, N.J., Bhattacharya, D., Leister, D., and Weber, A.P.M.
644 (2013). Proteomic analysis of the *Cyanophora paradoxa* muroplast provides clues on early events in plastid
645 endosymbiosis. *Planta* *237*, 637–651. <https://doi.org/10.1007/s00425-012-1819-3>.

646 32. Atteia, A., Adrait, A., Brugiere, S., Tardif, M., Van Lis, R., Deusch, O., Dagan, T., Kuhn, L., Gontero,
647 B., Martin, W., et al. (2009). A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light
648 on the metabolic plasticity of the organelle and on the nature of the α -proteobacterial mitochondrial ancestor.
649 *Mol. Biol. Evol.* *26*, 1533–1548. <https://doi.org/10.1093/molbev/msp068>.

650 33. Terashima, M., Specht, M., and Hippler, M. (2011). The chloroplast proteome: A survey from the
651 *Chlamydomonas reinhardtii* perspective with a focus on distinctive features. Preprint,
652 <https://doi.org/10.1007/s00294-011-0339-1> <https://doi.org/10.1007/s00294-011-0339-1>.

653 34. Mueller, S.J., Lang, D., Hoernstein, S.N.W., Lang, E.G.E., Schuessele, C., Schmidt, A., Fluck, M.,
654 Leisibach, D., Niegl, C., Zimmer, A.D., et al. (2014). Quantitative analysis of the mitochondrial and plastid
655 proteomes of the moss *Physcomitrella patens* reveals protein macrocompartmentation and
656 microcompartmentation. *Plant Physiol.* *164*, 2081–2095. <https://doi.org/10.1104/pp.114.235754>.

657 35. Hooper, C.M., Castleden, I.R., Aryamanesh, N., Black, K., Grasso, S. V., and Millar, A.H. (2020).
658 CropPAL for discovering divergence in protein subcellular location in crops to support strategies for
659 molecular crop breeding. *Plant Journal* *104*, 812–827. <https://doi.org/10.1111/tpj.14961>.

660 36. Hooper, C.M., Castleden, I.R., Tanz, S.K., Aryamanesh, N., and Millar, A.H. (2017). SUBA4: The
661 interactive data analysis centre for Arabidopsis subcellular protein locations. *Nucleic Acids Res.* *45*, D1064–
662 D1074. <https://doi.org/10.1093/nar/gkw1041>.

663 37. Mason, C.B., Bricker, T.M., and Moroney, J. V. (2006). A rapid method for chloroplast isolation
664 from the green alga *Chlamydomonas reinhardtii*. *Nat. Protoc.* *1*, 2227–2230.
665 <https://doi.org/10.1038/nprot.2006.348>.

666 38. Bengtson, S., Sallstedt, T., Belivanova, V., and Whitehouse, M. (2017). Three-dimensional
667 preservation of cellular and subcellular structures suggests 1.6 billion-year-old crown-group red algae. *PLoS*
668 *Biol.* *15*. <https://doi.org/10.1371/journal.pbio.2000735>.

- 669 39. Tanizawa, Y., Mochizuki, T., Yagura, M., Sakamoto, M., Fujisawa, T., Kawamura, S., Shimokawa,
670 E., Yamaoka, S., Nishihama, R., Bowman, J.L., et al. (2026). MarpolBase: genome database for *Marchantia*
671 *polymorpha* featuring high quality reference genome sequences. *Plant Cell Physiol.* 67, 377–388.
672 <https://doi.org/10.1093/pcp/pcf159>.
- 673 40. Holm, L., Laiho, A., Törönen, P., and Salgado, M. (2023). DALI shines a light on remote homologs:
674 One hundred discoveries. *Protein Science* 32. <https://doi.org/10.1002/pro.4519>.
- 675 41. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore,
676 L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with
677 AlphaFold 3. *Nature* 630, 493–500. <https://doi.org/10.1038/s41586-024-07487-w>.
- 678 42. Shen, L., Tang, K., Wang, W., Wang, C., Wu, H., Mao, Z., An, S., Chang, S., Kuang, T., Shen, J.R.,
679 et al. (2022). Architecture of the chloroplast PSI–NDH supercomplex in *Hordeum vulgare*. *Nature* 601, 649–
680 654. <https://doi.org/10.1038/s41586-021-04277-6>.
- 681 43. Su, X., Cao, D., Pan, X., Shi, L., Liu, Z., Dall’Osto, L., Bassi, R., Zhang, X., and Li, M. (2022).
682 Supramolecular assembly of chloroplast NADH dehydrogenase-like complex with photosystem I from
683 *Arabidopsis thaliana*. *Mol. Plant* 15, 454–467. <https://doi.org/10.1016/j.molp.2022.01.020>.
- 684 44. Peng, L., Fukao, Y., Fujiwara, M., Takami, T., and Shikanai, T. (2009). Efficient operation of
685 NAD(P)H dehydrogenase requires supercomplex formation with photosystem I via minor LHCI in
686 *arabidopsis*. *Plant Cell* 21, 3623–3640. <https://doi.org/10.1105/tpc.109.068791>.
- 687 45. Gould, S.B., Garg, S.G., Handrich, M., Nelson-Sathi, S., Gruenheit, N., Tielens, A.G.M., and
688 Martin, W.F. (2019). Adaptation to life on land at high O₂ via transition from ferredoxin-to NADH-dependent
689 redox balance. *Proceedings of the Royal Society B: Biological Sciences* 286.
690 <https://doi.org/10.1098/rspb.2019.1491>.
- 691 46. Opatíková, M., and Kouřil, R. (2024). Unique structural attributes of the PSI-NDH supercomplex
692 in *Physcomitrium patens*. *Plant Journal* 120, 2226–2237. <https://doi.org/10.1111/tpj.17116>.
- 693 47. Ueda, M., Kuniyoshi, T., Yamamoto, H., Sugimoto, K., Ishizaki, K., Kohchi, T., Nishimura, Y., and
694 Shikanai, T. (2012). Composition and physiological function of the chloroplast NADH dehydrogenase-like
695 complex in *Marchantia polymorpha*. *Plant Journal* 72, 683–693. <https://doi.org/10.1111/j.1365->
696 313X.2012.05115.x.
- 697 48. Iwai, M., and Yokono, M. (2017). Light-harvesting antenna complexes in the moss *Physcomitrella*
698 *patens*: implications for the evolutionary transition from green algae to land plants. Preprint at Elsevier Ltd,
699 <https://doi.org/10.1016/j.pbi.2017.04.002> <https://doi.org/10.1016/j.pbi.2017.04.002>.
- 700 49. Kato, Y., Odahara, M., Fukao, Y., and Shikanai, T. (2018). Stepwise evolution of supercomplex
701 formation with photosystem I is required for stabilization of chloroplast NADH dehydrogenase-like complex:
702 Lhca5-dependent supercomplex formation in *Physcomitrella patens*. *Plant Journal* 96, 937–948.
703 <https://doi.org/10.1111/tpj.14080>.
- 704 50. Gould, S.B., Magiera, J., García García, C., and Raval, P.K. (2024). Reliability of plastid and
705 mitochondrial localisation prediction declines rapidly with the evolutionary distance to the training set
706 increasing. *PLoS Comput. Biol.* 20, e1012575. <https://doi.org/10.1371/journal.pcbi.1012575>.

- 707 51. Rosenthal, P.B., and Henderson, R. (2003). Optimal determination of particle orientation, absolute
708 hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* 333, 721–745.
709 <https://doi.org/10.1016/j.jmb.2003.07.013>.
- 710 52. Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T.,
711 Ogura, Y., Kohchi, T., et al. (1992). Gene organization deduced from the complete sequence of liverwort
712 *Marchantia polymorpha* mitochondrial DNA. *J. Mol. Biol.* 223, 1–7. [https://doi.org/10.1016/0022-
713 2836\(92\)90708-R](https://doi.org/10.1016/0022-713 2836(92)90708-R).
- 714 53. Honkanen, S., Jones, V.A.S., Morieri, G., Champion, C., Hetherington, A.J., Kelly, S., Proust, H.,
715 Saint-Marcoux, D., Prescott, H., and Dolan, L. (2016). The Mechanism Forming the Cell Surface of Tip-
716 Growing Rooting Cells Is Conserved among Land Plants. *Current Biology* 26, 3238–3244.
717 <https://doi.org/10.1016/j.cub.2016.09.062>.
- 718 54. Fukuzawa, H., Kohchi, T., Sano, T., Shirai, H., Umesono, K., Inokuchi, H., Ozeki, H., and Ohyama,
719 K. (1988). Structure and organization of *Marchantia polymorpha* chloroplast genome. *J. Mol. Biol.* 203, 333–
720 351. [https://doi.org/10.1016/0022-2836\(88\)90003-4](https://doi.org/10.1016/0022-2836(88)90003-4).
- 721 55. Forrest, L.L., Long, D.G., Cargill, D.C., Hart, M.L., Milne, J., Schill, D.B., Seppelt, R.D., and
722 Villarreal, J.C. (2015). On *Monocarpus* (Monocarpaceae, Marchantiopsida), an isolated salt-pan complex
723 thalloid liverwort. *Aust. Syst. Bot.* 28, 137–144. <https://doi.org/10.1071/SB15012>.
- 724 56. Montgomery, S.A., Tanizawa, Y., Galik, B., Wang, N., Ito, T., Mochizuki, T., Akimcheva, S.,
725 Bowman, J.L., Cognat, V., Maréchal-Drouard, L., et al. (2020). Chromatin Organization in Early Land Plants
726 Reveals an Ancestral Association between H3K27me3, Transposons, and Constitutive Heterochromatin.
727 *Current Biology* 30, 573–588.e7. <https://doi.org/10.1016/j.cub.2019.12.015>.
- 728 57. Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama,
729 R., Nakamura, Y., Berger, F., et al. (2017). Insights into Land Plant Evolution Garnered from the *Marchantia*
730 *polymorpha* Genome. *Cell* 171, 287–304.e15. <https://doi.org/10.1016/j.cell.2017.09.030>.
- 731 58. Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y.,
732 Takeuchi, M., Chang, Z., et al. (1986). Chloroplast gene organization deduced from complete sequence of
733 liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* 322, 572–574. <https://doi.org/10.1038/322572a0>.
- 734 59. Graf, M., Arenz, S., Huter, P., Dönhöfer, A., Nováček, J., and Wilson, D.N. (2016). Cryo-EM
735 structure of the spinach chloroplast ribosome reveals the location of plastid-specific ribosomal proteins and
736 extensions. *Nucleic Acids Res.*, gkw1272. <https://doi.org/10.1093/nar/gkw1272>.
- 737 60. Arenz, S., Bock, L. V., Graf, M., Innis, C.A., Beckmann, R., Grubmüller, H., Vaiana, A.C., and
738 Wilson, D.N. (2016). A combined cryo-EM and molecular dynamics approach reveals the mechanism of
739 ErmBL-mediated translation arrest. *Nat. Commun.* 7, 12026. <https://doi.org/10.1038/ncomms12026>.
- 740 61. Noeske, J., Wasserman, M.R., Terry, D.S., Altman, R.B., Blanchard, S.C., and Cate, J.H.D. (2015).
741 High-resolution structure of the *Escherichia coli* ribosome. *Nat. Struct. Mol. Biol.* 22, 336–341.
742 <https://doi.org/10.1038/nsmb.2994>.

- 743 62. Waltz, F., Lehner, P.A., Van der Stappen, P., Kater, L., Pfeffer, S., and Engel, B.D. (2026).
744 Chloroplast-encoded small subunit extensions reshape the *Chlamydomonas* chlororibosome. *bioRxiv*.
745 <https://doi.org/10.64898/2026.02.07.704542>.
- 746 63. Prywes, N., Phillips, N.R., Tuck, O.T., Valentin-Alvarado, L.E., and Savage, D.F. (2023). Rubisco
747 Function, Evolution, and Engineering. *Annu. Rev. Biochem.* 92, 385–410. [https://doi.org/10.1146/annurev-](https://doi.org/10.1146/annurev-biochem-040320-101244)
748 [biochem-040320-101244](https://doi.org/10.1146/annurev-biochem-040320-101244).
- 749 64. Loewen, P.C., Didychuk, A.L., Switala, J., Perez-Luque, R., Fita, I., and Loewen, M.C. (2013).
750 Structure of *Pisum sativum* Rubisco with bound ribulose 1,5-bisphosphate. *Acta Crystallogr. Sect. F Struct.*
751 *Biol. Cryst. Commun.* 69, 10–14. <https://doi.org/10.1107/S1744309112047549>.
- 752 65. Voland, R.W., Coleman, R.E., and Lancaster, K.M. (2024). The structure of Mn(II)-bound Rubisco
753 from *Spinacia oleracea*. *J. Inorg. Biochem.* 260, 112682. <https://doi.org/10.1016/j.jinorgbio.2024.112682>.
- 754 66. Suh, S.W., Cascio, D., Chapman, M.S., and Eisenberg, D. (1987). A crystal form of ribulose-1,5-
755 bisphosphate carboxylase/oxygenase from *Nicotiana tabacum* in the activated state. *J. Mol. Biol.* 197, 363–
756 365. [https://doi.org/10.1016/0022-2836\(87\)90129-X](https://doi.org/10.1016/0022-2836(87)90129-X).
- 757 67. Askey, B., Ceminsky, M., Scott, E., Wang, Y., Oh, Z.G., Azinas, S., Laganowsky, A., and Gunn,
758 L.H. (2026). Rubisco kinetic acclimation at the holoenzyme level. *Proceedings of the National Academy of*
759 *Sciences* 123. <https://doi.org/10.1073/pnas.2519914123>.
- 760 68. Taylor, T.C., Backlund, A., Bjorhall, K., Spreitzer, R.J., and Andersson, I. (2001). First Crystal
761 Structure of Rubisco from a Green Alga, *Chlamydomonas reinhardtii*. *Journal of Biological Chemistry* 276,
762 48159–48164. <https://doi.org/10.1074/jbc.M107765200>.
- 763 69. Okano, Y., Mizohata, E., Xie, Y., Matsumura, H., Sugawara, H., Inoue, T., Yokota, A., and Kai, Y.
764 (2002). X-ray structure of *Galdieria* Rubisco complexed with one sulfate ion per active site. *FEBS Lett.* 527,
765 33–36. [https://doi.org/10.1016/S0014-5793\(02\)03148-4](https://doi.org/10.1016/S0014-5793(02)03148-4).
- 766 70. Sun, Y., Sheng, Y., Ni, T., Ge, X., Sarsby, J., Brownridge, P.J., Li, K., Hardenbrook, N., Dykes,
767 G.F., Rockliffe, N., et al. (2024). Rubisco packaging and stoichiometric composition of the native β -
768 carboxysome in *Synechococcus elongatus* PCC7942. *Plant Physiol.* 197.
769 <https://doi.org/10.1093/plphys/kiac665>.
- 770 71. Holm, L. (2020). Using Dali for Protein Structure Comparison. In, pp. 29–42.
771 https://doi.org/10.1007/978-1-0716-0270-6_3.
- 772 72. Harris, B.J., Harrison, C.J., Hetherington, A.M., and Williams, T.A. (2020). Phylogenomic
773 Evidence for the Monophyly of Bryophytes and the Reductive Evolution of Stomata. *Current Biology* 30,
774 2001–2012.e2. <https://doi.org/10.1016/j.cub.2020.03.048>.
- 775 73. Sibbald, S.J., and Archibald, J.M. (2020). Genomic Insights into Plastid Evolution. *Genome Biol.*
776 *Evol.* 12, 978–990. <https://doi.org/10.1093/gbe/evaa096>.
- 777 74. de Sousa, F., Foster, P.G., Donoghue, P.C.J., Schneider, H., and Cox, C.J. (2019). Nuclear protein
778 phylogenies support the monophyly of the three bryophyte groups (Bryophyta Schimp.). *New Phytologist*
779 222, 565–575. <https://doi.org/10.1111/nph.15587>.

- 780 75. Lozano-Quiles, M., Raval, P.K., Rensing, S.A., and Gould, S.B. (2026). SWI3A/B regulates the
781 transition from vegetative to reproductive phase in the liverwort *Marchantia polymorpha*. *Plant Reprod.* 39,
782 6. <https://doi.org/10.1007/s00497-026-00537-5>.
- 783 76. Renzaglia, K.S., and Garbary, D.J. (2001). Motile Gametes of Land Plants: Diversity,
784 Development, and Evolution. *CRC Crit. Rev. Plant Sci.* 20, 107–213.
785 <https://doi.org/10.1080/20013591099209>.
- 786 77. Nishiyama, T., Fujita, T., Shin-I, T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A., Carninci, P.,
787 Hayashizaki, Y., Shinozaki, K., et al. (2003). Comparative genomics of *Physcomitrella patens* gametophytic
788 transcriptome and *Arabidopsis thaliana*: Implication for land plant evolution. *Proceedings of the National*
789 *Academy of Sciences* 100, 8007–8012. <https://doi.org/10.1073/pnas.0932694100>.
- 790 78. Bohne, A., Schwenkert, S., Grimm, B., and Nickelsen, G. (2016). Roles of Tetratricopeptide Repeat
791 Proteins in Biogenesis of the Photosynthetic Apparatus. 187–227.
792 <https://doi.org/10.1016/bs.ircmb.2016.01.005>.
- 793 79. Zeytuni, N., and Zarivach, R. (2012). Structural and Functional Discussion of the Tetra-Trico-
794 Peptide Repeat, a Protein Interaction Module. *Structure* 20, 397–405.
795 <https://doi.org/10.1016/j.str.2012.01.006>.
- 796 80. Lozano-Quiles, M., Raval, P.K., and Gould, S.B. (2026). Alternative splicing of a TPR domain
797 determines mitochondrial versus plastid function of the only CLU family protein in *Marchantia polymorpha*.
798 Preprint, <https://doi.org/10.64898/2026.03.13.711532> <https://doi.org/10.64898/2026.03.13.711532>.
- 799 81. Larkin, R.M., Stefano, G., Ruckle, M.E., Stavoe, A.K., Sinkler, C.A., Brandizzi, F., Malmstrom,
800 C.M., and Osteryoung, K.W. (2016). REDUCED CHLOROPLAST COVERAGE genes from *Arabidopsis*
801 *thaliana* help to establish the size of the chloroplast compartment. *Proceedings of the National Academy of*
802 *Sciences* 113, E1116–E1125. <https://doi.org/10.1073/pnas.1515741113>.
- 803 82. Gutmann, B., Royan, S., Schallenberg-Rüdinger, M., Lenz, H., Castleden, I.R., McDowell, R.,
804 Vacher, M.A., Tonti-Filippini, J., Bond, C.S., Knoop, V., et al. (2020). The Expansion and Diversification of
805 Pentatricopeptide Repeat RNA-Editing Factors in Plants. *Mol. Plant* 13, 215–230.
806 <https://doi.org/10.1016/j.molp.2019.11.002>.
- 807 83. O'Toole, N., Hattori, M., Andres, C., Iida, K., Lurin, C., Schmitz-Linneweber, C., Sugita, M., and
808 Small, I. (2008). On the Expansion of the Pentatricopeptide Repeat Gene Family in Plants. *Mol. Biol. Evol.*
809 25, 1120–1128. <https://doi.org/10.1093/molbev/msn057>.
- 810 84. Fujii, S., and Small, I. (2011). The evolution of RNA editing and pentatricopeptide repeat genes.
811 *New Phytologist* 191, 37–47. <https://doi.org/10.1111/j.1469-8137.2011.03746.x>.
- 812 85. Tillich, M., Lehwark, P., Morton, B.R., and Maier, U.G. (2006). The Evolution of Chloroplast RNA
813 Editing. *Mol. Biol. Evol.* 23, 1912–1921. <https://doi.org/10.1093/molbev/msl054>.
- 814 86. Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B., and Brennicke, A. (2013). RNA Editing in
815 Plants and Its Evolution. *Annu. Rev. Genet.* 47, 335–352. [https://doi.org/10.1146/annurev-genet-111212-](https://doi.org/10.1146/annurev-genet-111212-816133519)
816 133519.

- 817 87. Knoop, V. (2023). C-to-U and U-to-C: RNA editing in plant organelles and beyond. *J. Exp. Bot.*
818 74, 2273–2294. <https://doi.org/10.1093/jxb/erac488>.
- 819 88. Shen, C., Xu, H., Huang, W., Zhao, Q., and Zhu, R. (2024). Is RNA editing truly absent in the
820 complex thalloid liverworts (Marchantiopsida)? *New Phytologist* 242, 2817–2831.
821 <https://doi.org/10.1111/nph.19750>.
- 822 89. Dong, S., Zhao, C., Zhang, S., Wu, H., Mu, W., Wei, T., Li, N., Wan, T., Liu, H., Cui, J., et al.
823 (2019). The Amount of RNA Editing Sites in Liverwort Organellar Genes Is Correlated with GC Content and
824 Nuclear PPR Protein Diversity. *Genome Biol. Evol.* 11, 3233–3239. <https://doi.org/10.1093/gbe/evz232>.
- 825 90. Freyer, R., Kiefer-Meyer, M.-C., and Kössel, H. (1997). Occurrence of plastid RNA editing in all
826 major lineages of land plants. *Proceedings of the National Academy of Sciences* 94, 6285–6290.
827 <https://doi.org/10.1073/pnas.94.12.6285>.
- 828 91. Waltz, F., Nguyen, T.-T., Arrivé, M., Boehler, A., Chicher, J., Hammann, P., Kuhn, L., Quadrado,
829 M., Mireau, H., Hashem, Y., et al. (2019). Small is big in Arabidopsis mitochondrial ribosome. *Nat. Plants* 5,
830 106–117. <https://doi.org/10.1038/s41477-018-0339-y>.
- 831 92. Barkan, A., and Small, I. (2014). Pentatricopeptide Repeat Proteins in Plants. *Annu. Rev. Plant*
832 *Biol.* 65, 415–442. <https://doi.org/10.1146/annurev-arplant-050213-040159>.
- 833 93. SCHMITZLINNEWEBER, C., and SMALL, I. (2008). Pentatricopeptide repeat proteins: a socket
834 set for organelle gene expression. *Trends Plant Sci.* 13, 663–670.
835 <https://doi.org/10.1016/j.tplants.2008.10.001>.
- 836 94. Okuda, K., Chateigner-Boutin, A.-L., Nakamura, T., Delannoy, E., Sugita, M., Myouga, F.,
837 Motohashi, R., Shinozaki, K., Small, I., and Shikanai, T. (2009). Pentatricopeptide Repeat Proteins with the
838 DYW Motif Have Distinct Molecular Functions in RNA Editing and RNA Cleavage in *Arabidopsis*
839 Chloroplasts. *Plant Cell* 21, 146–156. <https://doi.org/10.1105/tpc.108.064667>.
- 840 95. Meierhoff, K., Felder, S., Nakamura, T., Bechtold, N., and Schuster, G. (2003). HCF152, an
841 Arabidopsis RNA Binding Pentatricopeptide Repeat Protein Involved in the Processing of Chloroplast *psbB*-
842 *psbT-psbH-petB-petD* RNAs. *Plant Cell* 15, 1480–1495. <https://doi.org/10.1105/tpc.010397>.
- 843 96. Schmitz-Linneweber, C., Williams-Carrier, R.E., Williams-Voelker, P.M., Kroeger, T.S., Vichas,
844 A., and Barkan, A. (2006). A Pentatricopeptide Repeat Protein Facilitates the *trans*- Splicing of the Maize
845 Chloroplast *rps12* Pre-mRNA. *Plant Cell* 18, 2650–2663. <https://doi.org/10.1105/tpc.106.046110>.
- 846 97. Schmitz-Linneweber, C., Williams-Carrier, R., and Barkan, A. (2005). RNA Immunoprecipitation
847 and Microarray Analysis Show a Chloroplast Pentatricopeptide Repeat Protein to Be Associated with the 5'
848 Region of mRNAs Whose Translation It Activates. *Plant Cell* 17, 2791–2804.
849 <https://doi.org/10.1105/tpc.105.034454>.
- 850 98. Bieri, P., Leibundgut, M., Saurer, M., Boehringer, D., and Ban, N. (2017). The complete structure
851 of the chloroplast 70S ribosome in complex with translation factor pY. *EMBO J.* 36, 475–486.
852 <https://doi.org/10.15252/embj.201695959>.

- 853 99. Waltz, F., Lehner, P.A., Van der Stappen, P., Kater, L., Pfeffer, S., and Engel, B.D. (2026).
854 Chloroplast-encoded small subunit extensions reshape the Chlamydomonas chlororibosome. *bioRxiv*.
855 <https://doi.org/10.64898/2026.02.07.704542>.
- 856 100. Barrio-Hernandez, I., Yeo, J., Jänes, J., Mirdita, M., Gilchrist, C.L.M., Wein, T., Varadi, M.,
857 Velankar, S., Beltrao, P., and Steinegger, M. (2023). Clustering predicted structures at the scale of the known
858 protein universe. *Nature* 622, 637–645. <https://doi.org/10.1038/s41586-023-06510-w>.
- 859 101. Baltzis, A., Santus, L., Langer, B.E., Magis, C., de Vienne, D.M., Gascuel, O., Mansouri, L., and
860 Notredame, C. (2025). multistrap: boosting phylogenetic analyses with structural information. *Nat. Commun.*
861 16, 293. <https://doi.org/10.1038/s41467-024-55264-0>.
- 862 102. Kortemme, T. (2024). De novo protein design—From new structures to programmable functions.
863 *Cell* 187, 526–544. <https://doi.org/10.1016/j.cell.2023.12.028>.
- 864 103. Lauko, A., Pellock, S.J., Sumida, K.H., Anishchenko, I., Juergens, D., Ahern, W., Jeung, J., Shida,
865 A.F., Hunt, A., Kalvet, I., et al. (2025). Computational design of serine hydrolases. *Science* (1979). 388.
866 <https://doi.org/10.1126/science.adu2454>.
- 867 104. Listov, D., Vos, E., Hoffka, G., Hoch, S.Y., Berg, A., Hamer-Rogotner, S., Dym, O., Kamerlin,
868 S.C.L., and Fleishman, S.J. (2025). Complete computational design of high-efficiency Kemp elimination
869 enzymes. *Nature* 643, 1421–1427. <https://doi.org/10.1038/s41586-025-09136-2>.
- 870 105. Wu, D., Jospin, G., and Eisen, J.A. (2013). Systematic Identification of Gene Families for Use as
871 “Markers” for Phylogenetic and Phylogeny-Driven Ecological Studies of Bacteria and Archaea and Their
872 Major Subgroups. *PLoS One* 8, e77033. <https://doi.org/10.1371/journal.pone.0077033>.
- 873 106. Rivalta, A., Hiregange, D.-G., Bose, T., Rajan, K.S., Yonath, A., Zimmerman, E., Waghalter, M.,
874 Fridkin, G., Martinez-Roman, I., Rosenfield, L., et al. (2025). Ribosomes: from conserved origin to
875 functional/medical mobility and heterogeneity. *Philosophical Transactions of the Royal Society B: Biological*
876 *Sciences* 380. <https://doi.org/10.1098/rstb.2023.0393>.
- 877 107. Trost, K., Knopp, M.R., Wimmer, J.L.E., Tria, F.D.K., and Martin, W.F. (2024). A universal and
878 constant rate of gene content change traces pangenome flux to LUCA. *FEMS Microbiol. Lett.* 371.
879 <https://doi.org/10.1093/femsle/fnae068>.
- 880 108. Nagies, F.S.P., Brueckner, J., Tria, F.D.K., and Martin, W.F. (2020). A spectrum of verticality across
881 genes. *PLoS Genet.* 16, e1009200. <https://doi.org/10.1371/journal.pgen.1009200>.
- 882 109. de Pins, B., Malbrancke, C., Jabłońska, J., Shmuel, A., Sharon, I., Bitbol, A.-F., Mueller-Cajar, O.,
883 Noor, E., and Milo, R. (2025). Rubisco is slow across the tree of life. *Proceedings of the National Academy*
884 *of Sciences* 122. <https://doi.org/10.1073/pnas.2501433122>.
- 885 110. Prywes, N., Phillips, N.R., Oltrogge, L.M., Lindner, S., Taylor-Kearney, L.J., Tsai, Y.-C.C., de Pins,
886 B., Cowan, A.E., Chang, H.A., Wang, R.Z., et al. (2025). A map of the rubisco biochemical landscape. *Nature*
887 638, 823–828. <https://doi.org/10.1038/s41586-024-08455-0>.
- 888 111. Bar-On, Y.M., Phillips, R., and Milo, R. (2018). The biomass distribution on Earth. *Proc. Natl.*
889 *Acad. Sci. U. S. A.* 115, 6506–6511. <https://doi.org/10.1073/pnas.1711842115>.

890 112. Antonovsky, N., Gleizer, S., Noor, E., Zohar, Y., Herz, E., Barenholz, U., Zelcbuch, L., Amram, S.,
891 Wides, A., Tepper, N., et al. (2016). Sugar Synthesis from CO₂ in *Escherichia coli*. *Cell* *166*, 115–125.
892 <https://doi.org/10.1016/j.cell.2016.05.064>.

893 113. Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W. (1999). From molecular to modular
894 cell biology. *Nature* *402*, C47–C52. <https://doi.org/10.1038/35011540>.

895 114. Pereira-Leal, J.B., Levy, E.D., and Teichmann, S.A. (2006). The origins and evolution of functional
896 modules: lessons from protein complexes. *Philosophical Transactions of the Royal Society B: Biological*
897 *Sciences* *361*, 507–517. <https://doi.org/10.1098/rstb.2005.1807>.

898 115. Ishizaki, K., Nishihama, R., Ueda, M., Inoue, K., Ishida, S., Nishimura, Y., Shikanai, T., and
899 Kohchi, T. (2015). Development of gateway binary vector series with four different selection markers for the
900 liverwort *Marchantia polymorpha*. *PLoS One* *10*. <https://doi.org/10.1371/journal.pone.0138876>.

901 116. Ishizaki, K., Chiyoda, S., Yamato, K.T., and Kohchi, T. (2008). *Agrobacterium*-mediated
902 transformation of the haploid liverwort *Marchantia polymorpha* L., an emerging model for plant biology.
903 *Plant Cell Physiol.* *49*, 1084–1091. <https://doi.org/10.1093/pcp/pcn085>.

904 117. Emms, D.M., and Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative
905 genomics. *Genome Biol.* *20*, 1–14. <https://doi.org/10.1186/s13059-019-1832-y>.

906 118. Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for
907 rapid unsupervised cryo-EM structure determination. *Nat. Methods* *14*, 290–296.
908 <https://doi.org/10.1038/nmeth.4169>.

909 119. Punjani, A., Zhang, H., and Fleet, D.J. (2020). Non-uniform refinement: adaptive regularization
910 improves single-particle cryo-EM reconstruction. *Nat. Methods* *17*, 1214–1221.
911 <https://doi.org/10.1038/s41592-020-00990-8>.

912 120. Zivanov, J., Nakane, T., and Scheres, S.H.W. (2019). A Bayesian approach to beam-induced motion
913 correction in cryo-EM single-particle analysis. *IUCrJ* *6*, 5–17. <https://doi.org/10.1107/S205225251801463X>.

914 121. Meng, E.C., Goddard, T.D., Pettersen, E.F., Couch, G.S., Pearson, Z.J., Morris, J.H., and Ferrin,
915 T.E. (2023). UCSF ChimeraX: Tools for structure building and analysis. *Protein Science* *32*.
916 <https://doi.org/10.1002/pro.4792>.

917 122. Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A., Carazo, J.M., Sorzano, C.O.S., and Vargas, J.
918 (2021). DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* *4*,
919 874. <https://doi.org/10.1038/s42003-021-02399-1>.

920 123. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment
921 search tool. *J. Mol. Biol.* *215*, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).

922 124. Croll, T.I. (2018). ISOLDE: a physically realistic environment for model building into low-
923 resolution electron-density maps. *Acta Crystallogr. D Struct. Biol.* *74*, 519–530.
924 <https://doi.org/10.1107/S2059798318002425>.

925 125. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of *Coot*.
926 *Acta Crystallogr. D Biol. Crystallogr.* *66*, 486–501. <https://doi.org/10.1107/S0907444910007493>.