On the Phylogenetic Distribution of Plastid Developmental Components in the Chloroplastida

Inaugural dissertation

for the attainment of the title of doctor
in the Faculty of Mathematics and Natural Sciences
at the Heinrich Heine University of Düsseldorf

presented by

Alexander István MacLeod
From Târgu Mureș (Marosvásárhely), Transylvania, Romania
&
Isle of Skye, Scotland

Düsseldorf, September 2023

from the Institute of Molecular Evolution

at the Heinrich Heine University Düsseldorf

Published by the permission of the

Faculty of Mathematics and Natural Sciences at

Heinrich Heine University Düsseldorf

Supervisor: Prof. Dr. Sven B. Gould

Co-supervisor: Prof. Dr. William F. Martin

Date of examination: TBD

For my partner and family.

Without your love, help and encouragement, I would not be in this privileged position.

Thank you for everything.

Statement of authorship

This dissertation is the result of my own work. No other person's work has been used without due acknowledgement. This dissertation has not been submitted in the same or similar form to other institutions. I have not previously failed a doctoral examination procedure.

Alexander István MacLeod

Table of Contents

1 Summary 1

2 Zusammenfassung 2

3 List of publications included in this thesis 3

4 Introduction 4

4.1 The Chloroplastida and streptophyte terrestrialization 4

4.2 Plastid evolution 7

4.3 Plastid division and the monoplastidic bottleneck 9

4.4 Protein sorting into the thylakoid membranes 15

4.5 The TOC/TIC import machinery 17

4.6 Hornworts 19

4.7 The evo-devo pipeline in researching plant evolution 20

5 Aims of this thesis 24

6 Summary of obtained results 25

7 Conclusions 27

8 Results associated with this thesis 28

8.1 Publication I 28

8.2 Publication II 36

8.3 Publication III 43

9 Acknowledgements 71

10 References 72

1 Summary

Plastids are organelles found in all photosynthetic organisms that are descended from a primary endosymbiotic event involving a heterotrophic protist engulfing a free-living cyanobacterium. This organelle has paved the way for the emergence, evolution and diversification of all floral – and some faunal – lineages: most notably, the Chloroplastida. Also known as the "green lineage", the Chloroplastida are a diverse supergroup of organisms ranging from unicellular microplankton to giant redwood trees. Despite this morphological diversity, one defining cellular trait unifies all members of the Chloroplastida: the plastid, specifically the chloroplast. Recent advances in the development of new algorithms in phylogenomics, coupled with the publishing of novel plant and algal genome data, creates a tremendous opportunity to study the evolution and diversity of plastid biology in the green lineage: from algae to angiosperms. This thesis aims to do just that by focussing on plastid cellular pathways involved in regulating plastid development and biogenesis. Most land plants (embryophytes) are polyplastidic, and house multiple plastids per-cell: except for hornworts. Hornworts are a unique group of bryophytes, the non-vascular sister group to the vascular tracheophytes, and are cytologically distinct from most embryophytes. They are monoplastidic (or near-monoplastidic in some cases) and some genera possess a pyrenoid in their cells, which is a unique carbon concentrating mechanism found mainly in algae. Hornwort emergence and diversification was accompanied by significant instances of gene loss, including two genes involved in regulating plastid division: FtsZ2 and ARC3. Ancestral state reconstruction suggests that this differential loss of FtsZ2 and ARC3 correlates with hornworts reverting to a monoplastidic phenotype, which the ancestral embryophyte and bryophyte were able to escape. Indeed, *ftsz2* and *arc3* mutants from the cress *Arabidopsis thaliana* and the more closely related moss *Physcomitrium patens* result in monoplastidic/near-monoplastidic phenotypes, so it does not stand to reason that monoplastidy in hornworts is a result of the combined loss of these two genes. The peptidoglycan (PG) layer of chloroplasts was initially characterised in moss, but little was known about its phylogenetic distribution in the green lineage; the availability of novel genome data has changed this, showing that this cyanobacterial relic appears to be more widespread than previously thought in the Chloroplastida. While the full enzymatic toolkit for PG layer biosynthesis is conserved in all bryophytes and most streptophyte algae, its phylogenetic distribution becomes more patchier in vascular plants and chlorophyte algae, with it being present in at least three phylogenetically distant chlorophytes and three phylogenetically distant seed plants (spermatophytes). However, there appears to be strict structural conservation in the enzymes involved in PG layer biosynthesis – from cyanobacteria to spermatophytes – indicating that the biochemical function of these enzymes have remained unchanged from primary endosymbiosis. Furthermore, it is also unlikely that the chloroplast PG layer coevolved with the plastid division component FtsZ3, as previously once thought, as there are many plants and algae that encode a full biosynthetic toolkit for PG layer biosynthesis but lack FtsZ3.

2 Zusammenfassung

Plastiden sind Organellen, die in allen photosynthetischen Organismen vorkommen und von einem primären Endosymbiose-Ereignis abstammen, bei dem ein heterotropher Protist einen freilebenden Cyanobakterium verschluckt hat. Diese Organellen haben den Weg für das Aufkommen, die Evolution und Diversifizierung aller floralen - und einiger faunaler - Linien geebnet, insbesondere der Chloroplastida. Auch bekannt als die "grüne Linie", sind die Chloroplastida eine diverse Übergruppe von Organismen, die von einzelligen Mikroplankton bis hin zu riesigen Mammutbäumen reichen. Trotz dieser morphologischen Vielfalt vereint ein definiertes zelluläres Merkmal alle Mitglieder der Chloroplastida: die Plastiden, insbesondere die Chloroplasten. Fortschritte bei der Entwicklung neuer Algorithmen in der Phylogenomik, gekoppelt mit der Veröffentlichung neuer Pflanzen- und Algen-Genomdaten, schaffen eine enorme Chance, die Evolution und Vielfalt der Plastidenbiologie in der grünen Linie zu untersuchen: von Algen bis zu Angiospermen. Diese Arbeit zielt darauf ab, genau dies zu tun, indem sie sich auf Plastidzellwege konzentriert, die an der Regulierung der Plastidentwicklung und -biogenese beteiligt sind. Die meisten Landpflanzen (Embryophyten) sind polyplastidisch und beherbergen mehrere Plastiden pro Zelle, mit Ausnahme von Hornmoosen. Hornmoose sind eine einzigartige Gruppe von Bryophyten, der nichtgefäßbildenden Schwestergemeinschaft zu den gefäßbildenden Tracheophyten, und sind zytologisch von den meisten Embryophyten unterschiedlich. Sie sind monoplastidisch (oder in einigen Fällen nahezu monoplastidisch) und einige Gattungen besitzen einen Pyrenoiden in ihren Zellen, der ein einzigartiger Kohlenstoff-Konzentrationsmechanismus ist, der hauptsächlich in Algen vorkommt. Die Entstehung und Diversifizierung von Hornmoosen wurde von signifikanten Fällen von Genverlust begleitet, einschließlich zweier Gene, die an der Regulation der Plastidteilung beteiligt sind: FtsZ2 und ARC3. Die Rekonstruktion des ancestralen Zustands legt nahe, dass dieser unterschiedliche Verlust von FtsZ2 und ARC3 mit der Rückkehr von Hornmoosen zu einem monoplastidischen Phänotyp korreliert, den der ancestrale Embryophyt und Bryophyt entkommen konnten. Tatsächlich führen ftsz2- und arc3-Mutanten der Kresse Arabidopsis thaliana und des näher verwandten Mooses Physcomitrium patens zu monoplastidischen/nahezu monoplastidischen Phänotypen, sodass es nicht vernünftig erscheint, dass die Monoplastidie bei Hornmoosen auf den kombinierten Verlust dieser beiden Gene zurückzuführen ist. Die Peptidoglykanschicht (PG-Schicht) von Chloroplasten wurde zunächst in Moosen charakterisiert, aber wenig war über ihre phylogenetische Verteilung in der grünen Linie bekannt. Die Verfügbarkeit neuer Genomdaten hat dies geändert und gezeigt, dass dieses Cyanobakterien-Relikt in den Chloroplastida weit verbreiteter zu sein scheint als zuvor angenommen. Während das vollständige enzymatische Toolkit für die PG-Schicht-Biosynthese in allen Bryophyten und den meisten Streptophyten-Algen erhalten bleibt, wird ihre phylogenetische Verteilung in Gefäßpflanzen und Chlorophyten-Algen unregelmäßiger, wobei sie bei mindestens drei phylogenetisch entfernten Chlorophyten und drei phylogenetisch entfernten Samenpflanzen (Spermatophyten) vorhanden ist. Es gibt jedoch eine strikte strukturelle Konservierung in den Enzymen, die an der Biosynthese der PG-Schicht beteiligt sind - von Cyanobakterien bis zu Spermatophyten -, was darauf hinweist, dass die biochemische Funktion dieser Enzyme seit der primären Endosymbiose unverändert geblieben ist. Darüber hinaus ist es unwahrscheinlich, dass die Chloroplasten-PG-Schicht gemeinsam mit dem Plastidteilungsbestandteil FtsZ3 koevolviert ist, wie früher angenommen wurde, da es viele Pflanzen und Algen gibt, die ein vollständiges biosynthetisches Toolkit für die PG-Schicht-Biosynthese codieren, aber kein FtsZ3 besitzen.

3 List of publications included in this thesis

<u>Publication I</u>

MacLeod, A. I., Raval, P. K., Stockhorst, S., Knopp, M. R., Frangedakis, E., and Gould, S. B. (2022). Loss of Plastid Developmental Genes Coincides With a Reversion to Monoplastidy in Hornworts. *Frontiers in Plant Science* (doi: 10.3389/fpls.2022.86307)

<u>Publication II</u>

MacLeod, A. I., Knopp, M. R., and Gould, S. B. (2023). A mysterious cloak: the peptidoglycan layer of algal and plant plastids. *Protoplasma* (doi: 10.1007/s00709-023-01886-y)

<u>Publication III</u>

Raval, P. K., **MacLeod, A. I.**, and Gould, S. B. (2023). A molecular atlas of plastid and mitochondrial adaptations across the evolution from chlorophyte algae to angiosperms. *bioRxiv* preprint submitted to *PLOS Biology* as of 06.09.2023. (doi: 10.1101/2023.09.01.555919)

4 Introduction

4.1 The emergence and significance of the Chloroplastida, with special emphasis on the terrestrialization of Streptophytes. Fewer natural phenomena have been as radically transformative to Plant Earth as its global greening by plants and algae (Lenton et al., 2016; Leebens-Mack et al., 2019). Carried out via the emergence and expansion of the Chloroplastida, which are members of the larger Archaeplastida supergroup, this greening was arguably one of the most important events in Earth's history and one of the major transitions of evolution of life on this planet. Not only did the emergence and evolution of the green lineage change Earth's geochemistry and geobiology, but it also gave rise to the terrestrial and aquatic floral and faunal biodiversity we appreciate to this day (Delwiche and Timme, 2011; de Vries et al., 2016; Lenton et al., 2016, 2016; Morris et al., 2018; Leebens-Mack et al., 2019; Dahl and Arens, 2020; Li et al., 2020b; Su et al., 2021; Schreiber et al., 2022). Indeed, the key ecological role played by the Chloroplastida is the result of this supergroup's continuous expansion to fill and support terrestrial, and aquatic, habitats of an ever-changing world (Rensing, 2018). Notable examples of such adaptive radiations include the conquest of glacial habitats by the Anydrophyta in the Cryogenian period, to the conquest of land by embryophytes (land plants) between 600-950 million years ago (Žárský et al., 2022). The green supergroup is made up of three phyla: the Streptophyta, Chlorophyta and the basal-branching Prasinodermatophyta (Li et al., 2020b) (*Figure 1*).

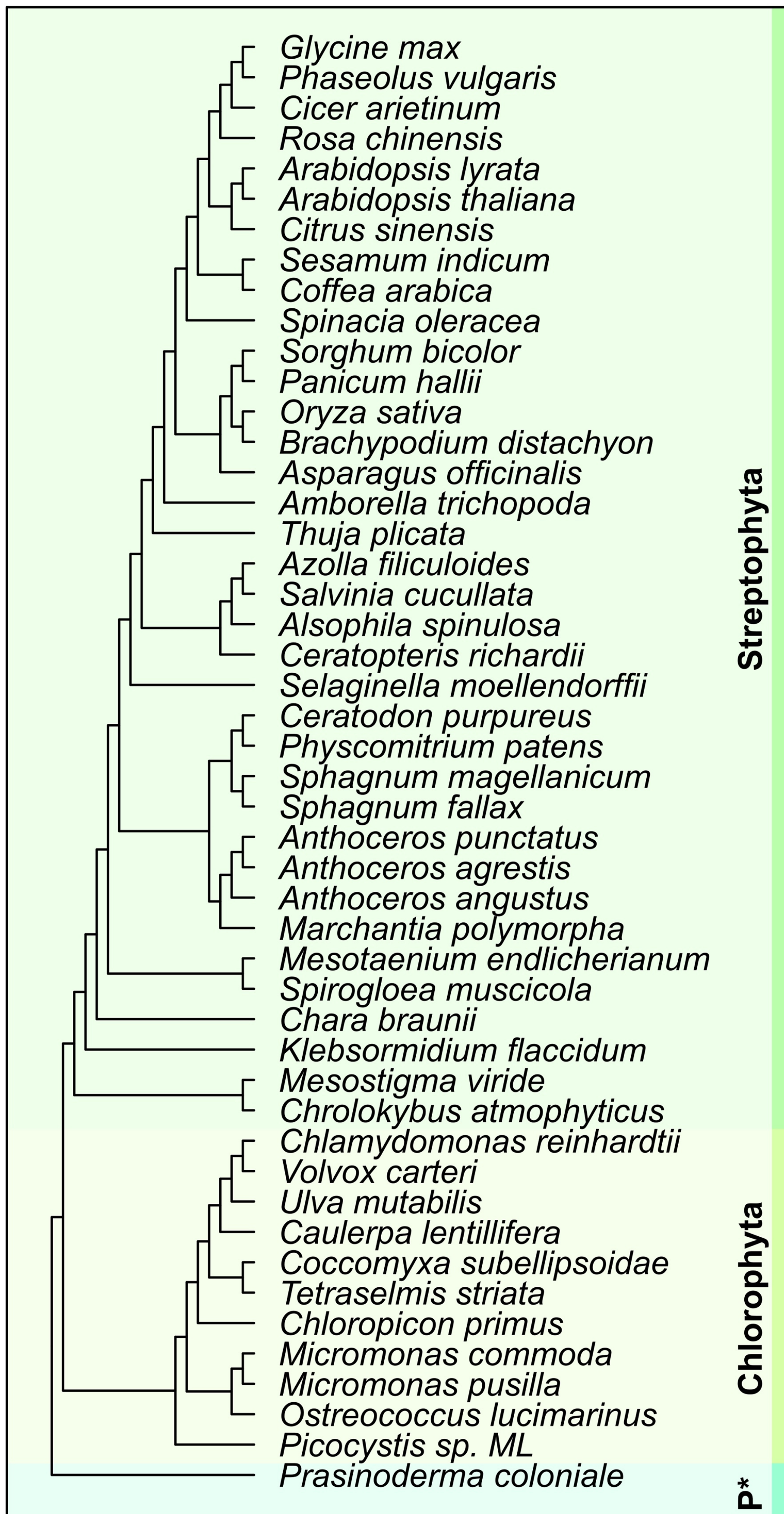
Glycine max
Phaseolus vulgaris
Cicer arietinum
Rosa chinensis
Arabidopsis lyrata
Arabidopsis thaliana
Citrus sinensis
Sesamum indicum
Coffea arabica
Spinacia oleracea
Sorghum bicolor
Panicum hallii
Oryza sativa
Brachypodium distachyon
Asparagus officinalis
Amborella trichopoda
Thuja plicata
Azolla filiculoides
Salvinia cucullata
Alsophila spinulosa
Ceratopteris richardii
Selaginella moellendorffii
Ceratodon purpureus
Physcomitrium patens
Sphagnum magellanicum
Sphagnum fallax
Anthoceros punctatus
Anthoceros agrestis
Anthoceros angustus
Marchantia polymorpha
Mesotaenium endlicherianum
Spirogloea muscicola
Chara braunii
Klebsormidium flaccidum
Mesostigma viride
Chrolokybus atmophyticus
Chlamydomonas reinhardtii
Volvox carteri
Ulva mutabilis
Caulerpa lentillifera
Coccomyxa subellipsoidae
Tetraselmis striata
Chloropicon primus
Micromonas commoda
Micromonas pusilla
Ostreococcus lucimarinus
Picocystis sp. ML
Prasinoderma coloniale
Streptophyta
Chlorophyta
P*

***Figure 1*: Cladogram showing the evolutionary relationships between the Chloroplastida.** Topology is based on Leebens-Mack et al. 2019 and Li et al. 2021. *P**; Prasinodermophyta

The Prasinodermatophyta are a group of unicellular green algae. They are believed to be one of the earliest branches of the green algae; as they form a monophyletic sister group to the Chlorophyta and Streptophyta (*Figure 1*). Prasinodermatophytes are only found in marine habitats, where they play an important role in the ecosystem by serving as primary producers in these marine communities, and where they form important components of phytoplankton communities (Li et al., 2020b). So far, only one species of this phylum, *Prasinoderma coloniale,* has been sequenced.

The Chlorophyta are also a group of green algae and, unlike their Prasinodermatophyta, can occupy freshwater, marine and terrestrial habitats (Li et al., 2021). In terms of size and morphology, they range from large multicellular seaweeds to tiny unicellular microplankton (Delwiche and Timme, 2011; Leliaert et al., 2012).

The Streptophyta are comprised of freshwater streptophyte algae and the (predominantly) terrestrial embryophytes. Streptophyte algae are morphologically diverse, with some members being unicellular, while others are multicellular. Embryophytes, on the other hand, are completely multicellular (Leebens-Mack et al., 2019). And one of the most significant moments in the evolutionary history of streptophytes was the conquest of land by plants.

The conquest of land by embryophytes was one of the most significant events of life on Earth. This moment marked a significant turning point in the history of life itself, as it allowed for the colonisation of land by complex multicellular organisms, leading to the eventual establishment of complex ecosystems (Lenton et al., 2016; de Vries and Archibald, 2018). Depending on the dating method, streptophyte terrestrialization is believed to have happened between 450-900 million years ago (Bowles et al., 2023). This event was initiated by streptophyte algae (specifically, an extinct strain related to the Zygnematophyceae), which gradually adapted to survive in the harsh terrestrial environment (Cheng et al., 2019; Žárský et al., 2022). Over evolutionary time, these algae evolved into multicellular embryophytes that were better adapted to cope with terrestrial stressors, leading to them dominating land (Rensing, 2018). Indeed, plants played a crucial role in the formation of the very first terrestrial ecosystems, as they were primary producers. And as more and more plants began to emerge and diversify on land, these ecosystems became more complex, and a lot more species began to interact with each other in different ways (de Vries and Archibald, 2018; Rensing, 2018).

The significance plant terrestrialization had on Earth's environment cannot be overstated. Four major components of Earth's geology and geochemistry were affected via the emergence of embryophytes: temperature, the water cycle, soil, and the atmosphere. Via the biochemical process of photosynthesis, plants absorb carbon dioxide from the atmosphere, which helps regulated the temperature of Earth, as this greenhouse gas is known to trap heat in Earth's atmosphere. Therefore, the more carbon dioxide that is present in Earth's atmosphere, the warmer the Earth's temperature will be. Therefore, as more plants conquered terrestrial habitats, the levels of carbon dioxide in the atmosphere decreased, leading to a significant cooling of Earth's climate (Arneth et al., 2017). Plant terrestrialization also had an impact on Earth's water cycle. This is because plants can absorb water from soil microenvironments and release it into the atmosphere; a physiological process known as transpiration (which plays crucial roles in the formation of clouds and the regulation of rain). Therefore, as more embryophytes established themselves on land, this led to a more efficient regulation of the water cycle (PITTERMANN, 2010). Furthermore, as plants died, decayed, and decomposed, the fertility of Earth's soil increased, due to the build-up of organic compounds. This allowed for more efficient growth of subsequent plants (Tiessen et al., 1994). Finally, plants had a major and significant impact on Earth's atmosphere. This is because oxygen is a by-product of photosynthesis, and as more plants dominated the terrestrial environment of Earth, this resulted in a subsequent increase in oxygen. Indeed, it is hypothesised that modern levels of oxygen are a result of Streptophyte terrestrialization (Lenton et al., 2016).

And while they might be morphologically, ecologically and cytologically diverse, there is one organelle that defines Chloroplastidal identity: the plastid (specifically, the chloroplast) (de Vries et al., 2016).

4.2 Evolution and diversity of plastids in the Chloroplastida. The first plastid emerged roughly between 1.2 to 2.7 billion years ago (Bowles et al., 2023), when a free-living cyanobacterium was engulfed by a photosynthetic protist, giving rise to the photosynthetic eukaryotes – the Archaeplastida. The Archaeplastida are mainly comprised of three ancient groups: Rhodophyta, Glaucophyta and the Chloroplastida (Keeling, 2010, 2013; Gawryluk et al., 2019; Schön et al., 2021) (*Figure 2*).

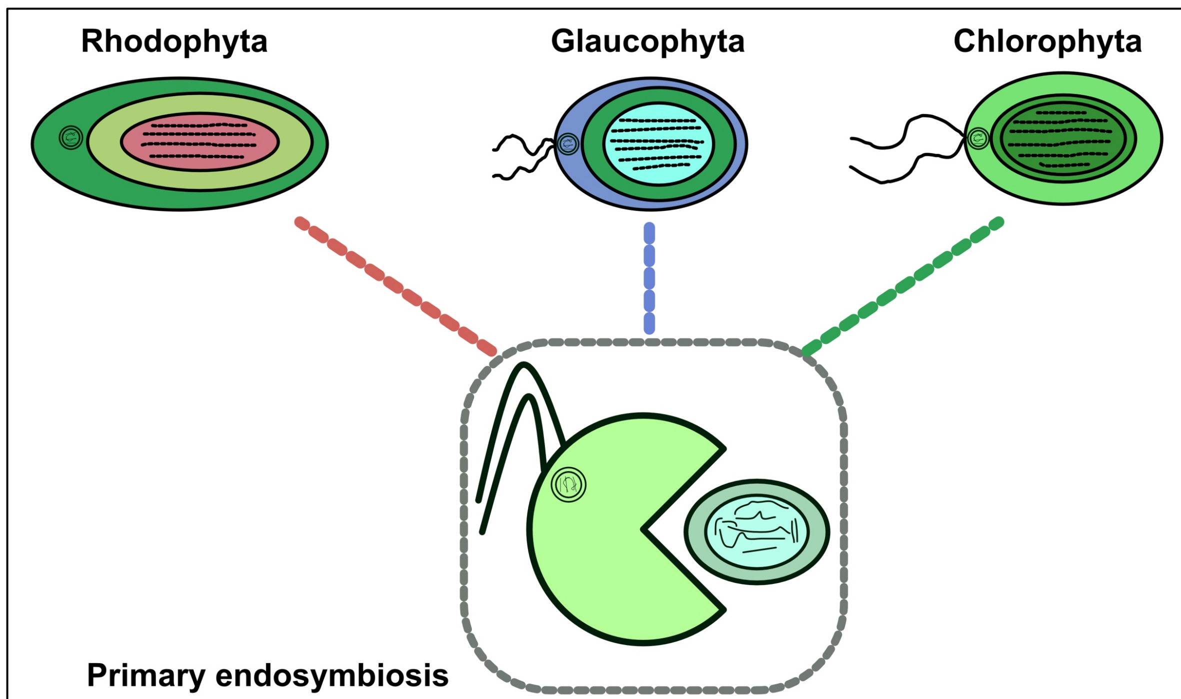

***Figure 2*. Origin of primary plastids and the emergence of the three main groups of the Archaeplastida.**

Plastids are ubiquitous organelles in most of the Archaeplastida, essential for photosynthesis and other physiological processes (de Vries et al., 2016; Figueroa-Martinez et al., 2017; Gawryluk et al., 2019). These unique organelles have paved successful trajectories in eukaryotes, a notable example being the greening ashore: the conquest of land by embryophytes (Keeling, 2013; de Vries et al., 2016; Schreiber et al., 2022). The plastids of the chloroplastida (chloroplasts) are characterised by the presence of photosynthetic pigments – chlorophylls – that are used to absorb light for photosynthesis (Delwiche and Timme, 2011). However, there are some notable exceptions to this phenomenon. The chlorophyte alga *Polytoma uvella*, for example, is non-photosynthetic and uses its plastid – the leucoplast – for starch storage and metabolism (Figueroa-Martinez et al., 2017).

Regardless, most chloroplastidal cells only house chloroplasts. However, ferns and seed plants evolved the ability for plastids to interconvert between different plastid types, of which there are at least six in spermatophytes: gerontoplasts, elaioplasts, etioplasts, chromoplasts, amyloplasts and chloroplasts (Duckett and Ligrone, 1993; Jarvis and López-Juez, 2013; de Vries et al., 2016; Choi et al., 2021). This is a phenomenon known as multifaceted polyplastidy (de Vries and Gould, 2018). These plastids, in turn, can be transported to various parts of the plant to fulfill their physiological roles (Jarvis and López-Juez, 2013). Chloroplasts, as discussed previously, function as the photosynthetic powerhouses of plant cells (Berg et al., 2015). Gerontoplasts are derived from chloroplasts in areas of the leaf undergoing senescence; while etioplasts function as developmental precursor structures, that form in dark conditions, from which chloroplasts can develop. The other three plastids play key roles in metabolism and

physiology. Chromoplasts are involved in carotenoid pigment accumulation; amyloplasts play important roles in carbohydrate metabolism and gravitropism in seeds, roots and seeds; while elailoplasts function as lipid droplet storage centers that play important roles in, for example, pollen development (Choi et al., 2021). Most of these multifaceted plastids, with the exception of gerontoplasts (which derive from chloroplasts), develop from a preliminary proplastid. However, it is also possible for some of these plastids to interconvert. For example, chloroplasts can develop from a proplastid *and* an etioplast. Similarly, chloroplasts can also revert to an etioplast, and gerontoplasts can revert to a chloroplast (Jarvis and López-Juez, 2013; de Vries et al., 2016; Choi et al., 2021).

4.3 Plastid division and the monoplastidic bottleneck. If plastids are central to embryophyte form and function, then the plastid division machinery (PDVM) is central to *plastid* identity and function, as plastid development, biochemistry and cell biology cannot be achieved without the unique PDVM evolved by land plants (Jarvis and López-Juez, 2013; Osteryoung and Pyke, 2014; MacLeod et al., 2022). Molecular studies of the PDVM in embryophytes have largely been restricted to the angiosperm *Arabidopsis thaliana*, and to a lesser extent the moss *Physcomitrium patens* (Osteryoung & Pyke, 2014) (*Figure 3*). In *A. thaliana*, the GTPases FtsZ1 and FtsZ2 are involved in regulating the formation of the inner plastid division ring, or Z-ring. Both FtsZ1 and FtsZ2, that evolved from a duplication event in an ancestral cyanobacterial FtsZ (Grosche and Rensing, 2017), form a dynamic heteropolymer to regulate inner division (Yoshida et al., 2016). A third protein, FtsZ3, also likely plays a crucial role in Z-ring formation in some species via similar methods as its other two FtsZ cousins (Grosche and Rensing, 2017). In embryophytes, FtsZ3 is only found in bryophytes and the vascular lycophytes, with it being differentially lost in ferns and spermatophytes (seed plants) (Grosche and Rensing, 2017; MacLeod et al., 2022, 2023). The proteins ARC3 and ARC6 also play key roles in regulating Z-ring formation and function (Glynn et al., 2008; Zhang et al., 2013). ARC3 is a chimera that unites an FtsZ and MORN domain (Shimada et al., 2004) and arose early in land plant evolution as it was likely present in the ancestral embryophyte (MacLeod et al., 2022). Yeast-two-hybrid and cell biological analyses suggest that ARC3 interacts directly with FtsZ2 to negatively regulate Z-ring formation, thereby ensuring synchronized organellar division (Zhang et al., 2013). ARC6 is a J-domain protein that evolved from the cyanobacterial division protein Ftn2 (Vitha et al., 2003), and like ARC3, acts as a negative regulator of Z-ring formation (Glynn et al., 2008). However, ARC6 also acts as a precursor of outer plastid division ring formation via its recruitment of the PDV2 protein (Glynn et al., 2008). PDV2 evolved in streptophyte algae (MacLeod et al., 2022) and acts downstream of ARC6 by regulating the formation of the outer division ring (Miyagishima et al., 2006; Sun et al., 2020). Together with the PDV1 protein, which likely evolved in tracheophytes (MacLeod et al., 2022), PDV2 recruits the dynamin-like ARC5. ARC5, in turn, polymerizes on the chloroplast's outer surface, leading to the formation of an outer division ring (Gao et al., 2003). In species that possess orthologues, FtsZ3 is also likely involved in outer ring division formation and regulation, although it is unknown whether it associates with ARC5 (Martin et al., 2009). Subsequent constriction via the synchronized action of both the Z- and outer rings leads to organelle division. The peptidoglycan (PG) layer of some embryophytes, a relic of the chloroplast's

cyanobacterial past, likely plays a role in the PDVM, like it does in mosses (Machida et al., 2006; Grosche and Rensing, 2017) (*Figure 3*).

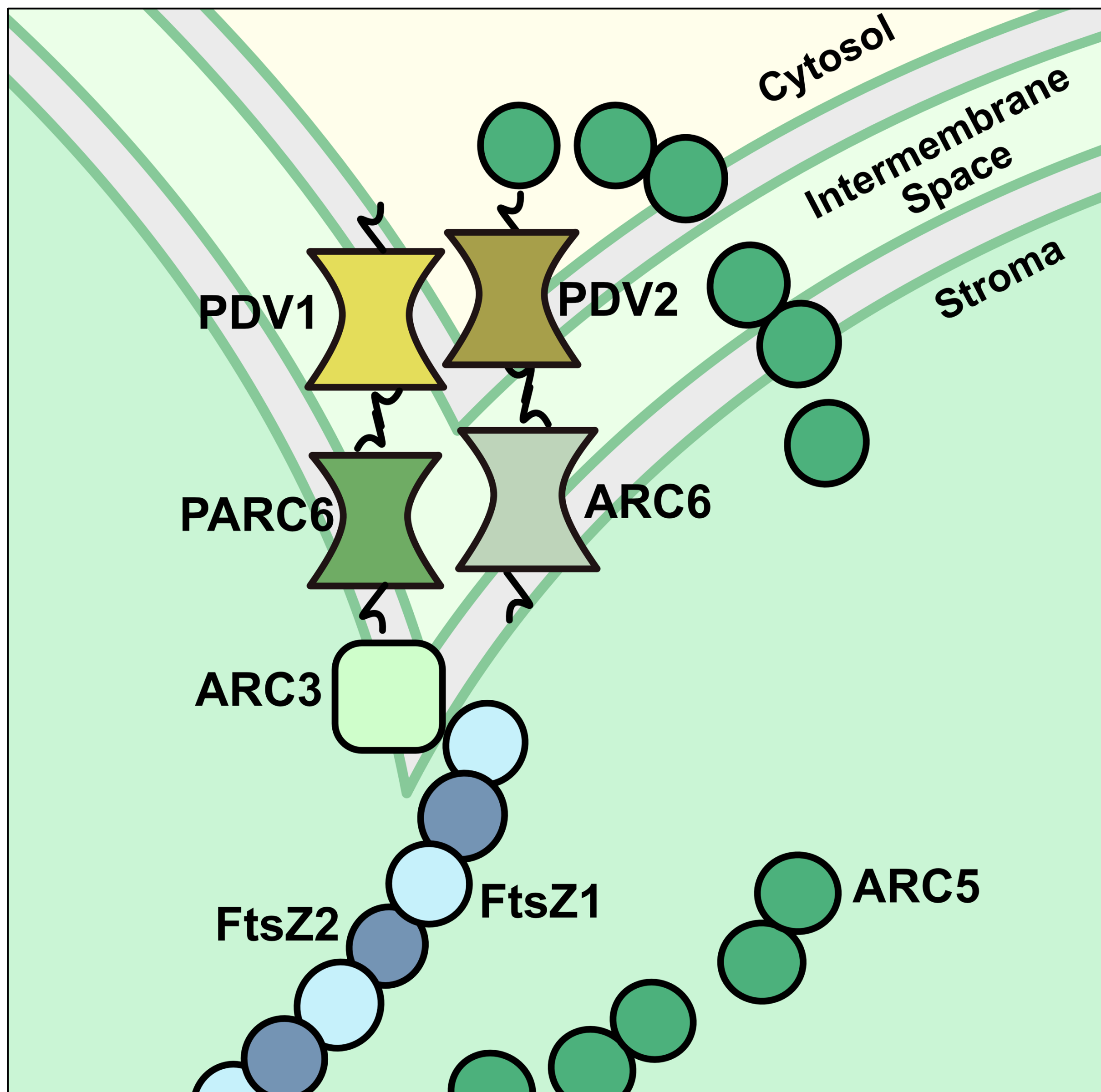

Figure 3*. PDVM of the thale cress *Arabidopsis thaliana

The increase in the availability of non-angiosperm streptophyte genome and transcriptome assemblies (Goodstein et al., 2012; Hori et al., 2014; Lang et al., 2018; Nishiyama et al., 2018; Cheng et al., 2019; Leebens-Mack et al., 2019; Li et al., 2020a; Wang et al., 2020; Zhang et al., 2020), coinciding with advances in phylogenomic and bioinformatics methods (Buchfink et al., 2015; Emms and Kelly, 2019, 2022), have provided tremendous opportunities to study the evolution of plastid division in the plants and algae, particularly in embryophytes (*Figure 4*).

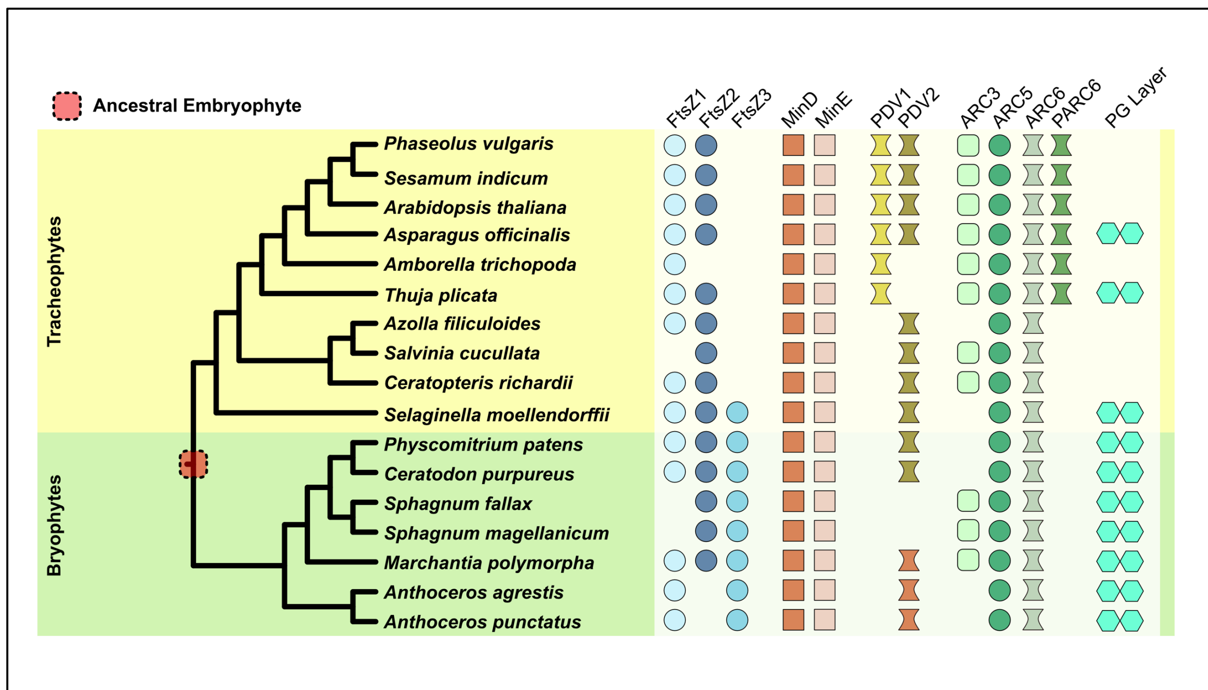

***Figure 4*. Phylogenetic distribution of PDVM components in the terrestrial clade**. Orthologue metadata are derived from MacLeod et al., 2022.

Two defining features of early land plants were: (a) the ability to house multiple chloroplasts per cell (polyplastidy), and (b) the ability of these chloroplasts to divide independently of the cell (de Vries et al., 2016; de Vries & Gould, 2018; MacLeod et al., 2022). Indeed, this tightly linked regulation of plastid division in most embryophytes ensures that there is an appropriate response to biotic and abiotic factors, such as desiccation (Pudelski et al., 2012; Park et al., 2018). As such, it was essential for a hypothetical ancestral embryophyte to be polyplastidic, and to escape the monoplastidic bottleneck – a phenomenon that constrains some organisms from possessing multiple, and better synchronized plastids, per cell (de Vries & Gould, 2018) – to maintain the fitness of plant and plastid, alike (de Vries et al., 2016; MacLeod et al., 2022).

The polyplastidic nature of the ancestral embryophyte is supported by ancestral state reconstruction (ASR) analyses (MacLeod et al., 2022), suggesting that whatever PDVM configuration this organism used was responsible for the emergence of polyplastidy in the terrestrial clade. While ASRs should always be taken with a pinch of salt, they can be an incredibly useful tool to study the origins and evolution of certain traits (Joy et al., 2016; Holland et al., 2020). Previous analyses were able to overcome some of these uncertainties that accompany ASR analyses via appropriate model selection and obtaining high probability states for surveyed taxa (MacLeod et al., 2022). These analyses showed that, in contrast to the PDVM of the ancestral embryophyte, streptophyte algae appear to completely lack ARC3 orthologues (*Figure 4*). As such, it is suggested that the transition of plants to land was accompanied via the independent recruitment of ARC3 to the PDVM: a PDVM that was also

comprised of FtsZ1, FtsZ2, FtsZ3, ARC5, ARC6, PDV2, MinD, MinE and the PG layer (MacLeod et al., 2022). This recruitment of ARC3 likely allowed the ancestral embryophyte to escape the monoplastidic bottleneck (MacLeod et al., 2022). This hypothesis is supported by molecular genetic analyses undertaken in *A. thaliana* along with phylogenomic analyses undertaken in hornworts, a topic that will be discussed next.

Despite embryophyte emergence being accompanied by a burst of independent gene recruitments, the bryophyte group of embryophytes appears to be significantly reduced in terms of gene context relative to their vascular sisters (Harris et al., 2022; MacLeod et al., 2022). This phenomenon can be observed in PDVM components as well. For example, FtsZ1 orthologues appear to be completely absent in the moss genus *Sphagnum* (*Figure 4*). However, even within bryophytes, one group appears to be significantly more reduced that the others: the hornworts. While hornwort genera such as *Megaceros,* along with some *Anthoceros* and *Nothoceros* species, are polyplastidic (as they can possess up to four chloroplasts per cell), the ancestral hornwort was most likely monoplastidic (MacLeod et al., 2022). While gene losses allowed mosses and liverworts, and likely the ancestral bryophyte, to maintain a polyplastidic phenotype, it is very likely that the monoplastidic – and near-monoplastidic – nature of hornworts, including the ancestor of this group, is a result of this group losing two key PDVM components: ARC3 and FtsZ2. Indeed, molecular genetic and cell biology studies undertaken in *A. thaliana* and *P. patens* suggest that *ftsz2* and *arc3* mutants either result in complete monoplastidy (*ftsz2*) (Pyke and Leech, 1992; Martin et al., 2009) or very near monoplastidy (*arc3*) (Pyke & Leech, 1992), suggesting that it does not stand to reason that the combined loss of these two genes could have contributed to the monoplastidic nature of the hornworts (MacLeod et al., 2022).

While ARC3 is indeed absent in some polyplastidic species, for example in some ferns, mosses and the lycophyte *Selaginella moellendorffii* (*Figure 4*), FtsZ2 is also present in these species (MacLeod et al., 2022). Therefore, it is likely that: (a) FtsZ2 could compensate for ARC3's absence in some way, a claim supported by molecular genetic analyses, highlighted previously (Martin et al., 2009; Pyke & Leech, 1992), and (b) ARC3 was necessary to "kickstart" the transition to polyplatidy in the terrestrial clade, but that FtsZ2 acted to maintain it. Regardless, comparative genomic analyses do indeed suggest that the PDVM of bryophytes and seedless tracheophytes underwent significant modifications, the consequences and implications of which will be discussed next.

Components of the embryophyte PDVM underwent multiple modifications in bryophytes and ferns, involving a series of gains and losses of PDVM proteins, before becoming relatively homogenous in the spermatophytes (*Figure 4*). Examples of these modifications include the differential loss of ARC3 in the Bryopsida class of mosses, the fern *Azolla filiculoides* and the lycophyte *S. moellendorffii*; the loss of FtsZ1 and PDV2 in Sphagnopsida class of mosses, along with the former gene's loss in the fern *Salvinia cucullata*; and the loss of the PG layer in ferns

(*Figure 4*). What accounts for the distribution of diverse PD machineries in seedless plants is unknown. One likely explanation is that the PDVM of early embryophytes could have undergone a "trial and error" period to maintain a polyplastidic phenotype, the importance of which has been discussed in a previous section of this section (de Vries et al., 2016; de Vries & Gould, 2018). Indeed, during this period, some lineages reverted to the monoplastidic nature of their algal ancestors, thereby explaining the monoplastidy of hornworts and some lycophytes; also discussed previously (Frangedakis et al., 2020; Liu et al., 2020; MacLeod et al., 2022).

However, there is a strong correlation between the emergence of a largely homogenous PDVM in spermatophytes and the emergence of a cytological and physiological trait that defines this clade: multifaceted polyplastidy (*Figure 4*). Multifaceted polyplastidy is the ability for plastids to interconvert between different plastid types, of which there are at least six in spermatophytes: gerontoplasts, elaioplasts, etioplasts, chromoplasts, amyloplasts and chloroplasts (Choi et al., 2021; de Vries et al., 2016; Jarvis & López-Juez, 2013). These plastids, in turn, can be transported to various parts of the plant to fulfill their physiological roles (Jarvis & López-Juez, 2013).

It is unknown what the underlying molecular genetics of multifaceted polyplastidy are, although retrograde signaling is hypothesized to play a key role in plastid interconversion (Jarvis & López-Juez, 2013). However, regardless of what the causes of these major modifications in the PDVM and developmental pathways of spermatophytes were, the consequences of obtaining this homogenous PDVM in spermatophytes may have been as far reaching facilitating the origin of multifaceted polyplastidy in the terrestrial clade. Indeed, the emergence of spermatophytes – and as such multifaceted polyplastidy – was accompanied via the independent recruitment of two new components to the PDVM that are unique to this clade: PDV1 and PARC6 (MacLeod et al., 2022; Osteryoung & Pyke, 2014) (*Figure 4*). PDV1 is an integral membrane protein that is a homologue of PDV2; it works with the latter in establishing a successful outer division ring (Sun et al., 2020). PARC6 is a paralogue of the ARC6 component of the PDVM and is involved in the recruitment of PDV1 (Glynn et al., 2009), suggesting that these two components co-evolved (Osteryoung & Pyke, 2014): a hypothesis supported by comparative genomic analyses (MacLeod et al., 2022). Mutations in both PDV1 and PARC6 cause significant decreases in plastid copy number per cell (Glynn et al., 2009; Sun et al., 2020), but the roles that PDV1 and PARC6 play – or played – in establishing a multifaceted polyplastidic phenotype in spermatophytes have yet to be addressed.

There are several studies suggesting that the PG layer plays a key role in regulating chloroplast division in bryophytes and streptophyte algae (Machida et al., 2006; Hirano et al., 2016; Grosche and Rensing, 2017; Dowson et al., 2022). Mutations in key genes involved in the *P. patens* PG biosynthetic pathway either cause one giant chloroplast to occur per or few chloroplasts per relative to the wild-type moss lines *(Homi et al., 2009)*. Furthermore,

comparative genomic analyses suggest that the PG layer is a trait that is present in all three phyla of the Chloroplastida (Viridiplantae) (van Baren et al., 2016; Li et al., 2020b; Bachy et al., 2022; Dowson et al., 2022; MacLeod et al., 2023), but whether or not this cyanobacterial relic is a component of the PDVM in all species has yet to be determined.

The PG layer of embryophytes was hypothesized to be only present in bryophytes and lycophytes (Grosche & Rensing, 2017), along with the ancestral embryophyte (MacLeod et al., 2022) – with it being differentially lost in ferns and spermatophytes (Grosche & Rensing, 2017) – and was first characterized in the moss *P. patens* via click chemistry methods (Liechti et al., 2014). Early cytological research found that if the moss *Phsycomitrium patens* is treated with d-Cycloserine, this causes gigantic chloroplasts to occur in moss cells (Katayama et al., 2003). Because d-Cycloserine inhibits the formation of D-Ala:D-Ala (DA-DA) peptidoglycan polymers, via its inhibition of DDL ligases, subsequent molecular genetic experiments found that mutations of the *DDL* gene in *Physcomitrium* yielded moss plants with giant, undivided chloroplasts in their cells. However, this *ddl*-induced plastidic phenotype, along with DA-DA biosynthesis, in mosses could be rescued by "feeding" these plants with a solution known as ethynyl-DA-DA (EDA-DA). After extracting these newly-rescued cells and fixing them, click chemistry could be used to attach an azide-modified fluorophore to the alkyne groups of EDA-DA, lighting up the murein layer in the intermembrane space and confirming its presence at the division layer of chloroplasts (Hirano et al., 2016). Furthermore, latter proteomic analyses confirmed that there is functional conservation between the peptidoglycan biosynthetic pathway of mosses and Gram-negative bacteria. But while the peptidoglycan layer of mosses has been functionally characterised, the same cannot be said for other embryophytes that possess orthologues for a full murein biosynthetic toolkit. However, cytological studies found if the cells of ferns (*Adiantum capillus-veneris*, *Ceratopteris richardii*, and *Equisetum arvense*) and lycophytes (*Selaginella nipponica*) are treated with the peptidoglycan-targeting antibiotic Fosfomycin, this results in a reduction in plastid copy numbers in treated species (Izumi et al., 2008). Therefore, these cytological studies showed that this moss's murein layer is comprised of D-amino acids – the same components that make up this cytoplasmic envelop of bacteria (Hirano et al., 2016). Indeed, subsequent biochemical and metabolomic analyses suggest that components of the moss PG biosynthetic pathway – specifically, the active sites of core ligase enzymes – display strict conservation in comparison to the PG biosynthetic pathway of cyanobacteria (Dowson et al., 2022). However, these latter experiments also found that a PDVM component, FtsZ3, is unlikely to play a role in PG layer biosynthesis in moss. Indeed, previous studies have suggested that the plastid division protein FtsZ3 plays a role in regulating the biogenesis of chloroplast PG, due to an alleged correlation between these two traits (Grosche & Rensing, 2017); however recently published comparative genomic data, along with biochemical and metabolomic analyses, suggest that this is unlikely to be the case, as there are multiple exceptions to this correlation (van Baren et al., 2016; MacLeod et al., 2023) (*Figure 4*).

In summary, it is very likely that the recruitment of ARC3 in the ancestral embryophyte allowed early land plants to escape the monoplastidic bottleneck, and hence gain fitness advantages needed for their survival on a terrestrial habitat. Furthermore, it is likely that the series of modifications in the PDVM of basal-branching embryophytes was necessary to maintain a polyplastidic phenotype in the terrestrial clade. As such, the emergence of a largely homogenous PDVM in spermatophytes was likely necessary for the emergence of multifaceted polyplastidy, but whether the recruitment of PDV1 and PARC6 contributed to this unique physiological innovation observed in seed plants has yet to be determined. We also discuss phylogenomic evidence suggesting that the monoplastidic phenotype of hornworts is a direct cause of this group differentially losing FtsZ2 and ARC3. Whether or not monoplastidy in other embryophytes, namely the lycophytes, is also a result of a combined loss of these two PDVM components has yet to be determined, as the absence of genomic and transcriptomic data from lycopod species-of-interest prevent this outstanding question from being addressed. Furthermore, it will be interesting to see if the reintroduction of ARC3 and FtsZ2 to hornwort genomes, via transfection protocols (Frangedakis et al., 2021; Neubauer et al., 2022), could result in a polyplastidic phenotype. Furthermore, it is also unlikely that the PG layer and FtsZ3 coevolved. While FtsZ3 is present in many PG layer-possessing embryophytes, metabolomic, genomic and biochemical data suggest that: (a) FtsZ3 likely does not play a role in PG layer biosynthesis, and (b) the presence of a PG layer is not dependent on the presence of FtsZ3.

4.4 Protein sorting to the thylakoid membranes. The ability for organisms to photosynthesize is achieved by a unique set of interconnected compartments known as the thylakoid membrane system (Frain et al., 2016; Flori et al., 2017; Wood et al., 2018, 2019; Wietrzynski et al., 2020; Xu et al., 2021). As such, it is essential for plants and green algae to possess a robust and highly coordinated thylakoidal proteome. Said proteome is assembled via three main pathways: the chloroplast signal recognition particle (cpSRP), chloroplast secretory (cpSec) and the chloroplast twin-arginine translocation (cpTat) pathways (Xu et al., 2021). The core components of these three pathways were inherited from cyanobacteria (Frain et al., 2016).

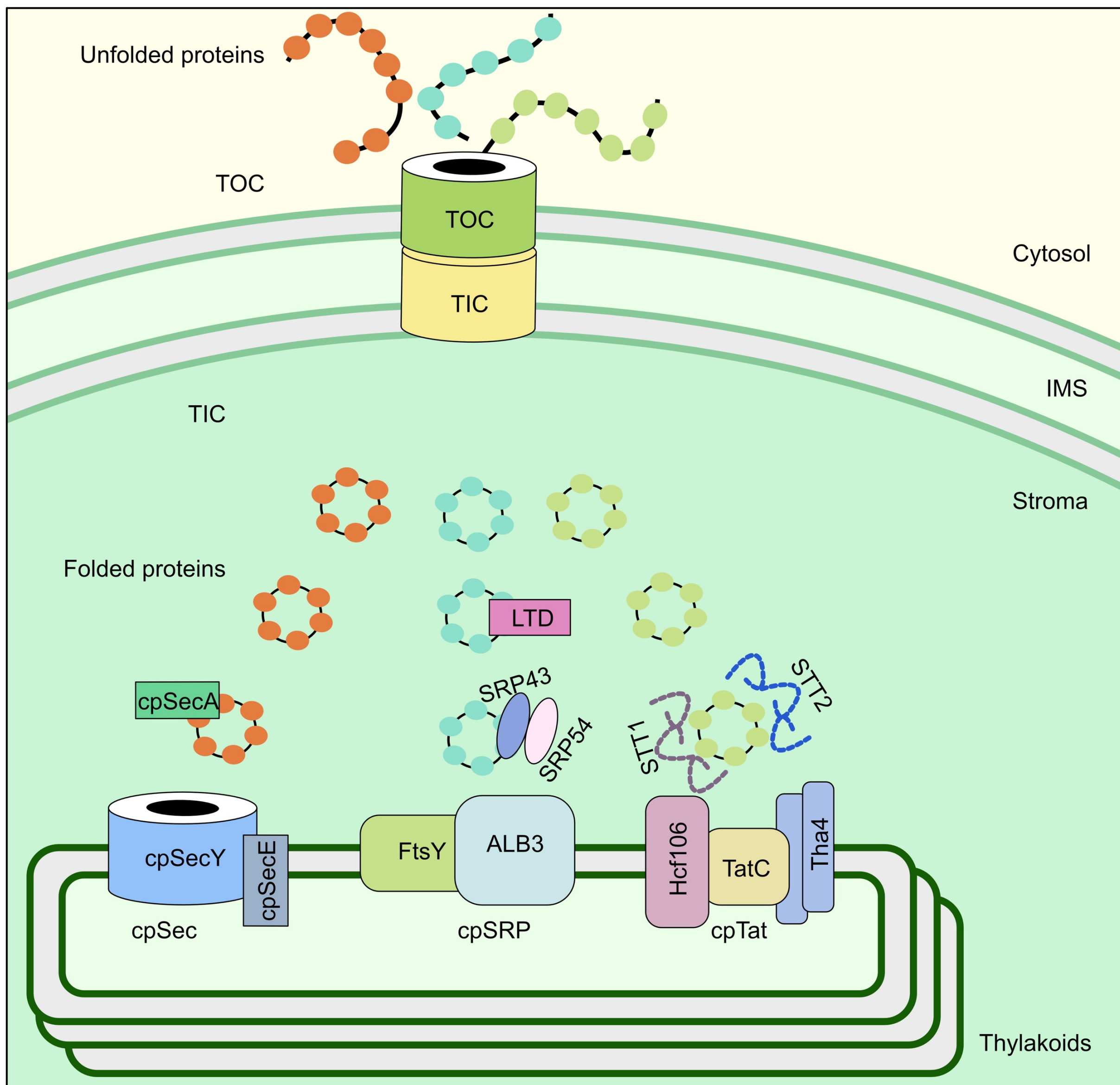

***Figure 5.* Overview of the three main pathways involved in protein sorting to the thylakoid in the thale cress *Arabidopsis thaliana*.** Figure adapted from MacLeod et al. 2022.

The cpSRP pathway in plastids is involved in sorting and targeting light harvesting complex proteins (LHCPs) to the thylakoid (Jarvis and López-Juez, 2013; Xu et al., 2021). LHCPs play key roles in the transfer of light energy during photosynthesis (Rochaix and Bassi, 2019; Xu et al., 2021). In the green lineage, after the rudimentary LHCP pre-protein is imported by a translocon of the outer/inner envelope of the chloroplast (TOC/TIC) super complex, it is accepted by the LTD protein, which transfers it to the cpSRP43/cpSRP54 complex. Afterwards, the LHCP is passed to the FtsY receptor, where subsequent GTP hydrolysis results in its integration to the thylakoid membrane via the ALB3 protein (*Figure 5*). In some embryophytes, a homologue of ALB3, ALB4, is the protein used to integrate some cpSRP substrates (for example cytochrome *f*) to the thylakoid membrane (Trösch et al., 2015).

The cpSec pathway is involved in importing unfolded proteins to the thylakoid membrane, and is comprised of three key components: cpSecA, cpSecE and cpSecY (Xu et al., 2021). After the unfolded preprotein is imported by the TOC/TIC transolocon, it interacts with the motor protein cpSecA, which chaperones it to a thylakoidal translocon channel comprised of cpSecY, and sometimes cpSecE; with the latter playing an accessory role in protein translocation, by tiling and rotating the N-terminal half of cpSecY, with subsequent passage of this preprotein leading to its integration into the thylakoidal proteome (Park et al., 2014) (*Figure 5*).

The cpTat pathway is involved in sorting *folded* proteins to the thylakoid membrane (Xu et al., 2021), and is regulated by the stromal proton motive force (PMF) (*Figure 5*). In embryophytes, once passed into the stroma via the TOC/TIC complex, the preprotein binds to a TatC-Hcf106 complex, leading to the subsequent recruitment of the Tha4 protein via the PMF. Tha4, in turn, undergoes a conformational change, leading to the passage of this preprotein into the thylakoid (MacLeod et al., 2023) (Xu et al., 2021). However, some embryophytes have evolved the unique ability for preproteins to undergo liquid-liquid phase transitions (LLPTs), where these biomolecules separate into membrane-less organelles (Ouyang et al., 2020), leading to more efficient sorting during the Tat pathway (*Figure 5*). These LLPTs are regulated by the STT1 and STT2 proteins.

4.5 The TOC/TIC import machinery. The TOC/TIC import machinery is a complex molecular structure that allows for the import of the chloroplast – and plastid – proteome (Richardson and Schnell, 2020). This macromolecular complex is comprised of two key protein translocation complexes: the TOC (translocon at the outer envelope membrane of the chloroplasts) complex and the TIC (translocon at the inner envelope membrane of chloroplasts) complex (*Figure 5*).

The TOC complex is comprised of the following subunits: TOC75, TOC34 and TOC159. Unfolded pre-proteins from the cytosol enter the TOC75 channel and are eventually received by the TIC complex in the intermembrane space (IMS) of the chloroplast. Recycling of the TOC complex is regulated via the E3 ubiquitin ligase SP1, which targets this complex for proteasomal degradation (Richardson and Schnell, 2020). The TIC complex is comprised of a massive 1 MDa complex comprised of the following components: TIC20, TIC21, TIC56, TIC100, TIC236, TIC110, TIC40, TIC22 and TIC214 (or YCF1). Once the pre-proteins pass from the IMS into TIC and then the stroma, they are folded and targeted to their respective locations within the chloroplast. (Xu et al., 2021).

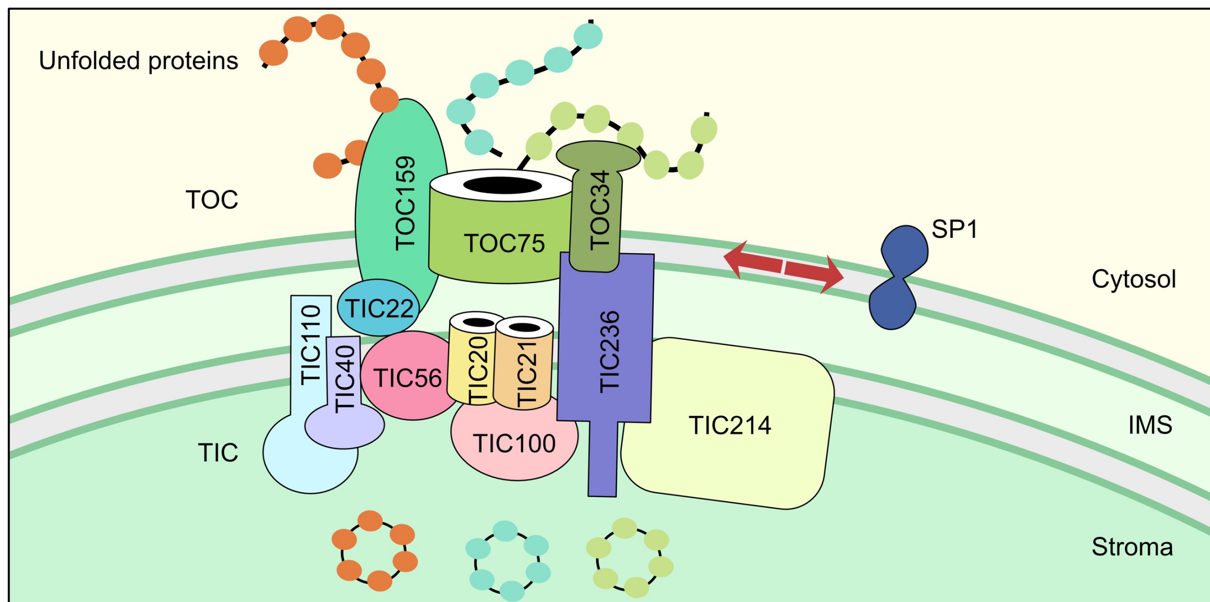

Figure 6. Overview of the TOC/TIC protein import machinery. Figure adapted from MacLeod et al. 2022.

The TOC/TIC import machinery is a complex system that allows for the import of proteins into the chloroplasts of plants and green algae. The system consists of two main protein translocation complexes, the TOC (translocon at the outer envelope membrane of chloroplasts) complex and the TIC (translocon at the inner envelope membrane of chloroplasts) complex, each with multiple subunits responsible for different aspects of the import process. The Toc complex consists of the following subunits: TOC34, TOC75, TOC159 and the SP1 ubiquitin ligase. TOC75 is involved in forming a channel in the outer membrane of the chloroplast that allows for the translocation of the precursor protein into the intermembrane space. TOC34 and TOC159, on the other hand, form a heterodimer that functions in the recognition and binding of precursor proteins in the cytosol for import into the chloroplast. The recycling of these TOC components – via proteasomal degradation – is regulated via the SP1 E3 ubiquitin ligase (Richardson and Schnell, 2020). The TIC complex, on the other hand, consists of the following subunits: TIC20, TIC22, TIC40, TIC110, TIC55, TIC62 and TIC214. These components are involved in the formation of a 1MDa multimer that receives the precursor proteins from the TOC channel in the intermembrane space (IMS). After being received by the TIC complex, these preproteins can pass into the stroma and be targeted to their respective plastidic compartments (Richardson and Schnell, 2020).

The evolution of the TOC/TIC import machinery in plants can be traced back to ancestral chloroplastid (Knopp et al., 2020). Indeed, cyanobacteria, which are the ancestors of chloroplasts, have a simpler import machinery that includes the ancestor of the TOC75 protein, OEMP85 (Knopp et al., 2020). But with the emergence of multicellular plants, the complexity of the TOC/TIC import machinery increased. Furthermore, the TOC/TIC import machinery in plants has co-evolved with precursor proteins to ensure efficient and accurate

import. For example, the N-terminal targeting sequences of precursor proteins have evolved to contain specific amino acid motifs that are recognized by the TOC receptors (Garg and Gould, 2016; Knopp et al., 2020; Richardson and Schnell, 2020).

4.6 Hornworts. A central theme of this thesis is a group of plants known as the hornworts. Hornworts are a group of bryophytes (non-vascular plants) that are commonly found in damp, shaded habitats around the world. They are a very ancient group of plants, making them an attractive group of plants for plant evolutionary developmental biology (evo-devo) studies (Li et al., 2017; Frangedakis et al., 2020; MacLeod et al., 2022). As will be discussed, hornworts have a distinct morphology that sets them apart from other embryophytes and have an interesting evolutionary history (*Figure 7*).

Belonging to the phylum Anthocerotophyta, hornworts are believed to have diverged from other embryophytes between 400-900 million years ago (Morris et al., 2018; Su et al., 2021). Hornworts have remained relatively unchanged for the past 200 million years, which make them one of the oldest plant lineages still in existence (Villarreal and Renner, 2012). Despite their ancient origins, hornworts occupy a wide range of habitats – from most tropical forest floors to arid deserts (Frangedakis et al., 2020).

However, hornworts are primarily found in damp, shaded habitats such as stream banks, swamps, and forest floors. They are often found growing in dense mats or carpets, which help to conserve moisture and provide a protective environment for other small plants and animals. Hornworts are known to form symbiotic relationships with cyanobacteria, which live within the plant's thallus and help to fix nitrogen from the atmosphere (Frangedakis et al., 2020, 2023). This allows hornworts to thrive in nutrient-poor soils, and they are often found in areas where other plants struggle to grow (Rahmatpour et al., 2021).

Hornworts have a distinct morphology that sets them apart from other plant groups. They are also the only species of land plants that have failed to escape the monoplastidic bottleneck; a phenomenon that prevents some photosynthetic organisms from possessing multiple plastids per cell (de Vries and Gould, 2018; Frangedakis et al., 2020, 2023; MacLeod et al., 2022). In addition, hornworts are the only group of embryophytes that possess a pyrenoid structure, a unique carbon concentrating mechanism (CCM) that is commonly found in algae (Villarreal and Renner, 2012; Frangedakis et al., 2020).

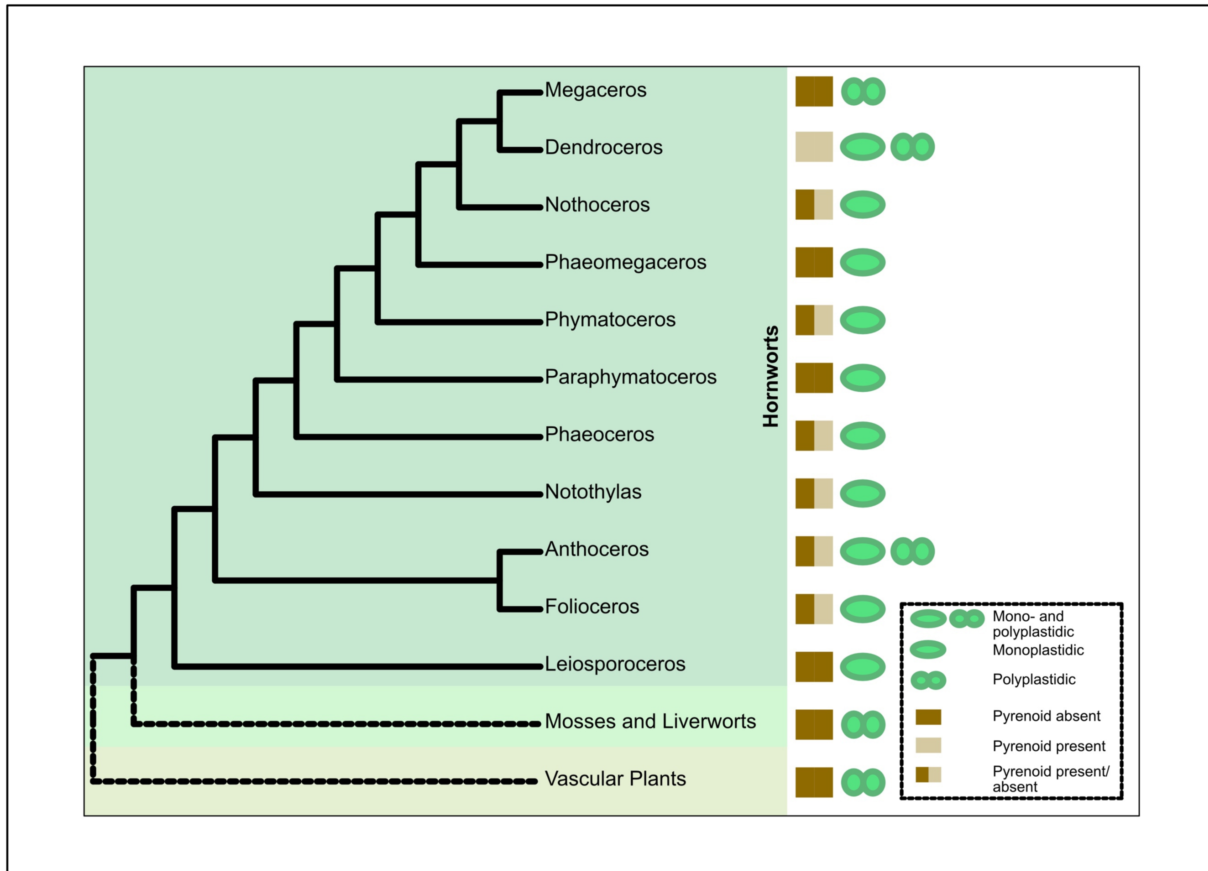

***Figure 7*. The evolution and diversity of hornworts, with an emphasis on cellular morphology.** Figure adapted from MacLeod et al. 2022.

4.7 Exploring the role of evolutionary developmental biology (Evo-Devo) in plant adaptation to land, and how methods in evo-devo were used to address the main aims of this thesis. Evolutionary developmental biology (evo-devo) is a discipline of biology that studies: (a) if certain developmental processes are conserved between phylogenetically distant taxa, (b) if yes, to what extent, and (c) if these developmental processes contributed to the evolution of organisms over time (Carroll, 2005; Delaux et al., 2019). Evo-devo is a very important research field because it provides deep molecular and cellular insights into how certain developmental processes have been modified and adapted over evolutionary time to produce a plethora of traits observed in nature today: from multicellularity to the origin of plant body plans (Delaux et al., 2019).

As mentioned previously in this thesis, the emergence of embryophytes from streptophyte algal ancestors was a crucial event in the history of life of Earth. And because had to overcome a wide range of terrestrial stressors – from desiccation, ultraviolet (UV) radiation and nutrient acquisition from the soil to name a few – studying the molecular and cellular bases by which plants occupied land is very important for the following two reasons (Schreiber et al., 2022):

1. **Preserving plant biodiversity:** Understanding the underlying molecular biology of plant adaptations to new habitats and niches can help further understanding of how plant species respond to environmental change in an a rapid and everchanging climate. This knowledge would be crucial for preserving plant biodiversity in the face of the climate crisis (Rensing, 2018; Schreiber et al., 2022).

2. **Improving crop yields:** Most molecular and developmental innovations that allowed plants to conquer land, such as the development of a complex root system and drought tolerance mechanisms, are incredibly valuable to agriculture and the field of plant biotechnology. Therefore, by delineating the molecular genetic underpinnings of these innovations that plants evolved to dominate terrestrial habitats, biotechnologists can potentially develop new strategies aimed at making agriculture more sustainable and efficient (Fricke et al., 2022).

A typical pipeline for undertaking an evo-devo study involves two steps. The primary step involves using *in silico* methods for characterizing orthologues for developmental genes-of-interest in a target species, group or phylum. Orthology is defined as a gene – or a family of genes – that: (a) evolved from a common ancestor via speciation, and (b) *likely* perform similar or identical biological functions (Altenhoff and Dessimoz, 2009). If orthologues for said genes-of-interest are present, the secondary step involves incorporating methods in molecular and cellular biology to determine *if* this orthology infers developmental and/or biochemical function (Carroll, 2005; Delaux et al., 2019). While this latter step relies of a variety of techniques in biochemistry, molecular and cell biology, the former step relies mainly on three techniques: comparative genomics, phylogenetics and ancestral state reconstructions (ASRs).

Comparative genomics incorporates methods in genome mining and hierarchical clustering to determine which genes are conserved between species, and what genes aren't (Edwards and Keller, 2004; Xia, 2011). Comparative genomic algorithms usually organize orthologous genes into "clusters" or "orthogroups" based on their similarity (Emms and Kelly, 2019). An example of one such program and algorithm used to determine a set of orthogroups between species – and the one used in studies discussed in this thesis – is OrthoFinder (Emms and Kelly, 2015, 2019). OrthoFinder is widely used in comparative genomic studies because it is very user-friendly and accurate for determining orthogroups. OrthoFinder uses a sequence-similarity approach to: (a) firstly determine homology between a set of protein sequences, and (b) finally use a graph-based approach to cluster these shared gene families into orthogroups (Emms and Kelly, 2015, 2019).

However, to determine the evolutionary origin of said clusters and orthogroups, it is often useful to construct phylogenetic trees for genes-of-interest. Phylogenetic trees are constructed by aligning sequences-of-interest to remove non-conserved aspects of genes, and then fed into a tree-building algorithm to infer the genes' evolutionary history (Thomas, 2001). A common way of generating a phylogenetic tree – and the one used in results discussed in this thesis – is the maximum-likelihood (ML) method. The likelihood function in ML analysis is a mathematical model that calculates the probability of observing the sequence data given a particular tree topology, branch lengths, and substitution model. The likelihood function estimates the likelihood of a given sequence alignment, given a particular tree topology, and is calculated using a combination of the nucleotide or amino acid substitution model and the branch lengths. The optimization of the likelihood function is performed using algorithms to identify the tree with the highest likelihood of producing the observed data. The generated tree represents the most probable evolutionary history of the organisms under study based on the available molecular data. ML trees are widely used in various fields of biology, including evolutionary biology, molecular ecology, and biogeography (Truszkowski and Goldman, 2016). The software package used to generate phylogenetic trees discussed in this thesis is IQ-TREE, version 2.0.3 (Minh et al., 2020), that employs several heuristics, including nearest-neighbour interchanges (NNI), subtree pruning and regrafting (SPR), and tree bisection and reconnection (TBR), to search for the optimal phylogenetic tree, with subsequent tree visualization carried out using FigTree and the R package ggtree (Yu et al., 2017; Rambaut and Drummond, 2018).

Ancestral state reconstruction, on the other hand, is a technique used in evolutionary biology to estimate the most likely states of a trait in ancestral organisms based on the states observed in their descendants (Joy et al., 2016; Holland et al., 2020). This technique can be applied to both discrete and continuous traits, although the methods used to calculate the ancestral states differ depending on the trait type. For discrete traits, such as presence or absence of a particular morphological feature, maximum likelihood methods are commonly used: the ancestral state with the highest probability is then considered the most likely state. For continuous traits, such as body size or metabolic rate, ancestral state reconstruction typically involves fitting a model of trait evolution to the observed data, and then using that model to estimate the most likely trait values for ancestral organisms. Maximum likelihood methods can also be used for continuous traits, with the likelihood function based on a model of trait evolution that accounts for factors such as natural selection and genetic drift (Joy et al., 2016). A variety of packages used for inferring ancestral state reconstructions exist. One such package – and the package used in studies described in this thesis is the Phytools package from the R programming language (Revell, 2012). The Phytools package provides a set of functions for ancestral state reconstruction, allowing users to estimate the most probable states of ancestral nodes of a phylogenetic tree based on the character states observed in the extant species. The package offers several methods for ancestral state reconstruction, including maximum likelihood, parsimony, and Bayesian approaches, for both discrete and continuous traits. The likelihood approach in Phytools is based on a Markov model, which assumes that the evolution of traits follows a stochastic process where the probability of a

character state at a given node depends on the states of its ancestor and the transition probabilities between states. The package also provides tools for visualizing the results of ancestral state reconstruction, including plotting the probabilities of different ancestral states across the tree (Revell, 2012). It is important to note that while ASRs can be very useful, they should be taken with a massive pinch of salt, as: (a) they are a statistical model, not a reality, and (b) it is virtually impossible to know for certain how a hypothetical and extinct ancestor of a given phylogenetic group looked. As such, to increase the accuracy for an ASR estimation, it is essential to have a robust species phylogeny (Holland et al., 2020).

The next stage of the evo-devo pipeline is to incorporate methods in molecular biology, cell biology and biochemistry to determine the functional significance of these genes, how the effect the physiology of a biological system, and if there is functional conservation between said system and other closely- or distantly-related species (Delaux et al., 2019).

5 Aims of this thesis

The purpose of this thesis is to discuss results from two publications that investigated plastid evolution in the Chloroplastida. These publications investigated the phylogenetic distribution of plastid developmental genes in the green lineage to determine whether key differences correlate with major changes in the evolutionary history of algae and plants.

Experiments discussed in Publication I deconstructed the plastid biology of 10 hornwort species to highlight that the emergence of this group of bryophytes was accompanied with the differential loss of two plastid division components: ARC3 and FtsZ2. Mutations in both ARC3 and FtsZ2 cause monoplastidic phenotypes to occur in *Arabidopsis thaliana* and *Physcomitrium patens*, with ancestral state reconstruction analyses suggesting that the differential loss of these two genes likely explains the monoplastidic nature of hornworts.

Publication II described experiments that investigated the phylogenetic distribution of genes involved in chloroplast peptidoglycan layer biosynthesis in 48 members of the Chloroplastida. Results from this publication indicate that this cyanobacterial relic is present in all three phyla of the green lineage but shows a very unusual and uneven distribution in this supergroup. Furthermore, there is a strict structural conservation in the enzymes involved in the murein layer biosynthetic machinery, from cyanobacteria to angiosperms, indicating that the function of these enzymes has also been maintained over evolutionary time.

Experiments described in Publication III examined organelle proteome data and genomes of both phototrophic and non-phototrophic eukaryotes, spanning this evolutionary timeframe. Of 331,571 protein families studied, 31,625 were unique to organisms with primary plastids. A deeper dive showed 1,961 protein families in plastids and 846 in mitochondria. These results reveal the extensive changes these endosymbiotic organelles underwent, from their inception in algae to their presence in land plants, highlighting crucial molecular adaptations that aided in plant diversification and the rise of angiosperms.

6 Summary of obtained results

Publication I

This paper combines methods in genome mining, gene clustering, and ancestral state reconstruction to investigate the plastid biology of hornworts. We initially found that hornworts are significantly reduced in terms of gene content, more so than other plants (including bryophytes). To speculate on whether monoplastidy in hornworts is a result of differential loss of certain genes, we decided to zoom into specific candidate genes-of-interest to identify any interesting patterns in the presence/absence of genes regulating plastid development: specifically, those involved in regulating protein import, thylakoid biogenesis, and chloroplast division. This is because this developmental toolkit is involved in determining plastid fate and identity in embryophytes. And while we didn't identify any significant patterns in the thylakoid biogenesis and import clusters, we found interesting correlations in the plastid division cluster: hornworts differentially lost the ARC3 and FtsZ2 plastid division proteins. This differential loss of ARC3 and FtsZ2 correlates with hornworts reverting to a monoplastidic phenotype. Indeed, previous studies have shown that generating mutant lines of ARC3 and FtsZ2 in the cress *Arabidopsis thaliana* and the moss *Physcomitrium patens* cause fewer plastids (in the case of *arc3* mutants) or one giant plastid per cell (in the case of *ftsz2* mutants). Furthermore, ancestral state reconstruction analyses demonstrate the ancestral embryophyte possessed both ARC3 and FtsZ2 and was polyplastidic; the opposite of which is true for the ancestral hornwort, highlighting that hornwort emergence was accompanied by the loss of ARC3 and FtsZ2.

Publication II

This paper combines methods in phylogenomics, protein domain analyses and gene clustering to investigate the phylogenetic distribution and molecular evolution of the peptidoglycan (PG) layer in the Chloroplastida lineage. We found that the distribution of a full PG enzymatic toolkit is concentrated and noticeable in streptophyte algae, bryophytes and the lycophytes. However, we also found that this toolkit is present in three chlorophyte algal species – reporting for this first time its presence in *Ulva mutabilis* and *Chloropicon primus*. In addition, we also find that this cyanobacterial relic is present in at least three seed plants: *Asparagus officinalis*, *Thuja plicata* and *Citrus sinensis*. Furthermore, there is a strong structural conservation in the PG layer biosynthetic toolkit from algae to angiosperms.

We also investigated the correlation between the PG layer and FtsZ plastid division proteins. Previous studies suggested that the FtsZ3 protein is involved in regulating PG layer biosynthesis, due to an alleged correlation between the two components – indicating that these two components likely coevolved. However, we found that this is unlikely to be the case. *T. plicata*, *A. officinalis*, *C. sinensis*, *U. mutabilis* and *C. primus* possess orthologues representing a full enzymatic toolkit for PG layer biosynthesis but lack FtsZ3. These results suggest that presence of the PG layer is not dependent on the presence of FtsZ3 (or any FtsZ protein involved in plastid division). As such, our results suggest that FtsZ3 may not have a direct role in the biogenesis of the PG layer, and other genes and networks may contribute to the formation and regulation of this cyanobacterial relic.

Publication III
This paper discusses methods in comparative genomics to explore the evolution and diversity of plant and algal biochemical and cellular pathways across a billion years of plant and algal evolution. Our analyses, encompassing 168 Archaeplastida species, unveiled 31,625 protein families unique to primary plastid-bearing eukaryotes. Of these, 1961 and 846 protein families were predicted to function within plastids and mitochondria, respectively. Tracing the evolutionary trajectory of these protein families, we discerned significant organelle remodelling from algae to land plants. The emergence of key adaptations, such as wax and cutin synthesis, RNA editing enhancements, and the transition from mono- to polyplastidy, were particularly noteworthy. These molecular adaptations in organelles were pivotal in facilitating major evolutionary transitions, especially the terrestrialization of plants.

Our findings further underscored the intricate interplay between plastids and mitochondria, emphasizing their collaborative role in the evolutionary success of the Chloroplastida. The evolution of components specific to chloroplast and mitochondrial biology, especially in the context of streptophyte evolution and terrestrialization, emerged as a focal point. Indeed, the significant changes observed at two evolutionary timepoints – namely the green lineage's inception and the water-to-land transition during the split between Zygnematophyceae and embryophytes – were particularly enlightening. The newly recruited protein families influenced a plethora of organellar processes, from carbon and lipid metabolism to information processing and organelle development. In summary, our study offers a rich tapestry of insights into the molecular changes experienced by plastids and mitochondria throughout streptophyte evolution, shedding light on major evolutionary transitions and the intricate dance of organelles in the conquest of land.

7 Conclusions

The conclusions drawn from the two publications suggest interesting avenues for future research in the field of plastid evolution. Our results from Publication I suggest that a consequence of some plastid-related gene losses that accompanied hornwort emergence, most notably the combined loss of FtsZ2 and ARC3, resulted in hornworts reverting to a monoplastidic phenotype, which the embryophyte ancestor was able to escape. We speculate if the knockout of ARC3 and FtsZ2 in *Arabidopsis thaliana* and moss results in monoplastidic phenotypes, one could potentially "reverse" evolution by expressing ARC3 and FtsZ2 in hornworts.

Our results from Publication II indicate that the peptidoglycan (PG) layer is more widespread in the green lineage than previously thought, highlighting its presence in three seed plants and three chlorophyte algae. Our results also suggest that penicillin-binding proteins (PBPs) are present in the seed clade; a protein family once thought to be absent from this group of embryophytes. However, we conclude the study by stating that more research is needed in the green lineage to elucidate (a) the biochemical nature of the PG layer, (b) its biological significance, and (c) how/if PG layer biosynthesis and plastid division are interlinked.

Our findings from Publication III underscore the profound evolutionary adaptations of plastids and mitochondria across the green lineage, spanning over a billion years. These molecular adaptations, especially during terrestrialization, were pivotal for the success of streptophytes in diverse habitats. However, we conclude by emphasizing the need for further research to (a) delve deeper into the functional roles of newly identified protein families, (b) understand the intricate crosstalk between plastids and mitochondria, and (c) elucidate how these organelle adaptations influenced macroevolutionary trends leading to the global greening and rise of angiosperms.

8 Results associated with this thesis

8.1 Publication I: Loss of plastid developmental genes coincides with a reversion to monoplastidy in Hornworts

Authors: **Alexander I. MacLeod**, Parth K. Raval, Simon Stockhorst, Michael R. Knopp, Eftychios Frangedakis and Sven B. Gould

This article was published in the journal *Frontiers in Plant Science* (in the section Plant Systematics and Evolution) on the 14th of March 2022

Contribution of Alexander István MacLeod:

MAJOR:

- Designed and undertook experiments
- Drafted the manuscript
- Created the outline for *Figure 1* and supplementary figure S1
- Created and edited *Figure 2* and supplementary figures S2-S4
- Joint-corresponding author

All data associated with this manuscript (including supplementary information) can be accessed via the journal's website:
https://www.frontiersin.org/articles/10.3389/fpls.2022.863076/full

The article is reprinted here from the Journal of Frontiers in Plant Science under the Creative Commons CC BY license.

frontiers in Plant Science

PERSPECTIVE
published: 14 March 2022
doi: 10.3389/fpls.2022.863076

Loss of Plastid Developmental Genes Coincides With a Reversion to Monoplastidy in Hornworts

Alexander I. MacLeod[1], Parth K. Raval[1], Simon Stockhorst[1], Michael R. Knopp[1], Eftychios Frangedakis[2] and Sven B. Gould[1]**

[1] Institute for Molecular Evolution, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany, [2] Department of Plant Sciences, University of Cambridge, Cambridge, United Kingdom

The first plastid evolved from an endosymbiotic cyanobacterium in the common ancestor of the Archaeplastida. The transformative steps from cyanobacterium to organelle included the transfer of control over developmental processes, a necessity for the host to orchestrate, for example, the fission of the organelle. The plastids of almost all embryophytes divide independently from nuclear division, leading to cells housing multiple plastids. Hornworts, however, are monoplastidic (or near-monoplastidic), and their photosynthetic organelles are a curious exception among embryophytes for reasons such as the occasional presence of pyrenoids. In this study, we screened genomic and transcriptomic data of eleven hornworts for components of plastid developmental pathways. We found intriguing differences among hornworts and specifically highlight that pathway components involved in regulating plastid development and biogenesis were differentially lost in this group of bryophytes. Our results also confirmed that hornworts underwent significant instances of gene loss, underpinning that the gene content of this group is significantly lower than other bryophytes and tracheophytes. In combination with ancestral state reconstruction, our data suggest that hornworts have reverted back to a monoplastidic phenotype due to the combined loss of two plastid division-associated genes, namely, ARC3 and FtsZ2.

Keywords: plastid evolution, bryophytes, hornworts, plant terrestrialization, plastid division

OPEN ACCESS

Edited by:
Michael Eric Schranz, Wageningen University & Research, Netherlands

Reviewed by:
Yoshihisa Hirakawa, University of Tsukuba, Japan

****Correspondence:***
Alexander I. MacLeod
macleod@hhu.de
Sven B. Gould
gould@hhu.de

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *26 January 2022*
Accepted: *04 February 2022*
Published: *14 March 2022*

Citation:
MacLeod AI, Raval PK, Stockhorst S, Knopp MR, Frangedakis E and Gould SB (2022) Loss of Plastid Developmental Genes Coincides With a Reversion to Monoplastidy in Hornworts. Front. Plant Sci. 13:863076. doi: 10.3389/fpls.2022.863076

INTRODUCTION

Hornworts are a unique group of bryophytes, the monophyletic non-vascular sister lineage to all vascular land plants (Harris et al., 2020). The phylogenetic position of hornworts and their putative phenotypic resemblance to what one might consider to represent the last common ancestor of all land plants make them an attractive model for evo-devo studies linked to events such as plant terrestrialization (Frangedakis et al., 2020). Hornworts are the only group of land plants known to form a pyrenoid, a unique carbon-concentrating mechanism (CCM), otherwise common in algae; however, these CCMs are not present in all hornworts and are hence a poor taxonomic marker (Villarreal and Renner, 2012; **Supplementary Figure 1**).

Hornworts are one of the only groups of embryophytes that have not escaped the monoplastidic bottleneck. This is a phenomenon associated with plastid origin and the organelle's integration into the host cell cycle, which constrains the majority of algae from possessing multiple plastids per cell (de Vries and Gould, 2018). One consequence is that the only plastids—of which there are five types in embryophytes (Jarvis and López-Juez, 2013)—hornwort cells house are chloroplasts,

whose size and morphology vary across genera (Vaughn et al., 1992; Raven and Edwards, 2014; Li et al., 2017). To address the reason, we screened the genomes and annotated transcriptomes of ten hornwort species to identify the presence/absence of genes that play key roles in regulating plastid development, such as those involved in protein import into the chloroplast, thylakoid biogenesis, and chloroplast division (Jarvis and López-Juez, 2013). We highlight key differences between the developmental plastid biology of hornworts and other established model organisms in the terrestrial clade. Furthermore, we argue that the major changes in plastid biology, that not only coincided with major checkpoints in the evolutionary history of hornworts but also facilitated them, are a consequence of multiple instances of gene loss observed in this unique group of embryophytes.

HORNWORTS UNDERWENT SIGNIFICANT INSTANCES OF GENE LOSS

We used the BUSCO version 5.2.2 (Manni et al., 2021) software to estimate the gene content of hornworts and compared them with other bryophyte and tracheophyte (vascular plant) outgroups (**Supplementary Table 1**). We found that the gene content of hornworts is significantly lower than tracheophytes and other bryophytes (ANOVA; $F = 129.5$; d.f. = 2,30; $p < 0.001$), thereby suggesting that hornwort diversification and speciation were accompanied by significant instances of gene loss (**Supplementary Figure 2**), even more than what is observed for bryophytes in general (Harris et al., 2021).

FULL CONSERVATION OF TRANSLOCON OF THE OUTER ENVELOPE OF THE CHLOROPLAST BUT ONLY PARTIAL CONSERVATION OF TRANSLOCON OF THE INNER ENVELOPE OF THE CHLOROPLAST IN HORNWORT CHLOROPLASTS

The vast majority of plastid proteins are encoded by the nuclear genome, and after their synthesis in the cytosol, are imported into the plastid by the translocon of the outer/inner envelope of the chloroplast (TOC/TIC) complex (Richardson and Schnell, 2020). Embryophytes have evolved the most sophisticated TOC/TIC complexes (Gould et al., 2008; Knopp et al., 2020) and our data confirm that the hornwort TOC complex is comprised of the same key proteins that are found in other embryophytes, mainly TOC75, TOC34, and TOC159 (Richardson and Schnell, 2020; **Figure 1**). The recycling of major TOC components is regulated by the RING-type ubiquitin E3 ligase SP1, which targets these proteins for proteasomal degradation (Ling et al., 2012; **Figure 1B**).

The TIC complex of embryophytes is comprised of a 1 MDa multimer that forms a pore that receives precursor proteins from the TOC complex in the intermembrane space (IMS) and finally mediates their passage to the stroma (Nakai, 2015a; Richardson and Schnell, 2020; **Figure 1**). The presence/absence of TOC/TIC components reveals no pattern with regard to mono-/polyplastidy or presence/absence of a pyrenoid (**Figure 1A** and **Supplementary Figure 1**). However, some TIC components appear to have undergone differential loss in some hornwort taxa (**Figure 1A**), most notably TIC21, TIC22, YCF1 (TIC214), and maybe even TIC20 in *Leiosporoceros dussii*. The latter species is the only member of our surveyed taxa that lacks a TIC20 ortholog (**Figure 1A**).

YCF1/TIC214, the only TOC/TIC component encoded by the plastid genome and unique to the green lineage, is absent in a significant number of hornworts (**Figure 1A**), such as in *Nothoceros aenigmaticus,* for which also the plastid genome is available (Villarreal et al., 2013).

DIFFERENTIAL LOSS OF AN ANCIENT THYLAKOID DEVELOPMENTAL PATHWAY IN MOST HORNWORTS

Thylakoid proteomes contain the bulk of photosynthesis-related proteins of plant cells (Xu et al., 2021). After their import *via* TOC/TIC, thylakoid proteins are recognized and sorted *via* one of three main pathways, the components of which are predominantly derived from the cyanobacterial endosymbiont or inserted spontaneously (Xu et al., 2021; **Figure 1**).

The chloroplast secretory (cpSec) pathway is involved in importing unfolded proteins to the thylakoid lumen. Powered by the motor protein cpSecA, unfolded subunits pass through a pore formed by cpSecY and cpSecE (Xu et al., 2021; **Figure 1B**). While the presence/absence of cpSec components reveals no pattern with regard to mono-/polyplastidy or presence/absence of a pyrenoid, half of surveyed hornworts lack cpSecE orthologs, with this distribution not showing any unique phylogenetic pattern (**Figure 1A** and **Supplementary Figure 1**).

The chloroplast twin-arginine translocation (cpTat) pathway can import folded proteins and is powered by the thylakoid's proton motive force (PMF; Xu et al., 2021). In those hornworts, for which we identified the cpTat pathway, it is comprised of three proteins, namely, Tha4, TatC, and Hcf106 (**Figure 1**). Precursor proteins initially bind to a TatC-Hcf106 complex. Tha4 is subsequently recruited *via* the action of the PMF, undergoing a conformational change, leading to the passage of the precursor protein (Xu et al., 2021; **Figure 1B**). The presence/absence of cpTat components reveals no pattern with regard to mono-/polyplastidy or presence/absence of a pyrenoid; however, the cpTat pathway seems only to be encoded by the Anthocerotaceae, having been differentially lost in other hornwort families (**Figure 1A**).

The third main pathway involved in sorting proteins for thylakoid biogenesis is the chloroplast signal recognition particle (cpSRP) pathway. This translocation complex is involved in targeting specifically light harvesting complex proteins (LHCPs) to the thylakoid membrane (Xu et al., 2021; **Figure 1B**). LHCP integration is initiated when a rudimentary LHCP is transferred from the TIC translocon to the SRP43/SRP54 complex by the

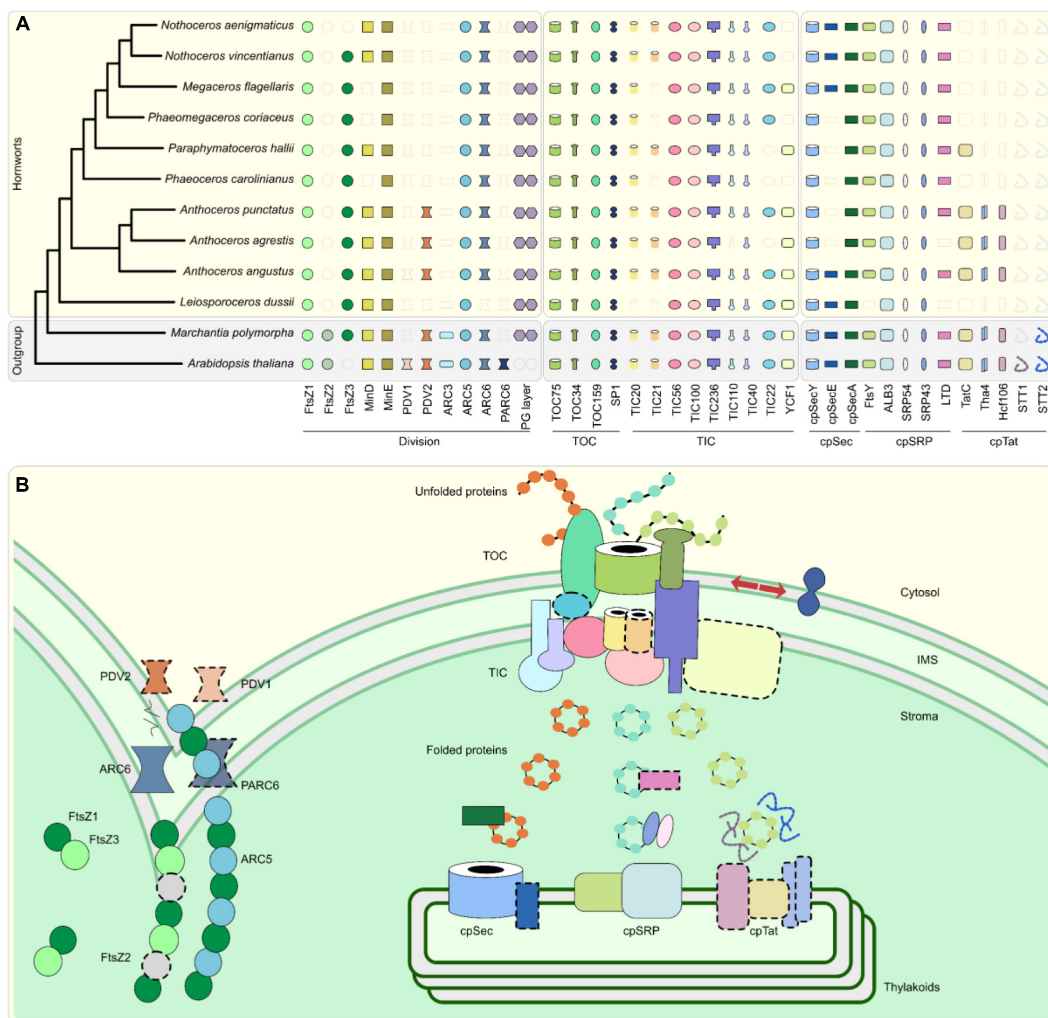

FIGURE 1 | Plastid development and biogenesis in hornworts. **(A)** A presence/absence pattern (PAP) of various plastid developmental components that are sorted into three categories based on whether they are associated with plastid division (PD) and protein translocation across the plastid envelope *via* TOC/TIC or the thylakoid membrane. Transparent icons indicate that no gene could be identified. **(B)** A combined schematic representation of plastid development in embryophytes. Components that are absent from more than two hornworts in our surveyed taxa, or absent in this group altogether, are highlighted by dotted outlines. ARC, accumulation and regulation of chloroplasts; FtsZ, filamentous temperature Z; IMS, intermembrane space; Sec, secretory; SRP, signal recognition particle; Tat; twin arginine translocation; TOC/TIC, translocator of the outer/inner chloroplast membrane; PDV, plastid division. While ARC5 is absent from the *Anthoceros agrestis* Bonn ecotype, which we included in our OrthoFinder analyses as the representative for this species, our reciprocal best hit pipeline confirmed that it is present in the Oxford ecotype, with its gene ID being AagrOXF_evm.TU.utg000081l.174. A maximum likelihood (ML) tree was constructed *via* the IQ-TREE version 2.0.3 software (Minh et al., 2020), using an automated selection model, by concatenating single-copy chloroplast and mitochondrial markers from 65 different hornwort species, and three outgroups (Villarreal and Renner, 2012). Said sequences were aligned with MUSCLE in AliView (Edgar, 2004; Laarson, 2014). Gene trees for orthologs listed on the PAP were generated using the PhyML version 3.0 and IQ-TREE version 2.0.3 softwares using automated selection models (Guindon et al., 2010; Lefort et al., 2017). We used the SHOOT framework (Emms and Kelly, 2021) to extract orthologous sequences from across the Archaeplastida for said trees. We analyzed the genomes and transcriptomes of ten hornworts, along with the genomes of *Arabidopsis thaliana* and *Marchantia polymorpha*, to determine the presence of various components involved in plastid development (Lamesch et al., 2012; Bowman et al., 2017; Leebens-Mack et al., 2019; Li et al., 2020; Zhang et al., 2020; **Supplementary Table 2**). These orthology clusters (orthogroups) were identified using the OrthoFinder version 2.5.4 software (Emms and Kelly, 2015, 2019; **Supplementary Table 2**). To validate orthogroup presence/absence, we checked for reciprocal best hits using DIAMOND (Buchfink et al., 2015). Due to the difficulty in identifying orthologs for the import protein YCF1 in the Archaeplastida (de Vries et al., 2015), we employed a different strategy to identify orthologs for this gene. We extracted established YCF1 sequences from GenBank and UniProt and used them as queries for DIAMOND.

LTD protein. Subsequently, this SRP43/SRP54 complex binds to the FtsY receptor. GTP hydrolysis results in LHCP integration *via* the action of the ALB3 integral translocase (Xu et al., 2021; **Figure 1B**). Our results suggest that the cpSRP pathway is ubiquitous in all hornworts, as the core components of this pathway are present in the vast majority of our surveyed taxa; therefore, presence/absence of cpSRP components reveals no pattern with regard to mono-/polyplastidy or presence/absence of a pyrenoid. However, FtsY is absent in *L. dussii*, and LTD is absent in both *Anthoceros angustus* and *L. dussii*.

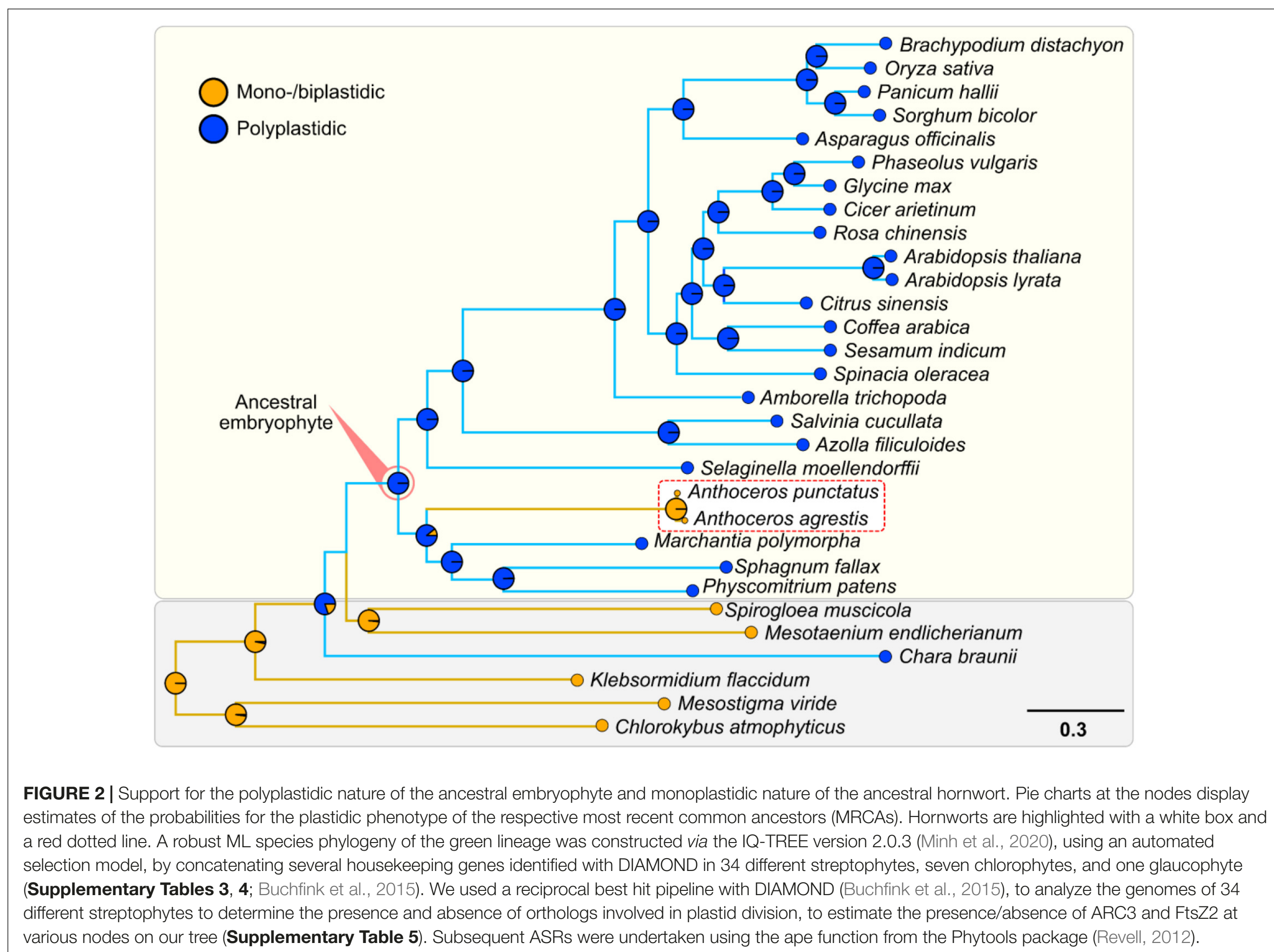

FIGURE 2 | Support for the polyplastidic nature of the ancestral embryophyte and monoplastidic nature of the ancestral hornwort. Pie charts at the nodes display estimates of the probabilities for the plastidic phenotype of the respective most recent common ancestors (MRCAs). Hornworts are highlighted with a white box and a red dotted line. A robust ML species phylogeny of the green lineage was constructed *via* the IQ-TREE version 2.0.3 (Minh et al., 2020), using an automated selection model, by concatenating several housekeeping genes identified with DIAMOND in 34 different streptophytes, seven chlorophytes, and one glaucophyte (**Supplementary Tables 3, 4**; Buchfink et al., 2015). We used a reciprocal best hit pipeline with DIAMOND (Buchfink et al., 2015), to analyze the genomes of 34 different streptophytes to determine the presence and absence of orthologs involved in plastid division, to estimate the presence/absence of ARC3 and FtsZ2 at various nodes on our tree (**Supplementary Table 5**). Subsequent ASRs were undertaken using the ape function from the Phytools package (Revell, 2012).

LOSS OF PLASTID DIVISION COMPONENTS COINCIDES WITH MONOPLASTIDY IN HORNWORTS

Plastid division in bryophytes is achieved by three components, namely, the outer and inner rings and most likely the peptidoglycan (PG) layer (**Figure 1**). The inner division ring (Z-ring) is comprised of FtsZ1, FtsZ2, and FtsZ3, while the outer division ring comprises ARC5 and FtsZ3 (Osteryoung and Pyke, 2014; Grosche and Rensing, 2017; **Figure 1B**). Z-ring and outer ring synchronization are achieved *via* an interplay of ARC6 and PDV2 (Osteryoung and Pyke, 2014). The PG layer is a relic of the chloroplast's cyanobacterial past, and it might be relevant in regulating chloroplast division in bryophytes and streptophyte algae (Hirano et al., 2016; Grosche and Rensing, 2017).

Hornworts appear to have differentially lost *both* ARC3 and FtsZ2 (**Figure 1A**). This differential loss correlates with this group of bryophytes reverting back to a monoplastidic, or near-monoplastidic, phenotype (**Figure 2** and **Supplementary Figures 3, 4**; Villarreal and Renner, 2012; Raven and Edwards, 2014; Li et al., 2017). Indeed, previous studies have shown that generating individual gene mutant lines of *ARC3* and *FtsZ2* in *A. thaliana* and the moss *Physcomitrium patens* causes fewer plastids (in the case of *arc3* mutants) or one giant plastid per cell (in the case of *ftsz2* mutants) (Pyke and Leech, 1992; Martin et al., 2009). ARC3 is part of the FtsZ family and unites an FtsZ domain with a C-terminal MORN domain (Zhang et al., 2013).

DISCUSSION

It is evident that hornwort—and bryophyte—emergence and diversification were accompanied by major instances of gene loss (Harris et al., 2021). Our results reinforce this hypothesis, specifically highlighting that the combined loss of certain genes may be responsible for the unique plastid phenotype observed in this group.

The absence of TIC20 in *L. dussii* could be the result of a transcriptome annotation and coverage issues (Cheon et al., 2020), since TIC20 is hypothesized to be a universal protein across the green lineage (Kalanon and McFadden, 2008; de Vries et al., 2015). Should this not be the case, then, maybe YCF1/TIC214 and TIC100 can compensate for TIC20's absence in a unique manner. Some putative absences of YCF1/TIC214 could also be the result of assembly

and/or annotation errors; however, the gene was lost without question in grasses, too (de Vries et al., 2015; Nakai, 2015b). The loss of this import protein does not lead to the loss of the entire import capacity (Bölter and Soll, 2017) and raises the question whether there is a functional, causative correlation between the loss of YCF1/TIC214 across these diverse embryophyte groups.

Considering cpSecE only plays an accessory role in protein translocation by tiling and rotating cpSecY's N-terminal half, its absence in some hornworts indicates that it might not be detrimental to the function of the cpSec pathway (**Figure 1B**; Park et al., 2014). If the cpTat pathway is indeed absent in most hornwort families, then this raises the question on how the thylakoids import folded proteins. Furthermore, all hornworts appear to lack STT proteins (**Figure 1A**), which mediate liquid-liquid phase transitions (LLPTs), allowing for more efficient sorting of cpTat substrates (**Figure 1**; Ouyang et al., 2020). cpTat-related LLPTs hence appear absent in hornworts or are regulated otherwise. The differential loss of FtsY and LTD in *L. dussii* could be a consequence of this species potentially losing TIC20, with this core TIC component being a key LTD interaction partner (Ouyang et al., 2011).

We found that the chloroplasts of all surveyed hornworts possess all the enzymes necessary for PG layer biosynthesis (**Figure 1A**), hinting toward a conserved function similar to that in the moss *P. patens* (Hirano et al., 2016). While ARC3 orthologs are absent in some polyplastidic seedless plants (such as *P. patens* and the lycophyte *Selaginella moellendorffii*), these species then possess orthologs for FtsZ2, which might compensate its loss to some degree (Rensing et al., 2008; Albert et al., 2011; Zhang et al., 2013). This is further supported by an ancestral state reconstruction analysis that demonstrates that the ancestral embryophyte possessed both ARC3 and FtsZ2 and was polyplastidic, the opposite of which is true for the ancestral hornwort (**Figure 2** and **Supplementary Figures 3**, **4**). We predict that the loss of both genes contributed to the monoplastidic nature of hornworts and that reintroducing them might induce a polyplastidic phenotype.

CONCLUSION AND OUTLOOK

We suggest that a consequence of some of plastid-related gene losses, including the combined loss of FtsZ2 and ARC3, resulted in hornworts reverting back to a monoplastidic phenotype, which the embryophyte ancestor was able to escape. If the knockout of ARC3 and FtsZ2 in *A. thaliana* and *P. patens* results in monoplastidic phenotypes, could one reverse evolution by expressing ARC3 and/or FtsZ2 in a hornwort? We anticipate our study to be a starting point for further experiments aimed at deconstructing bryophyte plastid biology and reconstructing new evolutionary hypotheses for outstanding questions in this topic. Next to exploring the monoplastidic bottleneck, hornworts might be able to shed new light on the import of folded proteins into the thylakoid of non-Anthocerotaceae hornworts, or the consequences of a potential TIC20 loss in *L. dussii* and the detailed function of YCF1; which, like all grasses, some hornworts appear to have lost.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

AM undertook the phylogenomic analyses, with help from SS and MK. AM, SG, PR, and EF wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

We are grateful for the support by the DFG (SFB 1208–2672 05415 and SPP2237–440043394) and VolkswagenStiftung (Life). AM is furthermore supported by the Moore and Simons Initiative grant (9743) of William F. Martin.

ACKNOWLEDGMENTS

We thank William F. Martin for his ongoing support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.863076/full#supplementary-material

REFERENCES

Albert, V. A., Aono, N., and Aoyama, T. (2011). Content Associated With the Evolution of Vascular Plants. *Science* 332, 960–963. doi: 10.1126/science.1203810

Bölter, B., and Soll, J. (2017). Ycf1/Tic214 Is Not Essential for the Accumulation of Plastid Proteins. *Mol. Plant* 10, 219–221. doi: 10.1016/j.molp.2016.10.012

Bowman, J. L., Kohchi, T., Yamato, K. T., Jenkins, J., Shu, S., Ishizaki, K., et al. (2017). Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. *Cell* 171, 287–304.e15. doi: 10.1016/j.cell.2017.09.030

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Cheon, S., Zhang, J., and Park, C. (2020). Is Phylotranscriptomics as Reliable as Phylogenomics? *Mol. Biol. Evol.* 37, 3672–3683. doi: 10.1093/molbev/msaa181

de Vries, J., and Gould, S. B. (2018). The monoplastidic bottleneck in algae and plant evolution. *J. Cell Sci.* 131:jcs203414. doi: 10.1242/jcs.203414

de Vries, J., Sousa, F. L., Bölter, B., Soll, J., and Goulda, S. B. (2015). YCF1: a green TIC? *Plant Cell* 27, 1827–1833. doi: 10.1105/tpc.114.135541

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157. doi: 10.1186/s13059-015-0721-2

Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238. doi: 10.1186/s13059-019-1832-y

Emms, D. M., and Kelly, S. (2021). SHOOT: phylogenetic gene search and ortholog inference. *bioRxiv* [Preprint]. doi: 10.1101/2021.09.01.458564

Frangedakis, E., Shimamura, M., Villarreal, J. C., Li, F. -W., Tomaselli, M., Waller, M., et al. (2020). The hornworts: morphology, evolution and development. *New Phytol.* 229, 735–754. doi: 10.1111/nph.16874

Gould, S. B., Waller, R. F., and McFadden, G. I. (2008). Plastid evolution. *Annu. Rev. Plant Biol.* 59, 491–517. doi: 10.1146/annurev.arplant.59.032607.092915

Grosche, C., and Rensing, S. A. (2017). Three rings for the evolution of plastid shape: a tale of land plant FtsZ. *Protoplasma* 254, 1879–1885. doi: 10.1007/s00709-017-1096-x

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., et al. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: assessing the Performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Harris, B. J., Clark, J. W., Schrempf, D., Szőllosi, G. J., Donoghue, P., Hetherington, A. M., et al. (2021). Divergent evolutionary trajectories of bryophytes and tracheophytes from a complex common ancestor of land plants. *bioRxiv* [Preprint]. doi: 10.1101/2021.10.28.466308

Harris, B. J., Harrison, C. J., Hetherington, A. M., and Williams, T. A. (2020). Phylogenomic Evidence for the Monophyly of Bryophytes and the Reductive Evolution of Stomata. *Curr. Biol.* 30, 2001–2012.e2. doi: 10.1016/j.cub.2020.03.048

Hirano, T., Tanidokoro, K., Shimizu, Y., Kawarabayasi, Y., Ohshima, T., Sato, M., et al. (2016). Moss chloroplasts are surrounded by a peptidoglycan wall containing D-amino acids. *Plant Cell* 28, 1521–1532. doi: 10.1105/tpc.16.00104

Jarvis, P., and López-Juez, E. (2013). Biogenesis and homeostasis of chloroplasts and other plastids. *Nat. Rev. Mol. Cell Biol.* 14, 787–802. doi: 10.1038/nrm3702

Kalanon, M., and McFadden, G. I. (2008). The chloroplast protein translocation complexes of *Chlamydomonas reinhardtii*: a bioinformatic comparison of Toc and Tic components in plants, green algae and red algae. *Genetics* 179, 95–112. doi: 10.1534/genetics.107.085704

Knopp, M., Garg, S. G., Handrich, M., and Gould, S. B. (2020). Major Changes in Plastid Protein Import and the Origin of the Chloroplastida. *iScience* 23:100896. doi: 10.1016/j.isci.2020.100896

Laarson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large data sets. *Bioinformatics* 30, 3276–3278. doi: 10.1093/bioinformatics/btu531

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–D1210. doi: 10.1093/nar/gkr1090

Leebens-Mack, J. H. M. S., Barker, E. J., Carpenter, M. K., Deyholos, M. A., Gitzendanner, S. W., Graham, I., et al. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574, 679–685. doi: 10.1038/s41586-019-1693-2

Lefort, V., Longueville, J.-E., and Gascuel, O. (2017). SMS: smart Model Selection in PhyML. *Mol. Biol. Evol.* 34, 2422–2424. doi: 10.1093/molbev/msx149

Li, F.-W., Nishiyama, T., Waller, M., Frangedakis, E., Keller, J., Li, Z., et al. (2020). Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. *Nat. Plants* 6, 259–272. doi: 10.1038/s41477-020-0618-2

Li, F.-W., Villarreal, A. J., and Szövényi, P. (2017). Hornworts: an Overlooked Window into Carbon-Concentrating Mechanisms. *Trends Plant Sci.* 22, 275–277. doi: 10.1016/j.tplants.2017.02.002

Ling, Q., Huang, W., Baldwin, A., and Jarvis, P. (2012). Chloroplast biogenesis is regulated by direct action of the ubiquitin-proteasome system. *Science* 338, 655–659. doi: 10.1126/science.1225053

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. (2021). BUSCO Update: novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* 38, 4647–4654. doi: 10.1093/molbev/msab199

Martin, A., Lang, D., Hanke, S. T., Mueller, S. J., Sarnighausen, E., Vervliet-Scheebaum, M., et al. (2009). Targeted gene knockouts reveal overlapping functions of the five physcomitrella patens ftsz isoforms in chloroplast division, chloroplast shaping, cell patterning, plant development, and gravity sensing. *Mol. Plant* 2, 1359–1372. doi: 10.1093/mp/ssp076

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015

Nakai, M. (2015a). The TIC complex uncovered: the alternative view on the molecular mechanism of protein translocation across the inner envelope membrane of chloroplasts. *Biochim. Biophys. Acta* 1847, 957–967. doi: 10.1016/j.bbabio.2015.02.011

Nakai, M. (2015b). Ycf1: a green TIC: response to the de Vries et al. Commentary. *Plant Cell* 27, 1834–1838. doi: 10.1105/tpc.15.00363

Osteryoung, K. W., and Pyke, K. A. (2014). Division and dynamic morphology of plastids. *Annu. Rev. Plant Biol.* 65, 443–472. doi: 10.1146/annurev-arplant-050213-035748

Ouyang, M., Li, X., Ma, J., Chi, W., Xiao, J., Zou, M., et al. (2011). LTD is a protein required for sorting light-harvesting chlorophyll-binding proteins to the chloroplast SRP pathway. *Nat. Commun.* 2:277. doi: 10.1038/ncomms1278

Ouyang, M., Li, X., Zhang, J., Feng, P., Pu, H., Kong, L., et al. (2020). Liquid-Liquid Phase Transition Drives Intra-chloroplast Cargo Sorting. *Cell* 180, 1144–1159.e20. doi: 10.1016/j.cell.2020.02.045

Park, E., Ménétret, J.-F., Gumbart, J. C., Ludtke, S. J., Li, W., Whynot, A., et al. (2014). Structure of the SecY channel during initiation of protein translocation. *Nature* 506, 102–106. doi: 10.1038/nature12720

Pyke, K. A., and Leech, R. M. (1992). Chloroplast division and expansion is radically altered by nuclear mutations in Arabidopsis thaliana. *Plant Physiol.* 99, 1005–1008. doi: 10.1104/pp.99.3.1005

Raven, J. A., and Edwards, D. (2014). Photosynthesis in Bryophytes and Early Land Plants. *Diversif. Evol. Environ.* 37, 29–58. doi: 10.1007/978-94-007-6988-5

Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., et al. (2008). The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* 319, 64–69. doi: 10.1126/science.1150646

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3, 217–223. doi: 10.1111/j.2041-210X.2011.00169.x

Richardson, L. G. L., and Schnell, D. J. (2020). Origins, function, and regulation of the TOC–TIC general protein import machinery of plastids. *J. Exp. Bot.* 71, 1226–1238. doi: 10.1093/jxb/erz517

Vaughn, K. C., Ligrone, R., Owen, H. A., Hasegawa, J., Campbell, E. O., Renzaglia, K. S., et al. (1992). The anthocerote chloroplast: a review. *New Phytol.* 120, 169–190. doi: 10.1111/j.1469-8137.1992.tb05653.x

Villarreal, J. C., Forrest, L. L., Wickett, N., and Goffinet, B. (2013). The plastid genome of the hornwort Nothoceros aenigmaticus (*Dendrocerotaceae*): phylogenetic signal in inverted repeat expansion, pseudogenization, and intron gain. *Am. J. Bot.* 100, 467–477. doi: 10.3732/ajb.1200429

Villarreal, J. C., and Renner, S. S. (2012). Hornwort pyrenoids, carbon-concentrating structures, evolved and were lost at least five times during the last 100 million years. *Proc. Natl. Acad. Sci. U. S. A.* 109, 18873–18878. doi: 10.1073/pnas.1213498109

Xu, X., Ouyang, M., Lu, D., Zheng, C., and Zhang, L. (2021). Protein Sorting within Chloroplasts. *Trends Cell Biol.* 31, 9–16. doi: 10.1016/j.tcb.2020.09.011

Zhang, J., Fu, X. X., Li, R. Q., Zhao, X., Liu, Y., Li, M. H., et al. (2020). The hornwort genome and early land plant evolution. *Nat. Plants* 6, 107–118. doi: 10.1038/s41477-019-0588-4

Zhang, M., Schmitz, A. J., Kadirjan-Kalbach, D. K., TerBush, A. D., and Osteryoung, K. W. (2013). Chloroplast division protein ARC3 regulates chloroplast FtsZ-ring assembly and positioning in Arabidopsis through interaction with FtsZ2. *Plant Cell* 25, 1787–1802. doi: 10.1105/tpc.113.111047

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 MacLeod, Raval, Stockhorst, Knopp, Frangedakis and Gould. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

8.2 Publication II: A mysterious cloak: the peptidoglycan layer of algal and plant plastids

Authors: **Alexander I. MacLeod**, Michael R. Knopp and Sven B. Gould

This article was published in the journal *Protoplasma* on the 21st of August 2023.

Contribution of Alexander István MacLeod:

MAJOR:

- Designed and undertook experiments
- Drafted the manuscript
- Created all figures
- Corresponding author

All data generated with this manuscript (including supplementary information) can be accessed here: https://link.springer.com/article/10.1007/s00709-023-01886-y

The article is reprinted here from the Journal of Protoplasma under the Creative Commons CC BY license.

Protoplasma
https://doi.org/10.1007/s00709-023-01886-y

SHORT COMMUNICATION

A mysterious cloak: the peptidoglycan layer of algal and plant plastids

Alexander I. MacLeod[1] · Michael R. Knopp[1] · Sven B. Gould[1]

Received: 3 May 2023 / Accepted: 23 July 2023
© The Author(s) 2023

Abstract

The plastids of algae and plants originated on a single occasion from an endosymbiotic cyanobacterium at least a billion years ago. Despite the divergent evolution that characterizes the plastids of different lineages, many traits such as membrane organization and means of fission are universal—they pay tribute to the cyanobacterial origin of the organelle. For one such trait, the peptidoglycan (PG) layer, the situation is more complicated. Our view on its distribution keeps on changing and little is known regarding its molecular relevance, especially for land plants. Here, we investigate the extent of PG presence across the Chloroplastida using a phylogenomic approach. Our data support the view of a PG layer being present in the last common ancestor of land plants and its remarkable conservation across bryophytes that are otherwise characterized by gene loss. In embryophytes, the occurrence of the PG layer biosynthetic toolkit becomes patchier and the availability of novel genome data questions previous predictions regarding a functional coevolution of the PG layer and the plastid division machinery-associated gene FtsZ3. Furthermore, our data confirm the presence of penicillin-binding protein (PBP) orthologs in seed plants, which were previously thought to be absent from this clade. The 5-7 nm thick, and seemingly unchanged, PG layer armoring the plastids of glaucophyte algae might still provide the original function of structural support, but the same can likely not be said about the only recently identified PG layer of bryophyte and tracheophyte plastids. There are several issues to be explored regarding the composition, exact function, and biosynthesis of the PG layer in land plants. These issues arise from the fact that land plants seemingly lack certain genes that are believed to be crucial for PG layer production, even though they probably synthesize a PG layer.

Keywords Plastid evolution · Peptidoglycan · Murein layer · Plant evolution · Chloroplastida

Introduction

Depending on the dating method, plastids emerged between 1.2 and 2.5 billion years ago (Bowles et al. 2023). Part of transforming an endosymbiont into an organelle involves transferring the majority of genetic information to the nucleus of the host cell through endosymbiotic gene transfer (EGT) (Deusch et al. 2008). Hence, the vast majority of photosynthesis- and plastid biogenesis-associated genes are cytosolically translated and then imported (Miyagishima et al. 2014; Knopp et al. 2020; Dowson et al. 2022). Enzymes of peptidoglycan (PG) layer biosynthesis, if present, are no exception. The peptidoglycan polymer provides bacterial cell walls with a structure and rigidity to protect themselves against biotic and abiotic stressors such as osmotic pressure, bacteriophages, heat, and salinity (Vollmer et al. 2008). The PG layer biosynthetic toolkit of the Chloroplastida is made up of ten key proteins: seven Mur enzymes (MurA-MurG), a DDL ligase (D-alanine:D-alanine ligase), the MraY enzyme, and penicillin-binding proteins (PBPs) (Dowson et al. 2022). PBPs have hitherto only been identified in algae, bryophytes, and seedless vascular plants (van Baren et al. 2016).

The presence of a PG layer in the plastids of the Archaeplastida has been documented for the Glaucophyta and Chloroplastida, but not yet for the Rhodophyta (Björn 2020). A cyanelle, the glaucophyte plastid, possesses a reduced yet still relatively thick PG layer between its two membranes with consequences for protein import (Steiner et al. 2005). The PG layer is also present in some members of the green lineage, such as mosses, but the degree to which this trait is conserved in multiple clades remains unresolved (Bachy

Communicated by Handling Editor: Jaideep Mathur.

✉ Alexander I. MacLeod
macleod@hhu.de

[1] Institute for Molecular Evolution, Heinrich Heine University of Düsseldorf, 40225 Düsseldorf, Germany

Published online: 21 August 2023

et al. 2022). And while the peptidoglycan layer of mosses has been functionally characterized in parts (Hirano et al. 2016; Dowson et al. 2022), the same cannot be said for other embryophytes that possess orthologs for a full PG biosynthetic toolkit.

Understanding how the PG layer has evolved within the green lineage can provide valuable insights into the evolution of various aspects of plastid development, such as the timing of specific gene losses and gains. This could potentially clarify why some organisms within the Chloroplastida have unique plastid characteristics (Hirano et al. 2016; Li et al. 2017; de Vries and Gould 2018; MacLeod et al. 2022). In this study, we undertake a comprehensive and evidence-based phylogenomic approach on 48 genomes from the Chloroplastida supergroup and aim to delineate the phylogenetic distribution and evolution of the PG layer in the green lineage. We highlight that genes encoding proteins associated with PG layer biosynthesis have an uncommon phylogenetic distribution in the Chloroplastida and that the PG layer did not evolve concurrently with a component of the plastid division machinery, FtsZ3, as recently suggested (Grosche and Rensing 2017). The results underscore the PG layer's existence in gymnosperms and spermatophytes, for which dedicated studies exploring its biological relevance for the plastid organelle are surprisingly sparse.

Material and methods

Determining the phylogenetic distribution of PG layer biosynthetic enzymes in the Chloroplastida, and plastid division components in chlorophyte algae

Protein sequence IDs of the ten key enzymes involved in peptidoglycan layer biosynthesis were retrieved from the genome of the liverwort *Marchantia polymorpha* and used as queries for the identification of orthology clusters (Bowman et al. 2017). OrthoFinder version 2.5.4 was used to identify orthologs among the input genomes from 48 Chloroplastida members (Goodstein et al. 2012; Hori et al. 2014; O'Leary et al. 2016; Bowman et al. 2017; Li et al. 2018, 2020b, 2020a ; Nishiyama et al. 2018; Cheng et al. 2019; Wang et al. 2020; Zhang et al. 2020; Grigoriev et al. 2021; Huang et al. 2022), with a BLASTp e-value threshold of 1×10^{-9}. The phylogenetic distribution of enzymes involved in peptidoglycan layer biosynthesis was determined by examining the presence or absence of orthologous groups (orthogroups) containing these proteins across different members of the Chloroplastida. This exact pipeline was replicated to determine the presence/absence of plastid division components in 37 chlorophyte algae and one Prasinodermophyta (Li et al. 2020b; Grigoriev et al. 2021). Furthermore, where a given orthology cluster contained a protein family, which was the case for FtsZ proteins, phylogenetic trees were constructed to separate each protein into a respective subfamily. Sequence alignments were undertaken using MAFFT v7.471 using the LINSI parameter, with tree building being undertaken using IQ-TREE v2.0.3 using an automated selection model, with a 100 non-parametric bootstraps (Katoh et al. 2002; Minh et al. 2020). Finally, we used the SHOOT.bio phylogenetic application to determine whether PG layer biosynthetic genes in some seed plants branch within the terrestrial clade (Emms and Kelly 2022).

Phylogenetic species tree construction

The species tree is based on a weighted concatenated alignment from 11 individual alignments. The first step was to calculate protein families including all sequences from the 48 analyzed genomes. Pairwise local identities were determined via DIAMOND (v2.0.1) and filtered for all reciprocal best blast hit pairs with at least 40% local sequence identity and a maximum e-value of 1×10^{-10} (Buchfink et al. 2015). A total of 210 protein clusters contained sequences from all 48 genomes; however, no single-copy gene cluster was found. To create a robust reference tree, 11 protein families were chosen in which only few genomes were represented by more than one sequence. For these clusters, alignments were calculated with MAFFT v7.471 using the LINSI parameter and the duplicate sequences were manually removed, favoring the copies that did not show major deletions or insertions to yield a robust phylogeny (Katoh et al. 2002). All 11 alignments were concatenated while equalizing their phylogenetic signal using a weighted concatenation approach. The final tree was built by IQ-TREE v2.0.3 (Minh et al. 2020) with 100 non-parametric bootstraps using the LG + F + R7 substitution model. Best-fit model identification was via IQ-TREE's model finder (Minh et al. 2020). Tree trimming and visualization was carried out using the ggtree R package (Yu et al. 2017). A species tree of chlorophytes—used to plot the phylogenetic distribution of plastid division machinery components in this phylum—was estimated using STAG in the OrthoFinder run (Emms and Kelly 2015, 2015, 2018).

Delineating domain architecture and function of orthologous sequences

InterProScan v5 (Jones et al. 2014) was used to delineate the basic domain architecture and function of protein sequences involved in peptidoglycan layer biosynthesis. The program was used to identify protein domains, annotate their functions, and determine the arrangement and composition of the domains in the protein sequences.

Results

Structural conservation of PG layer biosynthetic enzymes across the Chloroplastida

Protein domain and gene ontology analyses show a high level of structural conservation in PG layer biosynthetic enzymes, from algae to angiosperms, and confirm that these proteins likely play key roles in peptidoglycan biosynthesis in the species analyzed (Fig. 1). Furthermore, while previous studies have suggested that the PG layer was differentially lost in the MRCA of seed plants (spermatophytes) (Grosche and Rensing 2017), the full toolkit for the biosynthesis of peptidoglycan is identified in at least three phylogenetically distant members of the seed clade: *Thuja plicata* (Gymnosperms), *Asparagus officinalis* (Monocots), and *Citrus sinensis* (Eudicots) (Fig. 1). This includes the identification of PBP family orthologs in the seed clade.

No evident correlation between the presence of the PG layer and any of the three FtsZ proteins

The PG layer plays a key role in regulating chloroplast division in bryophytes and streptophyte algae (Machida et al. 2006; Homi et al. 2009; Hirano et al. 2016; Grosche and Rensing 2017; Dowson et al. 2022). The GTPases FtsZ1, FtsZ2, and FtsZ3 are central to plastid division, and form versatile heteropolymers that establish constriction sites, facilitating the division of plastids in a coordinated and efficient way (Martin et al. 2009; Yoshida et al. 2016). FtsZ3 was suggested to play a role in regulating the biogenesis of the PG layer due to an alleged correlation between these two traits (Grosche and Rensing 2017). There are, however, multiple exceptions to this correlation. For example, the spermatophytes *Thuja plicata*, *Asparagus officinalis*, and *Citrus sinensis* possess orthologs representing a full enzymatic toolkit for PG layer biosynthesis (Fig. 2). In addition, the hornwort *Nothoceros aenigmaticus* and the phylogenetically distant chlorophytes, *Chloropicon primus*, *Ulva mutabilis*, and *Micromonas pusilla*, all likely possess a PG layer between their chloroplast membranes (Bachy et al. 2022; MacLeod et al. 2022), but lack FtsZ3 (Fig. 2). In summary, the now-available genomes do not support a PG layer and FtsZ3 coevolution or functional connection. In fact, it appears that the presence of the chloroplast PG layer is not dependent on the presence of any one specific protein of the FtsZ family (Fig. 2).

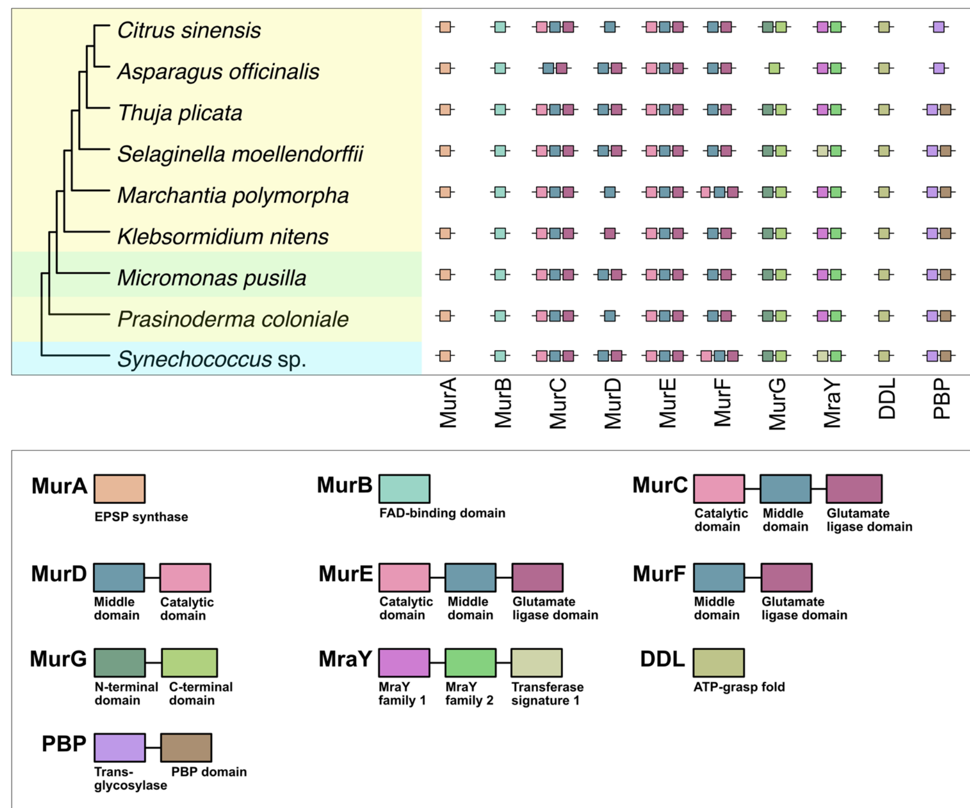

Fig. 1 The PG layer biosynthetic toolkit is structurally well conserved from cyanobacteria to angiosperms. Cy*, Cyanobacteria; P*, Prasinodermophyta; Chl*, Chlorophyta

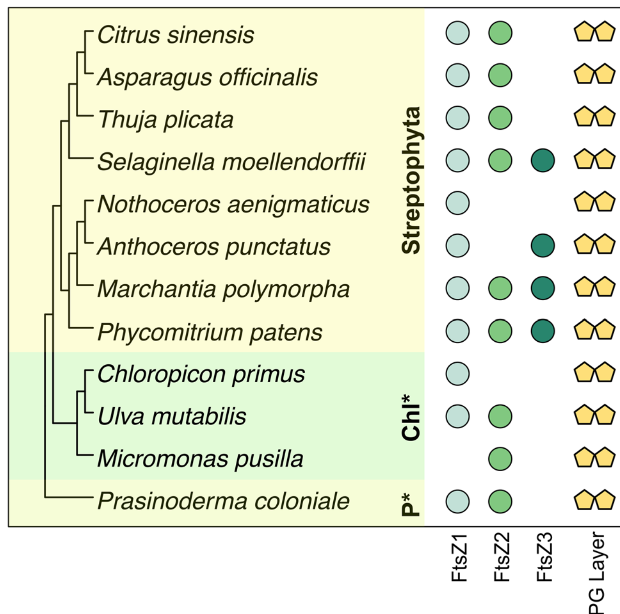

Fig. 2 Phylogenetic distribution of the FtsZ plastid division proteins in Chloroplastida that likely have a PG layer, highlighting the unlikely coevolution between any specific FtsZ protein and the plastid enveloping murein layer. P*, Prasinodermophyta; Chl*, Chlorophyta. Ortholog metadata for FtsZs in the Streptophyta was obtained from MacLeod et al. (2022)

Discussion

The unusual phylogenetic distributions of individual genes or even entire biosynthetic pathways are sometimes the result of the identification of bacterial false positive contaminations (Koutsovoulos et al. 2016; Husnik and McCutcheon 2018; Goig et al. 2020). This is unlikely to be the case regarding PG layer biosynthetic enzymes from *T. plicata*, *A. officinalis*, and *C. sinensis*, as they branch deep within the embryophyte clade (Emms and Kelly 2022). As such, all available data suggest a monophyletic origin of the pathway (Li et al. 2020b) and an independent differential loss in various taxa across the Archaeplastida (Bachy et al. 2022).

Most angiosperms do not seem to encode for a complete set of enzymes synthesizing the PG layer. However, they all share four enzymes related to the process (MurE, MurG, MraY, and DDL), called the "4-PGN" set, and recent experimental work suggests that two angiosperms, *Arabidopsis thaliana* and *Nicotiana benthamiana*, may have a PG layer surrounding their chloroplasts (Tran et al. 2023). If true, then it would suggest that these species use a different set of enzymes and biochemistry to synthesize parts of the PG layer, with the 4-PGN set playing a key role, therefore being retained. Intriguingly, the retention of the same set of genes (± 1) occurred independently in some chlorophyte algae such as *Micromonas commoda* (van Baren et al. 2016). It raises the question whether they have been retained for the same functional reason, which is likely but not proven.

While recent biochemical and metabolomic analyses suggest that components of the moss peptidoglycan biosynthetic pathway—specifically, the active sites of core ligase enzymes—display strict conservation in comparison to the PG layer biosynthetic pathway of cyanobacteria (Dowson et al. 2022), the FtsZ3 PDVM component is unlikely to play a role in PG layer biosynthesis in moss. Genome analyses, including our own, indicate that the PG layer exists in all three phyla of Chloroplastida (van Baren et al. 2016; Grosche and Rensing 2017; Li et al. 2020b). There is, however, no strict connection between FtsZ3, or any FtsZ gene, and the PG layer in terms of the FtsZ-based ring's association with this cyanobacterial relic. Therefore, any gene from the FtsZ family can likely perform its role in regulating the formation of plastid division rings, indicating functional redundancy within this family.

Conclusion

The peptidoglycan layer of chloroplasts was present in the MRCA of Chloroplastida and lost in most Chlorophyta and many Streptophyta, but retained in the Prasinodermophyta. Since the number of annotated genome assemblies for this latter phylum still stands at a mere one, it will be interesting for future genome mining experiments to elucidate whether the PG layer can be characterized—either biochemically or genomically—in this basal-branching green phylum. One can conclude that the PG layer is present in the chloroplasts of at least three phylogenetically distant spermatophytes, likely more, suggesting that peptidoglycan is more widespread in the chloroplasts of this phylum than previously thought. Moreover, the correlation between the presence of the PG layer and the plastid division protein FtsZ3 is no longer supported. Based on this evidence, upcoming studies should now focus on clarifying both the biochemical characteristics and the biological significance of the PG layer in angiosperms, which were previously thought to lack this ancient cyanobacterial feature.

Acknowledgements We thank Professor William F. Martin for his ongoing support.

Author contributions AIM designed the experiments, undertook all comparative genomic analyses, and wrote the manuscript. MRK generated the main species tree. SBG suggested the topic to AIM and contributed to the improvement of the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. We are grateful for the support by the DFG (SPP2237–440043394). AM is furthermore supported by the Moore and Simons Initiative grant (9743) of William F. Martin.

Data availability For questions about accessing the data from this study, please contact the corresponding author directly.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Bachy C, Wittmers F, Muschiol J, Hamilton M, Henrissat B, Worden AZ (2022) The land–sea connection: insights into the plant lineage from a green algal perspective. Annu Rev Plant Biol 73:585–616. https://doi.org/10.1146/annurev-arplant-071921-100530

Björn LO (2020) Peptidoglycan in eukaryotes: unanswered questions. Phytochemistry 175:112370. https://doi.org/10.1016/j.phytochem.2020.112370

Bowles AMC, Williamson CJ, Williams TA, Lenton TM, Donoghue PCJ (2023) The origin and early evolution of plants. Trends Plant Sci 28:312–329. https://doi.org/10.1016/j.tplants.2022.09.009

Bowman JL, Kohchi T, Yamato KT, Jenkins J, Shu S, Ishizaki K et al (2017) Insights into land plant evolution garnered from the Marchantiapolymorpha genome. Cellhttps://doi.org/10.1016/j.cell.2017.09.030

Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176

Bowman JL, Kohchi T, Yamato KT, Jenkins J, Shu S, Ishizaki K et al (2017) Insights into land plant evolution garnered from the Marchantiapolymorpha genome. Cellhttps://doi.org/10.1016/j.cell.2019.10.019

Deusch O, Landan G, Roettger M, Gruenheit N, Kowallik KV, Allen JF et al (2008) Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. Mol Biol Evol. 10.1093https://doi.org/10.1093/molbev/msn022

de Vries J, and Gould SB (2018) Themonoplastidic bottleneck in algae and plant evolution. J Cell Scihttps://doi.org/10.1242/jcs.203414

Dowson AJ, Lloyd AJ, Cuming AC, Roper DI, Frigerio L, Dowson CG (2022) Plant peptidoglycan precursor biosynthesis: conservation between moss chloroplasts and Gram negative bacteria. Plant Physiol 190(1):165–179. https://doi.org/10.1093/plphys/kiac176

Emms DM, and Kelly S (2018) STAG: species tree inference from all genes. *bioRxiv*, 267914–267914. https://doi.org/10.1101/267914.

Emms DM, Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 16:157–157. https://doi.org/10.1186/s13059-015-0721-2

Emms DM, Kelly S (2022) SHOOT: phylogenetic gene search and ortholog inference. Genome Biol 23:85–85. https://doi.org/10.1186/s13059-022-02652-8

Goig GA, Blanco S, Garcia-Basteiro AL, Comas I (2020) Contaminant DNA in bacterial sequencing experiments is a major source of false genetic variability. BMC Biol 18:24. https://doi.org/10.1186/s12915-020-0748-z

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178–D1186. https://doi.org/10.1093/nar/gkr944

Grigoriev IV, Hayes RD, Calhoun S, Kamel B, Wang A, Ahrendt S et al (2021) PhycoCosm, a comparative algal genomics resource. Nucleic Acids Res 49:D1004–D1011. https://doi.org/10.1093/nar/gkaa898

Grosche C, Rensing SA (2017) Three rings for the evolution of plastid shape: a tale of land plant FtsZ. Protoplasma. https://doi.org/10.1007/s00709-017-1096-x

Hirano T, Tanidokoro K, Shimizu Y, Kawarabayasi Y, Ohshima T, Sato M et al (2016) Moss chloroplasts are surrounded by a peptidoglycan wall containing D-amino acids. Plant Cell. https://doi.org/10.1105/tpc.16.00104

Homi S, Takechi K, Tanidokoro K, Sato H, Takio S, and Takano H (2009) The peptidoglycan biosynthesis genes MurA and mray are related to chloroplast division in the moss Physcomitrella patens. Plant Cell Physiolhttps://doi.org/10.1093/pcp/pcp158

Hori K, Maruyama F, Fujisawa T, Togashi T, Yamamoto N, Seo M et al (2014) Klebsormidiumflaccidum genome reveals primary factors for plant terrestrial adaptation. Nat Communhttps://doi.org/10.1038/ncomms4978

Huang X, Wang W, Gong T, Wickell D, Kuo L-Y, Zhang X et al (2022) The flying spider-monkey tree fern genome provides insights into fern evolution and arborescence. Nature Plants 8:500–512. https://doi.org/10.1038/s41477-022-01146-6

Husnik F, McCutcheon JP (2018) Functional horizontal gene transfer from bacteria to eukaryotes. Nat Rev Microbiol 16:67–79. https://doi.org/10.1038/nrmicro.2017.137

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C et al (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Katoh K, Misawa K, Kuma KI, and Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucl Acid Reshttps://doi.org/10.1093/nar/gkf436

Knopp M, Garg SG, Handrich M, Gould SB (2020) Major changes in plastid protein import and the origin of the Chloroplastida. iScience 23(3):100896. https://doi.org/10.1016/j.isci.2020.100896

Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C et al (2016) No evidence for extensive horizontal gene transfer in the genome of the tardigrade Hypsibius dujardini. Proc Natl Acad Sci 113:5053–5058. https://doi.org/10.1073/pnas.1600338113

Li F-W, Nishiyama T, Waller M, Frangedakis E, Keller J, Li Z et al (2020a) Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. Nature Plants 6:259–272. https://doi.org/10.1038/s41477-020-0618-2

Li L, Wang S, Wang H, Sahu SK, Marin B, Li H et al (2020b) The genome of Prasinoderma coloniale unveils the existence of a third phylum within green plants. Nature Ecol Evol 4:1220–1231. https://doi.org/10.1038/s41559-020-1221-7

Li FW, Villarreal AJ, and Szövényi P (2017) Hornworts: an overlooked window into carbon-concentrating mechanisms. Trends in Plant Science 22. https://doi.org/10.1016/j.tplants.2017.02.002.

Li FW, Brouwer P, Carretero-Paulet L, Cheng S, De Vries J, Delaux PM et al (2018) Fern genomes elucidate land plant evolution and cyanobacterial symbioses. Nature Plantshttps://doi.org/10.1038/s41477-018-0188-8

Machida M, Takechi K, Sato H, Chung SJ, Kuroiwa H, Takio S et al (2006) Genes for the peptidoglycan synthesis pathway are essential for chloroplast division in moss. Proc Natl Acad Sci USA. https://doi.org/10.1073/pnas.0510693103

MacLeod AI, Raval PK, Stockhorst S, Knopp MR, Frangedakis E, and Gould SB (2022) Loss of plastid developmental genes coincides with a reversion to monoplastidy in hornworts. Front Plant Sci 13. https://doi.org/10.3389/fpls.2022.863076.

Martin A, Lang D, Heckmann J, Zimmer AD, Vervliet-Scheebaum M, and Reski R (2009) A uniquely high number of ftsZ genes in the moss Physcomitrella patens. Plant Biol. https://doi.org/10.1111/j.1438-8677.2008.00174.x.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A et al (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37:1530–1534. https://doi.org/10.1093/molbev/msaa015

Miyagishima S, Nakamura M, Uzuka A, and Era A (2014) FtsZ-less prokaryotic cell division as well as FtsZ- and dynamin-less chloroplast and non-photosynthetic plastid division. Front Plant Sci 5. https://doi.org/10.3389/fpls.2014.00459.

Nishiyama T, Sakayama H, de Vries J, Buschmann H, Saint-Marcoux D, Ullrich KK et al (2018) The Chara genome: secondary complexity and implications for plant terrestrialization. Cell. https://doi.org/10.1016/j.cell.2018.06.033

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R et al (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44:D733–D745. https://doi.org/10.1093/nar/gkv1189

Steiner JM, Yusa F, Pompe JA, Löffelhardt W (2005) Homologous protein import machineries in chloroplasts and cyanelles1. Plant J 44:646–652. https://doi.org/10.1111/j.1365-313X.2005.02559.x

Tran X, Keskin E, Winkler P, Braun M, and Kolukisaoglu Ü. (2023) The chloroplast envelope of angiosperms contains a peptidoglycan layer. Cells 12. https://doi.org/10.3390/cells12040563.

van Baren MJ, Bachy C, Reistetter EN, Purvine SO, Grimwood J, Sudek S et al (2016) Evidence-based green algal genomics reveals marine diversity and ancestral characteristics of land plants. BMC Genomics 17:267–267. https://doi.org/10.1186/s12864-016-2585-6

Vollmer W, Blanot D, De Pedro MA (2008) Peptidoglycan structure and architecture. FEMS Microbiol Rev 32:149–167. https://doi.org/10.1111/j.1574-6976.2007.00094.x

Wang S, Li L, Li H, Sahu SK, Wang H, Xu Y et al (2020) Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. Nature Plantshttps://doi.org/10.1038/s41477-019-0560-3

Yoshida Y, Mogi Y, TerBush AD, and Osteryoung KW (2016) Chloroplast FtsZ assembles into a contractible ring via tubulin-like heteropolymerization. Nature Plantshttps://doi.org/10.1038/NPLANTS.2016.95

Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y (2017) ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol 8:28–36. https://doi.org/10.1111/2041-210X.12628

Zhang J, Fu XX, Li RQ, Zhao X, Liu Y, Li MH et al (2020) The hornwort genome and early land plant evolution. Nature Plants. https://doi.org/10.1038/s41477-019-0588-4

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

8.3 Publication III: A molecular atlas of plastid and mitochondrial adaptations across the evolution from chlorophyte algae to angiosperms

Authors: Parth K. Raval, **Alexander I. MacLeod** and Sven B. Gould

This article was uploaded to the *bioRxiv* preprint server on the 6th of September 2023, and submitted to the journal *PLOS Biology* for independent review as of the 12th of September 2023

Contribution of Alexander István MacLeod:

MINOR:

- Drafted the Introduction to the manuscript
- Significantly contributed to the Results and Discussion sections on plastid development

All data generated with this manuscript (including supplementary information) can be accessed here: https://www.biorxiv.org/content/10.1101/2023.09.01.555919v1

The article is reprinted here from bioRxiv under the Creative Commons CC BY license.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

A molecular atlas of plastid and mitochondrial adaptations across the evolution from chlorophyte algae to angiosperms

Parth K. Raval, Alexander I. MacLeod, Sven B. Gould

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

Abstract

Algae and plants carry two organelles of endosymbiotic origin that have been co-evolving in their host cells for more than a billion years. The biology of plastids and mitochondria can differ significantly across major lineages and organelle adaptation likely accompanied the conquering of new ecological niches including land. Based on organelle proteome data, and the genomes of 168 phototrophic (Archaeplastida) versus a broad range of 518 non-phototrophic eukaryotes, we screened for changes in plastid and mitochondrial biology across one billion years of evolution. Taking into account 331,571 protein families (or orthogroups), we identify 31,625 protein families that are unique to primary plastid-bearing eukaryotes. 1961 and 846 protein families are predicted to operate in plastids and mitochondria, respectively. Tracing the evolutionary history of these protein families through evolutionary time uncovers the significant remodeling the organelles experienced from algae to land plants. The analyses of gained orthogroups identifies molecular adaptations of organelle biology that connect to the diversification of major lineages and facilitated major transitions from chlorophytes en route to the global greening and origin of angiosperms.

Keywords: plastid evolution, plant mitochondria, terrestrialization, organelle proteomes

Author of correspondence: gould@hhu.de

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

Fewer natural phenomena have been as transformative to planet Earth as the global greening through plants [1,2]. The conquest of land rests on the emergence and expansion of the Chloroplastida, also referred to as the Viridiplantae or simply the green lineage. The Chloroplastida are made up of three phyla: chlorophytes, streptophytes and the prasinodermophytes that are thought to be the sister lineage to the two former [3]. Chloro- and prasinodermophytes are represented by algae only, whereas streptophytes are made up of algae and embryophytes, the latter uniting all land plants [3–5]. The list of key adaptations that fostered land plant expansion in a macroevolutionary context are multiple: roots, a mutualistic symbiosis with fungi, stomata, a cuticle, polyplastidy, and an expansion of many metabolite families such as flavonoids to name a few [1,3–10]. These innovations provided embryophytes with decisive fitness advantages relative to their non-terrestrial chloro-, prasinodermato- and streptophyte algal relatives [1,11].

The eponymous organelle of plants, the chloroplast, underwent various changes, too. It adapted in multiple ways to the challenges characterizing the habitat the last common ancestor of land plants (LCA) encountered. Improving a stress response was necessary to deal for instance with increased levels of ultraviolet (UV) high light stress and to cope with temperature shifts that change rapidly on land in contrast to in water [12–14]. Polyplastidy, a phenomenon that separates plastid from nuclear division, leading to cells that can harbor more than one plastid per cell, was part of being able to develop larger body plans [12,15,16]. To communicate stress and the need for component biosynthesis, an elaborate retrograde signaling evolved on the basis of messenger proteins such as GUN1 or WHIRLY [17]. In combination, these adaptations were decisive for the success of streptophytes, which is evident in the number of species they have evolved and their domination of biomass [1,18].

Plastids do not operate autonomously, but are part of an intricate metabolic network and even physically interact with other compartments such as the endoplasmic reticulum and peroxisomes [19,20]. Pronounced metabolic and physical interactions of plastids also concern the only other plant compartment of ancient endosymbiotic origin: the mitochondrion. Plant mitochondria are much less in the focus of plant research. Next to their canonical functions, they are known to be involved in immunity, lipid metabolism and other (eco)physiological processes often in crosstalk with the photosynthetic organelle [21,22]. Like plastids, mitochondria were critical in the evolution and continued adaptation of important physiological traits in the green lineage. A notable example includes malate decarboxylation in the C4 photosynthetic pathway [23] – a convergent trait that improves plant photosynthetic efficiency in warm and dry habitats [24].

In spite of the importance of these two organelles of endosymbiotic origin in coordinating their duties, the evolution of components specific to chloroplast and mitochondrial biology has not been explicitly studied in light of streptophyte evolution and in particular plant terrestrialization. Previous work has determined genes specific to certain plant clades, including valuable resources such as the "GreenCut" [25]. Such analyses, however, are not focused on organelle biology nor clustered protein families. They were also limited by a low number of archaeplastidal genomes and insufficient methods for orthology inference available at that time. Since then, genome assemblies of members from previously unsampled clades has increased manyfold [11,26–34] and more organelle proteomes and better functional annotations are available. Similarly, and concomitantly, the development of novel and accurate algorithms for orthology inference [35–38] allow to now identify critical changes and adaptations in an eco-evo context of plastid and mitochondrial biology that underpin the success of the Chloroplastida.

Here, we curate a database of protein families unique to the green lineage. We plot their evolution across the major splits in the evolutionary history of streptophytes, focusing on the biology of the two organelles of endosymbiotic origin. We report that the number of plastid- and mitochondria-associated

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

protein families changes most significantly at two evolutionary bifurcations: firstly, at the green lineage itself, and secondly at the split between Zygnematophyceae and embryophytes at the water to land transition. The newly recruited protein families influenced organellar processes such as carbon and lipid metabolism, information processing and organelle development. We provide an extensive catalogue of the changes the proteomes of plastid and mitochondria experienced throughout streptophyte evolution, which offers multiple angles from which to explore major evolutionary transitions such as the conquest of land and embryophyte diversification.

Results

Half of the chloroplastida protein families are unique to embryophytes

Out of a total of 12,862,035 proteins, 95% were categorized from 686 eukaryotes (Table S1A) and grouped into 331,570 orthogroups (Table S1B). From these, 31,650 were present only in chloroplastida, and classified as Green Ortho Groups (GOGs) (Fig. S1 and Table S1C-D). An examination of GOG distribution among green species revealed that around half of all GOGs were unique to terrestrial plants (Fig. 1A). Approximately 400 GOGs appeared in more than 90% of species, termed here as 'core GOGs' (Fig. 1B). For only 5% of all GOGs, a functional annotation could be identified (Fig. 1C, Table S1E). Embryophyte-specific GOGs reflected a comparable percentage of this functional ambiguity, yet maintained a consistent distribution of identified functions, except increased fraction of membrane trafficking proteins (Fig. 1D, Table S1F). Notably, for core GOGs the number is higher, with 30% having functional annotations spanning areas such as photosynthesis, mitochondrial formation, trafficking, and information processing (Fig. 1E, Table S1G). The functions of a vast majority of the GOGs remain elusive (Table S1H), numbers that mirror those of previous studies [25], however, and hence provide an excellent ground for experimental exploration.

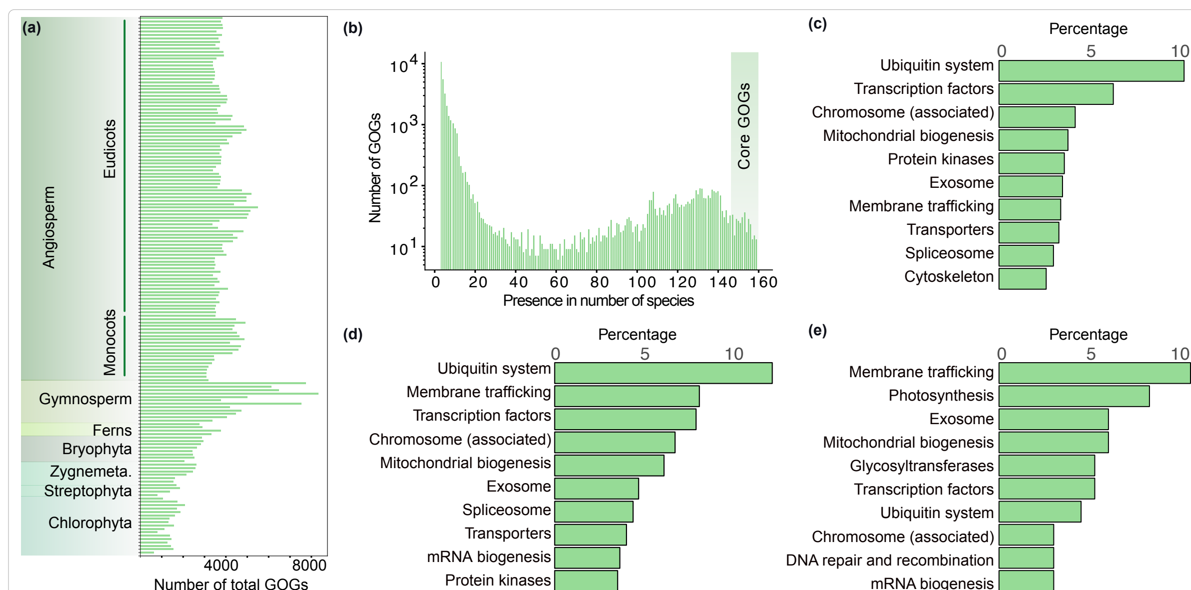

Fig. 1: Distribution and functional annotation of green orthogroups (GOGs). (a) Total number of GOGs present in each species from major Chloroplastida taxa. (b) Number of GOGs as a function of their presence across 159 Chloroplastida species. Major functional categories of 4.71% of all GOGs (c), 3.96% of the embryophyte GOGs (d) and 27.9 % of the core GOGs (e).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The origin and diversification of the Chloroplastida accompanied expansion of their organelle proteomes

To investigate changes in the proteomes of plastids and mitochondria, we curated 1961 plastid and 846 mitochondrial orthogroups (POGs and MOGs, respectively) based on published proteome data and homology-based protein clustering (Fig. S1B, Table S2A-C). In comparison to rhodophytes and glaucophytes, the green lineage has almost twice as many POGs (Fig. 2A). Within the green lineage, from the Zygnematophyceae and embryophytes onwards, plastid proteomes further expanded both in terms of the number of proteins within each POG and the number of unique POGs. The former is likely a consequence of genome duplications but the latter underscores functional divergence that followed gene duplications. The distribution of MOGs appears qualitatively similar to that of POGs (Fig. 2B, Table S2D-E). 60% of the POGs could be functionally annotated, predominantly operating in biosynthetic and other metabolic pathways such as photosynthesis (Fig. 2C, Table S2F). Around 75% of the MOGs could be functionally annotated and they predominantly operate in mitochondrial biogenesis, membrane trafficking and translation (Fig. 2D, Table S2G). Protein synthesis related proteins are abundant in both, POGs and MOGs, highlighting the biosynthetic capacity of the organelles. Proteins for mitochondrial biogenesis also appear in both. This encompasses numerous PPR and mTERF proteins (crucial for RNA editing) and proteins involved in various other information processing activities, probable to function in both organelles. Overall, the trends show that embryophytes have increased numbers of endosymbiotic organelle protein families.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

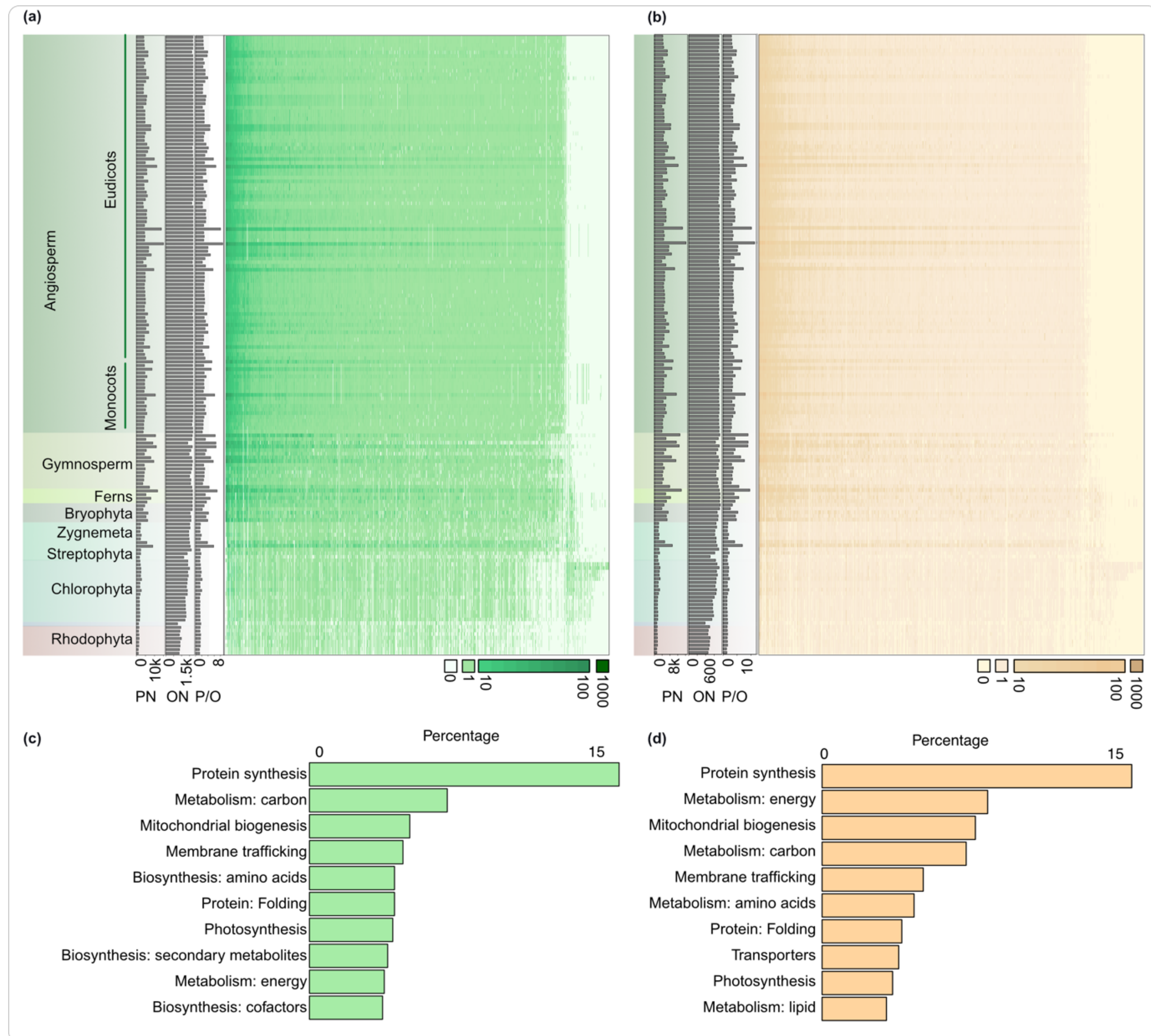

Fig. 2: Mitochondrial and plastid orthogroups across archaeplastidal species. Distribution of plastid (POGs, a) and mitochondrial orthogroups (MOGs, b). Protein copy numbers within each POG or MOG across species is shown in the heat-map as per the key on the bottom right of the heatmaps. Horizontal bars on the left side of the heatmaps show the total protein numbers (PN) likely localised to organelles, total POG or MOG numbers (ON) and distribution of protein number per OG (P/O) for a given species. Major functional categories of POGs and MOGs in (c) and (d), respectively.

The increased number of POGs and MOGs in the green lineage is explained by a combination of two phenomena: (a) new gains by the green ancestor, and (b) secondary losses at the origin of rhodophytes [39]. We used ancestral state reconstruction (ASR) to resolve between these two possibilities. The branching order of the archaeplastidal lineages remains challenging [40], as sometimes glaucophytes [41] and sometimes rhodophytes come out as the sister to the other remaining archaeplastidal lineages [4,42]. Hence, we inferred Archaeplastida phylogeny for both these scenarios (Fig. S2-3) and the ASR outcomes discussed are consistent with both (Fig. S4-6). The result suggests that the plastid proteome of the last common ancestor of Archaeplastida united 1150-1200 POGs (Fig. 3A, Fig. S4, Table S3). This inferred proteome witnessed significant gains of protein families at the emergence of the green ancestor (and later speciation). 50% of these newly gained POGs could be functionally annotated (Fig. 3A, Table S4), showing that the emergence of the green lineage was accompanied by an increase in the recruitment of photosynthesis- and metabolism-related POGs, while the transition to land (Z/E and

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

embryophyte ancestors) added metabolism related, as well as protein synthesis- and ubiquitin-related POGs to the toolkit. The mitochondrial proteome followed a qualitatively similar trend of expansion (Fig. 3B, Fig. S5, Table S5-6). 524 MOGs trace back to the archaeplastidal ancestor, while 632 MOGs were identified at the root of angiosperms. Around 50% of the newly gained MOGs could be functionally annotated, showing that the chloroplastidal gains contribute to carbon metabolism, protein synthesis and membrane trafficking. Terrestrialization also witnessed a similar gain of MOGs, most of which function in metabolism as well as mitochondrial biogenesis.

In summary, across plant species, plastid and mitochondrial proteomes have gained a significant number of protein families reflecting the dynamic nature of organellar proteomes post-endosymbiosis [43,44]. A closer look at the function of the newly gained organelle proteins shows a wide variety of functions (e.g. lipid and carbon metabolism, information processing, development and division of organelles).

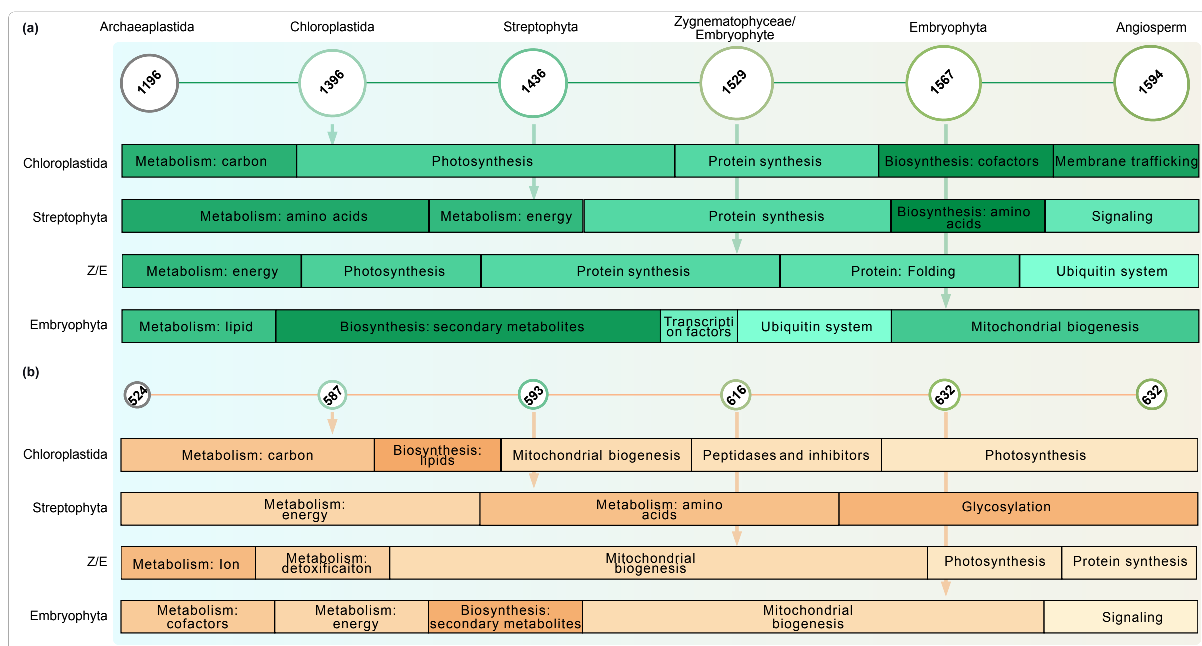

Fig. 3: Evolution of organelle proteomes in Archaeplastida. Gains in plastid (POGs, Fig. 3A) and mitochondrial orthogroups (MOGs, Fig. 3B) across the main nodes of archaeplastidal evolution. Each circle represents an ancestor, with predicted number of protein families present in that ancestor shown in the circle (also represented by the size of circles). Major gains were in the ancestor Chloroplastida, the common ancestor of Zygnematophyceae-embryophytes (Z/E) and embryophyte. Their functions are shown in the pie chart below the ancestors. The numbers of OGs across all ancestor, on a phylogenetic tree, is shown in Fig. S4-5.

Improved RNA processing and photosynthetic adaptability

RNA editing intercepts the linear information flow from mRNA to protein and is crucial for organelles to function [45,46]. Two main domains, the PPR and mTERF domain, are associated with RNA editing. We first screened for organelle orthogroups containing either of these two domains in at least 60% of all proteins within each respective orthogroup (Fig. S1C). A total of 50 POGs and 20 MOGs were found. More than 80% of them were restricted to embryophytes, only few were present in some algae (Fig. 4). A closer look revealed that most of the algal homologues lacked PPR and mTERF domains and they are hence unlikely true orthologues. More generally, this shows that any detailed interpretation regarding an inferred orthogroup's function should be supported by screening for functionally relevant

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

domains. True PPR or mTERF domain-containing RNA-editing proteins increased significantly in number by recruiting new orthogroups, also through adding the two domains to proteins that did not contain these in their algal ancestor. A presence-absence pattern also shows that >90% of proteins containing PPR/mTERF domains are present only in the land plants, except Chara and Klebsormidium (Fig. 4B). These proteins include, but are not limited to, OTP51 and SOT5 (present in embryophytes and Chara) as well as SOT1, SVR7, THA8, PDM4 (present only in embryophytes; Fig. S7). Target transcripts of these RNA editing factors point to the synthesis and assembly of photosynthesis-related proteins and to proteins of the thylakoid membrane (Fig. 6B). Likewise, mTERFs, which are crucial for plastid and leaf development, are also uniquely expanded in the terrestrial clade with examples of protein re-targeting across organelles [47]. The dual targeted (plastid and mitochondrion) mTERF6, unique to the land plants (Fig. S7) and the streptophyte alga *Klebsormidium*, takes part in retrograde signalling to the nucleus via ABA and imparts abiotic stress tolerance [48]. Overall, RNA editing across plants has undergone major changes and has a significant impact on photosynthesis, improvement of which was key to colonising land (Fig. 6B).

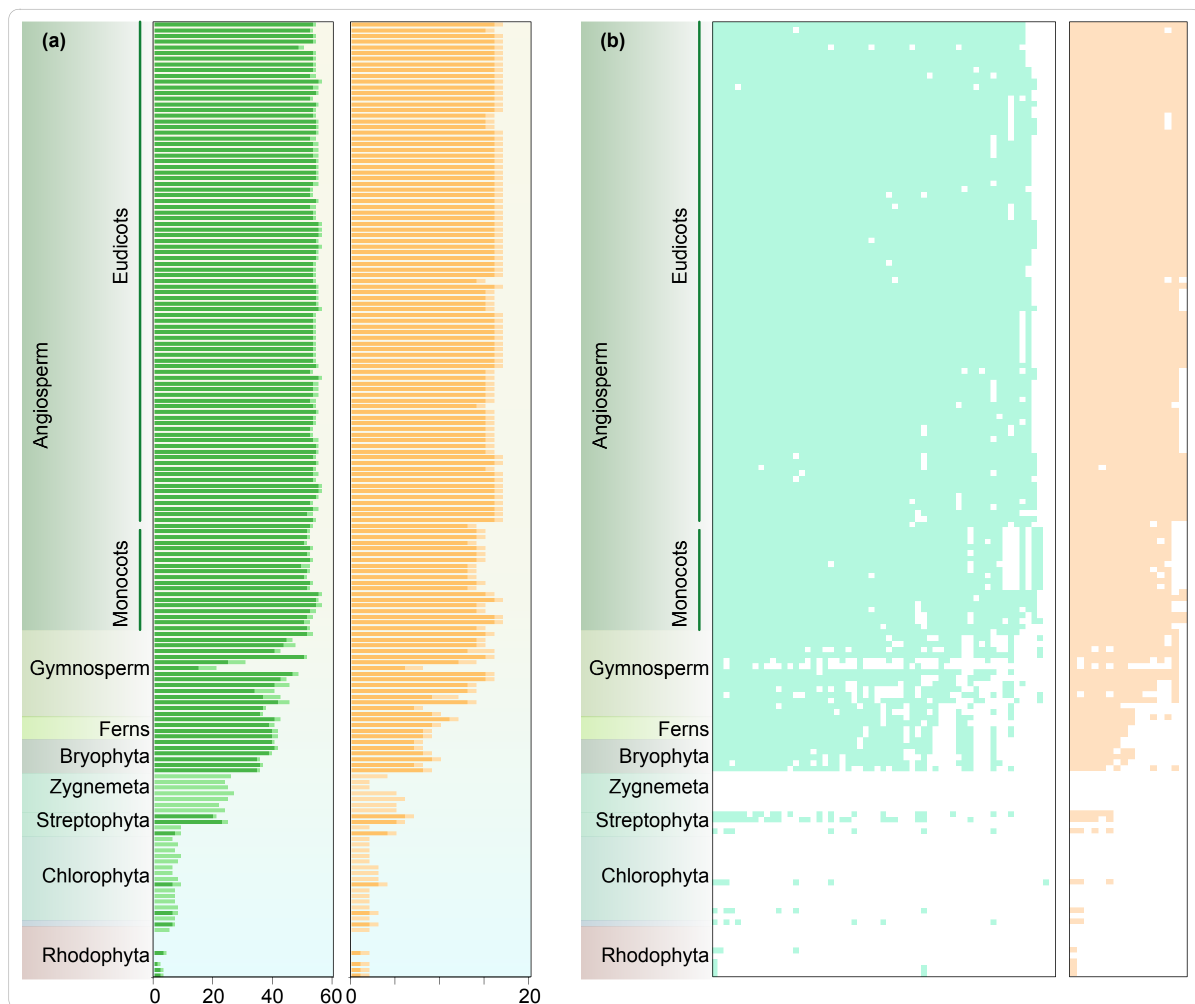

Fig. 4: Recruitment of PPR and mTERF domains in organelle proteins. (a) Number of POGs (left) and MOGs (right) where at least one protein contains a PPR/mTERF domain is shown in bars with dark shades of colors. Total number of orthogroups (regardless of presence or absence of PPR/mTERF domain in that particular spices) is shown in lighter shade. It shows the presence of the orthogroups in question in algae, but that they only later, in embryophytes, obtained PPR/mTERF domains. (b) Each cell represents an orthogroups and a coloured cell indicates the presence of a PPR or mTERF domain in the protein family (column) of a respective species (rows).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Adaptation to the terrestrial habitat and changes in plastid biochemistry

Main terrestrial stresses include draught, high (UV)light and swift temperature changes. Cutin and suberin, two of the most abundant lipid polymers on Earth [49], evolved as one countermeasure [50]. We find that cutin and suberin evolution was enabled by the embryophytic recruitment of an organelle-specific GPAT (Glycerol-3-phosphate acyltransferases) family (Fig. 5), which includes GPAT1 (mitochondrial), GPAT 4,6 and 8 of the endoplasmic reticulum [51,52]. Trafficking of these lipids across organelles was made possible by a dual targeted TGD4 [53] that appeared in the Z/E ancestor (Fig. 5). Acyl carrier thioesterases, responsible for the export of fatty acids from the plastid, acyl carrier protein desaturases (ACP-desaturase) and acyl-carrier proteins co-factors of fatty acid bio-synthesis were uniquely retained and expanded in the green lineage (Fig. S7). Duplication and divergence of ACP desaturases in embryo- and spermatophytes (Fig. S7) played an important role in regulating lipid composition shifts in response to temperature and drought, the regulation of seed oil content and development [54]. Likewise, acyl-carrier proteins also increased in copy number (Fig. S7) and adapted towards a light-induced expression and regulation of the seed fatty acid content [55,56]. These changes in organelle lipid synthesis and trafficking underpinned embryophyte adaptations to cope with draught and high temperature stress (wax biosynthesis, deposition on the layer of leaves and cuticle development), as well as seed development and germination in spermatophytes (Fig. 6D).

Changes in starch metabolism mostly pertain to its regulation. ADP-glucose pyrophoshorylase (AGPase), an enzyme responsible for a rate-limiting step in starch metabolism, is uniquely retained in the green lineage and increased in copy number in streptophytes (Fig. S7). AGPases diverged to regulate starch metabolism under osmotic and light stress, as well as the differential regulation of starch synthesis and degradation [57–61]. Another key regulatory enzyme, PGI (phosphoglucose isomerase) evolved a distinct family (PGI1) in Zygnematophyceae (Fig. S7). It likely kickstarted the regulation of starch metabolism at the water-to-land interface and later assumed significant roles in embryophyte fatty acid content regulation and the yield of seeds [62]. PTST3 also emerged around the time of terrestrialization (Fig. S7), which evolved to regulate starch synthesis with significant impact on plastid development [63]. In contrast to the flow of carbon through glycolysis, GSM2, which originated in streptophytes (Fig. 7), shunts carbon towards the pentose-phosphate pathway and protects plastids from oxidative stress in *Arabidopsis* [64].

Emergence of sophisticated antero- and retrograde communication cascades

Communication across compartments is critical for a concerted response to environmental stimuli. Plastids are key environmental sensors that interconnect cellular metabolism with physiological requirements and stress responses, and terrestrial stressors are key triggers of plastid-to-nucleus retrograde signalling [12,13,21]. We screened for the origin and diversification of EXECUTOR, WHIRLY and SVR4, all critical organellar components of retrograde signalling. The EXECUTOR, key to regulating retrograde signalling, oxygen and light stress regulation [65–67], originated in the ancestor of the Chloroplastida and so did WHIRLY (Fig. 5); the latter underwent copy number expansion in embryophytes and likely loss in some bryophytes (Fig. S7). Divergence of these copies led to a localisation across multiple organelles and today they are crucial for maintaining functional respiration, photosynthesis and the response of mitochondria and plastids to biotic and abiotic stresses [68–70]. These two key players, EXECUTOR and WHIRLY, accompanied the origin of the Chloroplastida and additional paralogs with specific functions in the main green lineages likely aided in the conquer of land (Fig. 6B).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

SVR4, a dual targeted (plastid and nucleus) recruitment of embryophytes (Fig. 5), communicates required gene expression changes needed for light-induced plastid development, thylakoid stacking and thermomorphogenesis [71,72]. In combination, this facilitates light-induced photomorphogenesis, a process key for surviving on land. An increase in the complexity of retrograde signaling was a precursor for terrestrialization [12], for instance via innovations such as the 3′-phosphoadenosine-5′-phosphate family, which facilitated the emergence of stomatal closing in land plants [73]. The recruitment and diversification of the proteins we highlight, were quintessential for responding to two major stressors more pronounced and more rapidly changing on land than in water: light and temperature (Fig. 6B).

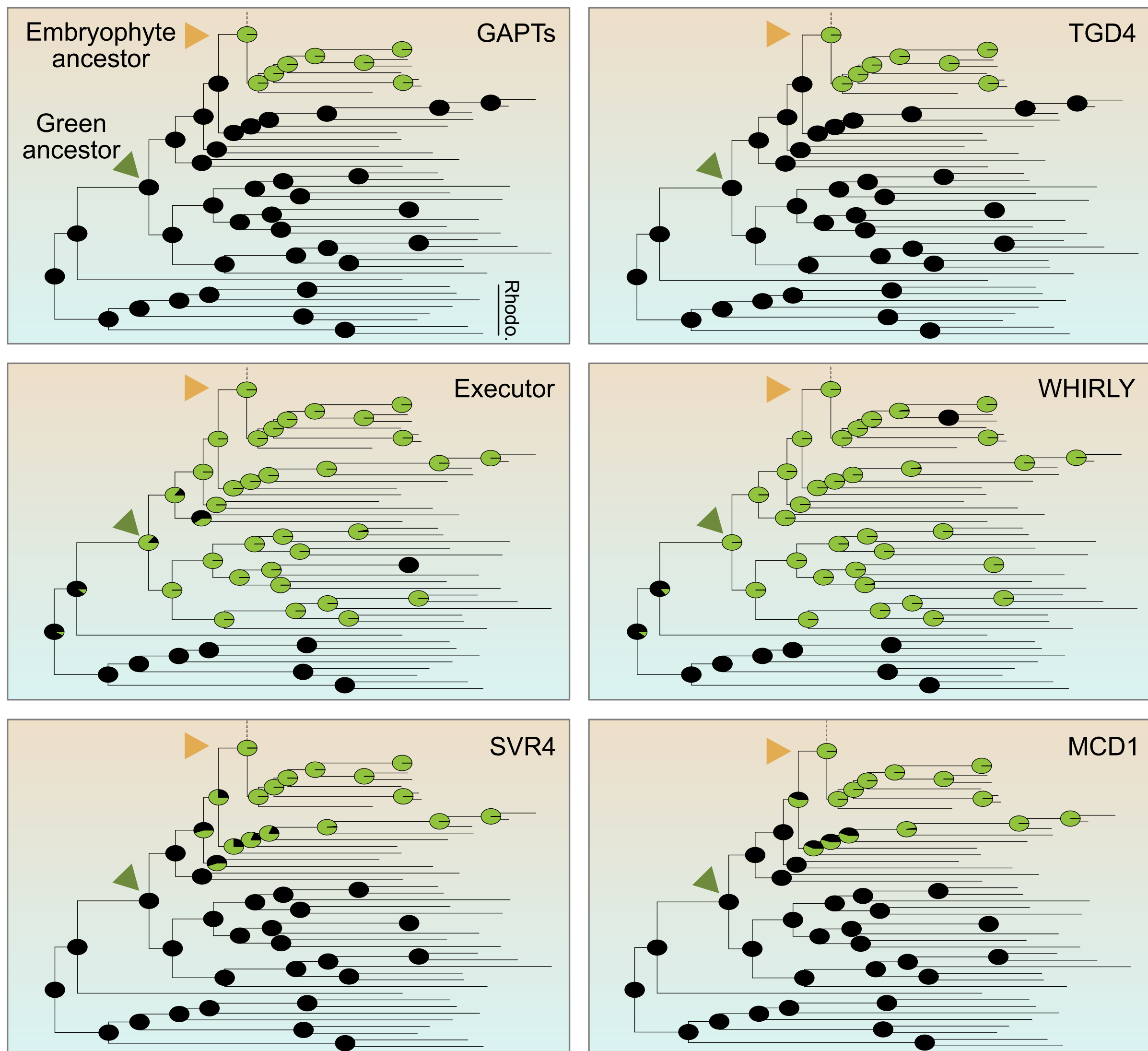

Fig. 5: Origins of key proteins involved in metabolism, communication and development. Ancestor state reconstruction (ASR) for selected lipid metabolism (GAPT and TGD4), retrograde signalling (Executor and Whirly), plastid development (SVR4) and division (MCD1) related proteins. The pie charts at each node on the tree represent probability of presence (green) or absence (black) of a protein family in that node.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Recruitment of new proteins and changes in organelle development

The coordination of tissue and plastid development is linked to ensure an appropriate response to biotic and abiotic factors, especially in morphologically complex plants [74–76]. Polyplastidy is a land plant trait [77] and known molecular determinants include MinD, MinE, ARC3 and the FtsZ proteins [16,75]. Our data supports that MULTIPLE CHLOROPLAST DIVISION SITE 1 (MCD1), a core component of the plastid division machinery [78], originated in the ancestral embryophyte (Fig. 5). The cotyledon chloroplast biogenesis factor CYO1 and the transcriptionally active chromosome factor 7 (TAC7) are core components of thylakoid biogenesis and the plastid translation machinery, respectively. Both originated in the streptophyte ancestor (Fig. S7) and, in *Arabidopsis*, play key roles in chloroplast, cotyledon, thylakoid and leaf development [79–81]. Lastly, CRUMPLED LEAF (CRL), a protein residing in the outer plastid membrane, emerged during terrestrialization, too (Fig. S7), likely for regulating plastid division and securing correct plastid inheritance during embryogenesis [82,83].

Crucial for plastid biogenesis, especially in light of an expanding proteome, is the import of proteins. The membrane GTPase TOC159 is essential for chloroplast biogenesis via the selective recognition and import of the photosynthetic proteins [84] and is unique to the green lineage (Fig. S7). The membrane recruitment of this protein requires TOC75, of which a special variant evolved in the green ancestor after the duplication of OEP80 [14]. The copy number of TOC159 expanded from the Zygnematophyceae onwards (Fig. S7), hinting at its functional diversification. Unlike in the chlorophyte alga *Chlamydomonas*, land plant TOC159 homologues possess an N-terminal acidic domain that gets phosphorylated to alter substrate specificity [84,85]. Furthermore, TOC159, along with TOC132 and TOC120, play important roles in regulating plastid lipid synthesis and membrane fluidity [86–88]. Further on the course of evolution, the J-domain-containing protein TOC12 [89] was likely recruited in the ancestral embryophyte for supporting the import machinery at the intermembrane space (Fig. S7). The terrestrial habitat demands a highly efficient and fluid import of proteins, for example upon high light and other abiotic stresses [14,90]. The expansion of the TOC/TIC system in the embryophyte ancestor reflects how the organelle dealt with an ever-increasing diversity of substrates that were required to be processed.

Discussion

The conquest of land by a streptophyte alga and the propagation of plants (Fig. 6A) was pivotal in the transformation of the terrestrial habitat and it laid the foundation for the concomitant colonization of land by animals [1,2]. Throughout the hundreds of millions of years of this endeavor, both plant organelles of endosymbiotic origin underwent a multitude of molecular adaptations, hereby evolving into the plastid and mitochondrion of modern plants. We identified 31,650 protein families unique to the green lineage, approximately 50% of which are unique to embryophytes. It demonstrates an expansion and divergence of protein families around plant terrestrialization, in line with a recent study that identified around 10,000 duplications at the birth of embryophytes [91].

Expansion of proteins families is evident in both organellar proteomes at the origin of the green lineage itself and at the water-to-land transition. The gain of protein families at the origin of the green lineage needs to be treated with caution due to the documented genetic bottleneck that characterizes rhodophyte origin [92–96] and the sparse availability of glaucophyte genome data. Some of the newly recruited protein families at the origin of the green lineage might rather be explained by a loss in rhodophytes and a retention in the chloroplastidal ancestor instead of a gain. Regardless, this has a little bearing on the biological significance of a given protein family with respect to the overall increase in complexity of organelle biology – both concerning the variety as well as the number of proteins targeted to plastids

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and mitochondria – throughout streptophyte evolution that affected the organelles metabolic, informational and developmental complexity, and facilitating the conquest of land (Fig. 6).

Changes in organelle lipid biochemistry contributed to one of the key adaptations in land plants: cuticles. Land plant GPATs (Glycerol-3-phosphate acyltransferases, crucial to lipid synthesis for cutin and suberin) contribute to increased hydrophobicity and water retention in embryophytes [50] and show activity distinct from algae [97,98]. Our analyses underpin origins of organelle specific GAPTs (GAPT 1,4,6,8) to the ancestor embryophyte, of which, deleting GPAT4 and GPAT8 distorts cuticles and increases water loss by several fold [51,52]. In parallel, lipid trafficking was mediated by the recruitment or divergence of proteins such as TGD4 and acyl carrier thioesterases, which contributed to wax biosynthesis and deposition on leaves, cuticle development, thylakoid membrane stacking [53], seed development and germination [54]. As for starch metabolism, the archaeplastidal ancestor likely stored starch in the cytosol [99], but the red and green lineage experienced different fates from there on. Rhodophytes continued to store starch in the cytosol in the form of Floridean starch [100], while in the green lineage, particularly in complex plants, more localized control of starch synthesis and degradation was facilitated by changes in regulatory proteins (eg AGPase). Together, organelle metabolism evolved to serve key roles in the synthesis, regulation and trafficking of lipids involved in wax coating to prevent water loss in the land plant ancestor, as well as synthesis and storage of starch (Fig. 6D).

RNA processing and editing is a crucial component of information processing and overall functionality of plant organelles [45,46]. Changes in RNA editing are evident from the origin of the green lineage itself, where RNAse-P (tRNA maturation) was replaced by protein only RNAse P or PROPs [101,102]. Subsequent expansion of PROPs in embryophytes (Fig. S7) led to organelle-localised copies of which some are essential for maintaining organelle morphology, functions and plant viability [103]. RNA editing of C to U is not found in algae, however, and editing sites in embryophytes are unlike those of any other eukaryote, suggesting they emerged independently [46]. Of the many RNA-editing proteins we find that were gained during terrestrialization, known targets are transcripts involved in photosynthesis and stress tolerance-related transcripts, both key to colonising the land (Fig. 6B). For instance, THA8, PDM4, SVR7 and SOT1 edit transcripts such as ycf2 and ycf3, and contribute to thylakoid development and biogenesis [104], the generation of photosynthetic complex proteins, granal stacking, embryo [105] and plastid development [106,107]. OTP51 and SOT5 edit transcripts related to chlorophyll synthesis, photosynthesis and thylakoid membranes (ycf3, TRNK and RPL2) [108–110], whereas DOG1 is important to high temperatures response and chloroplast development[111]. This elaborate RNA processing in organelles, especially plastids, appears to serve photosynthesis (and thylakoid) related transcripts. It is feasible that by benefitting photosynthesis, organelle RNA editing continued to be positively selected for and was expanded.

One adaptation towards efficient photosynthesis, where RNA editing also plays a key role, are grana [74]. The evolutionary origins of grana remain elusive, along with the underlying developmental pathways involved in regulating its formation and maintenance [74,112,113]. Grana are observed in embryophytes and some Zygnematophyceae (e.g. the *Cosmarium* genus) [114], but not chlorophytes such as *Chlamydomonas* [115]. We noticed a patchy distribution of grana morphology associated proteins such as CURT1, RIQ1 and RIQ2 (Fig. S7), with both RIQs being present in all streptophytes and some chlorophytes but excluding *Chlamydomonas*. In light of the many key adaptations in Zygnematophyceae discussed here and elsewhere [11,116], we speculate that grana originated in streptophytes and were important in conquering land through photosynthesis optimization, in particular with respect to photosystem repair and the separation of the photosystems and the ATP synthase [117,118].

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

This expansion of an organelle proteome necessitates improving the capacity to import proteins. Changes in some import receptors within the green lineage and in targeting sequences at its origins are known, with phosphorylation likely emerging as a key regulator for sorting the newly expanded proteome differentially to plastid and mitochondria (Fig. 6C) [14,119]. Despite such adaptations, protein sorting is never perfect and some mistargeting might be positively selected for. A regulated distribution of newly recruited proteins (e.g. WHIRLY, TGD4, mTERF6; Fig. 6B) to multiple organelles (with distinct organellar functions) hints at adaptive values of this apparent mis-sorting. How many of newly recruited proteins get 'mis-sorted' owing to biological adaptability versus stochasticity remains to be explored together with obtaining a more comprehensive picture of (regulatory) mechanisms associated with sorting in general.

Embryophyte cells target proteins not to a single plastid, but many simultaneously. The presence of multiple plastids per cell, polyplastidy, evolved in embryophytes, likely through changes in plastid fission and a decoupling of organelle fission from the cell cycle [15,16]. We find that MCD1 a core regulator of the plastid division proteins FtsZ2 and ARC3, emerged in the embryophyte ancestor, which corroborates the idea of a mono- to polyplastidy switch during terrestrialization [16,77,120,121]. A change in the copy number of plastids also requires a mechanism that maintains a functional organelle to cell volume ratio and resource allocation (Fig. 6C). The REDUCED CHLOROPLAST COVERAGE (REC) protein is involved in such a mechanism in *Arabidopsis* [122] and the phylogenetically related protein FRIENDLY regulates the distribution of mitochondria, also in plants and non-photosynthetic organisms [123,124]. REC and FRIENDLY share almost all of their domains. How they exactly function and differentiate between the two organelles remains elusive. From what we can tell, FRIENDLY emerged during eukaryogenesis and the origin of mitochondria. REC we trace back to the streptophyte ancestor (Fig. S7) and after a likely duplication event of FRIENDLY. We speculate that the origin of REC helped to establish polyplastidy, which itself supports larger body plans and the diversification of different plastid types [15]. Lastly, an increase in organelle copy number also requires an overall increase in the capacity to synthesize proteins. The largest fraction of organelle proteins operate in tRNA, amino acid and ribosomal biosynthesis and undergird the biosynthetic capacity of organelles, an adaptation strategy akin to their bacterial ancestor [125,126].

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

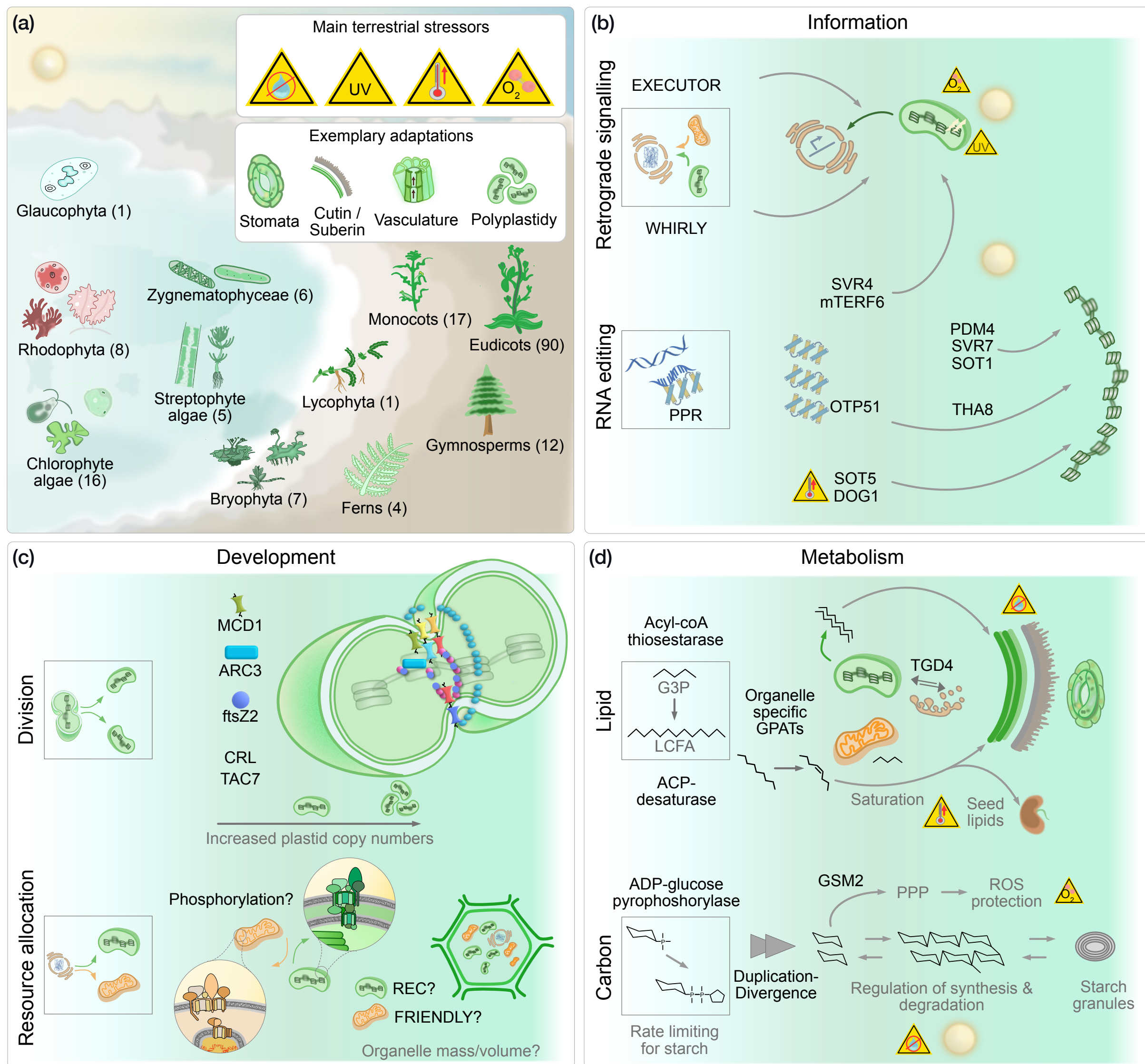

Fig. 6: The global greening and endosymbiotic organelles. (a) After the endosymbiotic origin of the plastid, three aboriginal lineages emerged that form the Archaeplastida: the glaucophytes, rhodophytes and chlorophytes. From the latter, streptophyte algae evolved, including the zygnematophyceae, that represent the algal sister clade to land plants (embryophytes). Abiotic stresses encountered during terrestrialization (water scarcity, high UV, swiftly altering temperatures and higher levels of O_2) selected for adaptive features such as stomata and a cutin layer. The numbers in parenthesis indicate the number of genomes from each major group we screened. Recruitment of new organelle proteins improved three key aspects of organelle biology in light of terrestrialization: (b) information processing, (c) development and (d) metabolism. Details for each tile are discussed in the main text.

The accommodation of the early mitochondrial endosymbiont was key to the emergence of the endomembrane system and complex eukaryotic traits including mito- and autophagy [127–129]. Our analyses show that the integration of a subsequent endosymbiont, the plastid, coincided with the emergence of proteins that work for the endomembrane system. Particularly, changes in the ubiquitin system were prominent around terrestrialization, when polyplastidy in the green lineage also emerged (Fig. 3). Ubiquitination is key to proteosome-mediated degradation and is performed chiefly by a family of proteins called E3 ubiquitin ligases which are important in land plants also for

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

photomorphogenesis[130]. RING (Really interesting new gene) E3 ligases contribute to growth, development and stress response via also mediating protein-protein interactions [131–134]. We trace a number of RING finger (and related) proteins to terrestrialization (Fig. S7) that include, but are not limited to, *DAL1* and *DAL2* (for *Drosophila* DIAP1 like 1 and 2), KEG (Keep on going), and NIP1 and NIP2. *DAL1* and *DAL2* play a key role in regulation of programmed cell death [135], peroxisome and chloroplast biogenesis [136–138]. KEG contributes to stress mitigation [139,140], while NIP1 and NIP2 play a role in plastid development by docking plastid RNA polymerase to the thylakoid membrane [141]. The regulated degradation of plastids and other changes in the endomembrane system were key to housing multiple plastids per cell and we find many more recruitments broadly affiliated with the endomembrane system, with no well characterised functions. Exploring the functions of these proteins will add valuable insights into the cell biological changes that endosymbiosis stipulates.

In closing, although experimentally reported plant plastid and mitochondrial proteomes are scarce, we were able to generate a first comprehensive molecular atlas of the changes of plastid and mitochondrial protein families in the evolution of the green lineage. Ancestral state reconstruction (ASR) allows to map the organelle transformations that facilitated the major transitions such as terrestrialization and which will improve with every new proteome that is added. By inferring plastid and mitochondrial proteomes for 168 species, we set testable expectations for new proteomes to come and provide a solid database, where origins and across species orthologues of any known (organelle) protein can be searched easily (Table S2B-C). We identify numerous mitochondrial protein recruitments, the physiological roles and adaptive values of which help to better understand plant mitochondrial biology. For plastid proteins, we infer their functions and physiological importance based on the extensively studied *Arabidopsis* system. Utilizing an advanced orthology search technique [37], we postulate that orthologues of *Arabidopsis* are likely to exhibit similar functions in other species. Our methodologically robust approach offers various changes in evolution, associated in particular with terrestrialization, that can now be experimentally explored across selected models and with a focus on less-well studied streptophyte algal and bryophyte species [142,143].

Conclusions

Endosymbiotic organelles have a distinct place in the evolutionary tapestry of life. Through the combination of organelle proteome data and phylogeny, we trace the evolution of mitochondria and plastids over a span of a billion years of plant evolution by inferring their proteomes for over a hundred Archaeplastida species. Our comprehensive molecular atlas identifies main changes in their metabolism, communication, information processing and biogenesis. Key adaptations in plant organelles fostered the emergence of wax and cutin (see organelle lipid synthesis and transport), improved the photosynthetic yield (see organelle RNA editing and grana) and the response to abiotic stressors (see inter-organelle communication), and mediated the transition from mono- to polyplastidy (see division and volume control). By connecting the molecular adaptations of mitochondria and plastids to macroevolutionary trends, we show how important changes in organelles of endosymbiotic origin were for the speciation that gave rise to the Chloroplastida and later the origin of land plants from a charophyte algal ancestor.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Material and Methods

Curation of green orthogroups (GOGs). Input protein sequences from 686 proteomes (from KEGG and Phytozome, Table S1A) were clustered using Orthofinder version 2.5.4 [37], after all vs all blasts were conducted (E-value cutoff 10e-10) using diamond blast version2.011 [35]. From orthogroups (OGs) recovered, OGs with at least 3 Chloroplastida species green species and less than 3 species other than Chloroplastida were annotated as green orthogroup (GOGs). Schematic in Fig. S1A.

Curation of plastid and mitochondria orthogroups (POGs and MOGs). 4,915,150 proteins from 168 Archaeplastida species were clustered using Orthofinder as described above. Orthogroups that contained at least one experimentally verified organelle protein from any one of the four experimentally verified organelle proteome of *C. reinhardtii* [144], *P. patens* [145] , *Z. mays* [146], *A. thaliana* [146], were annotated as organelle (plastid and mitochondria) orthogroups. Schematic in Fig. S1B.

Functional annotation of orthogroups. The source of >90% species was Kyoto Encyclopedia of Genes and Genomes (KEGG), which included KEGG orthology identification (KOID) for protein sequences. For all proteins within each GOG, KOIDs were retrieved and the most frequent KOID (i.e. majority rule) was annotated to each GOG (Fig. S1C). From the assigned KOIDs, their KO BRITE functional category was assigned to each GOG. KOIDs for POGs and MOGs were retrieved the same way. For each KOID, the pathway names and BRITE categories at various level of resolutions were used for assigning functional categories manually to each OG. Manual assignment was necessary since BRITE names included a large fraction of categories such as 'enzymes' and 'exosomes'. These were either not very informative or were misleading as many of 'exosome' annotated proteins took part in protein synthesis or folding. Lastly, for OGs or proteins discussed with respect to their physiological relevance, the functions were retrieved from the literature (cited in the text).

Inference of ancestral states. A phylogeny of Archaeplastidal species was inferred based on all genes conserved in all species, using 'Species tree inference from all genes (STAG)' method [147], as a part of orthofinder analysis. STAG infers a species tree by taking greedy consensus of gene trees from each protein family (including that of multigene families). This phylogeny was rooted using minimal ancestral deviation [148] which places Rhodophyta as the sister to all others. Independently, the same unrooted phylogeny was manually rooted using FigTree (v1.4.4) [149] such that Glaucophyta were at the base. Ancestor state of presence and absence of organelle protein families across nodes, were inferred using Phytool [150] package 0.7.80. Based on character state at the tips of the tree, Phytool inferred Bayesian posterior probabilities under a single rate model [151,152] of the character state across nodes of the tree. All OGs that were present in major ancestors of plant groups with probability higher than 0.75 and absent in the preceding ancestor, were considered as newly recruited in that lineage. OGs or proteins discussed with respect to its physiological role in a given clade, their absence outside the group was verified in our copy number database as well as on homologue database available on TAIR.

Searching for potential RNA editing POGs and MOGs. Hidden Markov models (HMM) of PPR and mTERF domains were downloaded from pFAM [153] with the IDs: PF01535, PF12854, PF13041, PF13812, PF02536. Each of these HMMS was used as a query to search against the full sequences of all proteins within each POG and MOG. If a given OG had more than 60% of individual proteins containing PPR or mTERF, the OG was annotated as RNA editing OG. Origin of such OGs were traced using ASR as described above.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Author contributions

PKR: Conceptualization; Experimental design; Methodology; Investigation, Data curation; Formal analysis; Visualization; Writing - original draft, review and editing. AIM: Methodology, Investigation, Formal analysis; Writing - original draft, review and editing SBG: Conceptualization; Project administration; Funding acquisition; Resources; Supervision; Visualization; Writing - original draft, review and editing.

Funding

We thank the Deutsche Forschungsgemeinschaft (SFB 1208-2672 05415 and SPP2237–440043394) and the Moore and Simons Initiative grant (9743) for financial support.

Acknowledgments

We acknowledge support from the high-performance computing cluster (HILBERT) at the Heinrich Heine University Düsseldorf and thank Michael Knopp from the Institute of Molecular Evolution at the Heinrich Heine University Düsseldorf.

References

1. Schreiber M, Rensing SA, Gould SB. The greening ashore. Trends in Plant Science. 2022. doi:10.1016/j.tplants.2022.05.005

2. Bowles AMC, Williamson CJ, Williams TA, Lenton TM, Donoghue PCJ. The origin and early evolution of plants. Trends in Plant Science. 2023;28: 312–329. doi:10.1016/j.tplants.2022.09.009

3. Li L, Wang S, Wang H, Sahu SK, Marin B, Li H, et al. The genome of Prasinoderma coloniale unveils the existence of a third phylum within green plants. Nature Ecology & Evolution. 2020;4: 1220–1231. doi:10.1038/s41559-020-1221-7

4. Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, et al. One thousand plant transcriptomes and the phylogenomics of green plants. Nature. 2019;574: 679–685. doi:10.1038/s41586-019-1693-2

5. Li X, Hou Z, Xu C, Shi X, Yang L, Lewis LA, et al. Large Phylogenomic Data sets Reveal Deep Relationships and Trait Evolution in Chlorophyte Green Algae. Genome Biology and Evolution. 2021;13: evab101–evab101. doi:10.1093/gbe/evab101

6. Puginier C, Keller J, Delaux P-M. Plant–microbe interactions that have impacted plant terrestrializations. Plant Physiology. 2022;190: 72–84. doi:10.1093/plphys/kiac258

7. de Vries S, Fürst-Jansen JMR, Irisarri I, Dhabalia Ashok A, Ischebeck T, Feussner K, et al. The evolution of the phenylpropanoid pathway entailed pronounced radiations and divergences of enzyme families. Plant J. 2021;107: 975–1002. doi:10.1111/tpj.15387

8. Bowman JL. Stomata: Active Portals for Flourishing on Land. Current Biology. 2011;21: R540–R541. doi:https://doi.org/10.1016/j.cub.2011.06.021

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

9. Kong L, Liu Y, Zhi P, Wang X, Xu B, Gong Z, et al. Origins and Evolution of Cuticle Biosynthetic Machinery in Land Plants. Plant Physiol. 2020;184: 1998–2010. doi:10.1104/pp.20.00913

10. Davies KM, Jibran R, Zhou Y, Albert NW, Brummell DA, Jordan BR, et al. The Evolution of Flavonoid Biosynthesis: A Bryophyte Perspective. Frontiers in Plant Science. 2020;11. doi:10.3389/fpls.2020.00007

11. Cheng S, Xian W, Fu Y, Marin B, Keller J, Wu T, et al. Genomes of Subaerial Zygnematophyceae Provide Insights into Land Plant Evolution. Cell. 2019. doi:10.1016/j.cell.2019.10.019

12. de Vries J, Stanton A, Archibald JM, Gould SB. Streptophyte Terrestrialization in Light of Plastid Evolution. Trends in Plant Science. 2016. doi:10.1016/j.tplants.2016.01.021

13. de Vries J, Archibald JM. Plant evolution: landmarks on the path to terrestrial life. New Phytologist. 2018. doi:10.1111/nph.14975

14. Knopp M, Garg SG, Handrich M, Gould SB. Major Changes in Plastid Protein Import and the Origin of the Chloroplastida. iScience. 2020. doi:10.1016/j.isci.2020.100896

15. de Vries J, Gould SB. The monoplastidic bottleneck in algae and plant evolution. Journal of Cell Science. 2018. doi:10.1242/jcs.203414

16. MacLeod AI, Raval PK, Stockhorst S, Knopp MR, Frangedakis E, Gould SB. Loss of Plastid Developmental Genes Coincides With a Reversion to Monoplastidy in Hornworts. Frontiers in Plant Science. 2022;13. Available: https://www.frontiersin.org/article/10.3389/fpls.2022.863076

17. Shimizu T, Kacprzak SM, Mochizuki N, Nagatani A, Watanabe S, Shimada T, et al. The retrograde signaling protein GUN1 regulates tetrapyrrole biosynthesis. Proceedings of the National Academy of Sciences. 2019;116: 24900–24906. doi:10.1073/pnas.1911251116

18. Bar-On YM, Phillips R, Milo R. The biomass distribution on Earth. Proceedings of the National Academy of Sciences. 2018;115: 6506–6511. doi:10.1073/pnas.1711842115

19. Pérez-Sancho J, Vanneste S, Lee E, McFarlane HE, Esteban del Valle A, Valpuesta V, et al. The Arabidopsis Synaptotagmin1 Is Enriched in Endoplasmic Reticulum-Plasma Membrane Contact Sites and Confers Cellular Resistance to Mechanical Stresses. Plant Physiology. 2015;168: 132–143. doi:10.1104/pp.15.00260

20. Gao H, Metz J, Teanby NA, Ward AD, Botchway SW, Coles B, et al. In Vivo Quantification of Peroxisome Tethering to Chloroplasts in Tobacco Epidermal Cells Using Optical Tweezers. Plant Physiology. 2016;170: 263–272. doi:10.1104/pp.15.01529

21. Wang Y, Selinski J, Mao C, Zhu Y, Berkowitz O, Whelan J. Linking mitochondrial and chloroplast retrograde signalling in plants. Philosophical Transactions of the Royal Society B: Biological Sciences. 2020;375: 20190410–20190410. doi:10.1098/rstb.2019.0410

22. Møller IM, Rasmusson AG, Van Aken O. Plant mitochondria – past, present and future. The Plant Journal. 2021;108: 912–959. doi:10.1111/tpj.15495

23. Kanai R, Edwards GE. The Biochemistry of C4 Photosynthesis. C4 Plant Biology. 1999. doi:10.1016/b978-012614440-6/50004-5

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

24. Christin PA, Osborne CP. The evolutionary ecology of C4 plants. New Phytologist. 2014. doi:10.1111/nph.13033

25. Heinnickel ML, Grossman AR. The GreenCut: re-evaluation of physiological role of previously studied proteins and potential novel protein functions. Photosynthesis Research. 2013;116: 427–436. doi:10.1007/s11120-013-9882-6

26. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research. 2012;40: D1178–D1186. doi:10.1093/nar/gkr944

27. Hori K, Maruyama F, Fujisawa T, Togashi T, Yamamoto N, Seo M, et al. Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. Nature Communications. 2014. doi:10.1038/ncomms4978

28. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Research. 2016;44: D733–D745. doi:10.1093/nar/gkv1189

29. Bowman JL, Kohchi T, Yamato KT, Jenkins J, Shu S, Ishizaki K, et al. Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. Cell. 2017. doi:10.1016/j.cell.2017.09.030

30. Lang D, Ullrich KK, Murat F, Fuchs J, Jenkins J, Haas FB, et al. The Physcomitrella patens chromosome-scale assembly reveals moss genome structure and evolution. The Plant Journal. 2018;93: 515–533. doi:10.1111/tpj.13801

31. Nishiyama T, Sakayama H, de Vries J, Buschmann H, Saint-Marcoux D, Ullrich KK, et al. The Chara Genome: Secondary Complexity and Implications for Plant Terrestrialization. Cell. 2018. doi:10.1016/j.cell.2018.06.033

32. Li F-W, Nishiyama T, Waller M, Frangedakis E, Keller J, Li Z, et al. Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. Nature Plants. 2020;6: 259–272. doi:10.1038/s41477-020-0618-2

33. Wang S, Li L, Li H, Sahu SK, Wang H, Xu Y, et al. Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. Nature Plants. 2020. doi:10.1038/s41477-019-0560-3

34. Zhang J, Fu XX, Li RQ, Zhao X, Liu Y, Li MH, et al. The hornwort genome and early land plant evolution. Nature Plants. 2020. doi:10.1038/s41477-019-0588-4

35. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nature Methods. 2015;12: 59–60. doi:10.1038/nmeth.3176

36. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome biology. 2015;16: 157–157. doi:10.1186/s13059-015-0721-2

37. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology. 2019;20: 238–238. doi:10.1186/s13059-019-1832-y

38. Emms DM, Kelly S. SHOOT: phylogenetic gene search and ortholog inference. Genome Biology. 2022;23: 85–85. doi:10.1186/s13059-022-02652-8

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

39. Janouškovec J, Liu S-L, Martone PT, Carré W, Leblanc C, Collén J, et al. Evolution of Red Algal Plastid Genomes: Ancient Architectures, Introns, Horizontal Gene Transfer, and Taxonomic Utility of Plastid Markers. PLOS ONE. 2013;8: e59001. doi:10.1371/journal.pone.0059001

40. Strassert JFH, Irisarri I, Williams TA, Burki F. A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. Nature Communications. 2021;12: 1879. doi:10.1038/s41467-021-22044-z

41. Price DC, Goodenough UW, Roth R, Lee J-H, Kariyawasam T, Mutwil M, et al. Analysis of an improved Cyanophora paradoxa genome assembly. DNA Research. 2019;26: 287–299. doi:10.1093/dnares/dsz009

42. Wang S, Liang H, Xu Y, Li L, Wang H, Sahu DN, et al. Genome-wide analyses across Viridiplantae reveal the origin and diversification of small RNA pathway-related genes. Communications Biology. 2021;4: 412. doi:10.1038/s42003-021-01933-5

43. Keeling PJ. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. Annual Review of Plant Biology. 2013. doi:10.1146/annurev-arplant-050312-120144

44. Ku C, Nelson-Sathi S, Roettger M, Sousa FL, Lockhart PJ, Bryant D, et al. Endosymbiotic origin and differential loss of eukaryotic genes. Nature. 2015;524: 427–432. doi:10.1038/nature14963

45. Hao W, Liu G, Wang W, Shen W, Zhao Y, Sun J, et al. RNA Editing and Its Roles in Plant Organelles. Frontiers in Genetics. 2021;12. Available: https://www.frontiersin.org/articles/10.3389/fgene.2021.757109

46. Small ID, Schallenberg-Rüdinger M, Takenaka M, Mireau H, Ostersetzer-Biran O. Plant organellar RNA editing: what 30 years of research has revealed. Plant J. 2020;101: 1040–1056. doi:10.1111/tpj.14578

47. Robles P, Quesada V. Research Progress in the Molecular Functions of Plant mTERF Proteins. Cells. 2021;10. doi:10.3390/cells10020205

48. Robles P, Núñez-Delegido E, Ferrández-Ayela A, Sarmiento-Mañús R, Micol JL, Quesada V. Arabidopsis mTERF6 is required for leaf patterning. Plant Science. 2018;266: 117–129. doi:10.1016/j.plantsci.2017.11.003

49. Wu L, Zhou Z-Y, Zhang C-G, Chai J, Zhou Q, Wang L, et al. Functional Roles of Three Cutin Biosynthetic Acyltransferases in Cytokinin Responses and Skotomorphogenesis. PLOS ONE. 2015;10: e0121943.

50. Kong L, Liu Y, Zhi P, Wang X, Xu B, Gong Z, et al. Origins and evolution of cuticle biosynthetic machinery in land plants. Plant Physiology. 2020;184: 1998–2010. doi:10.1104/pp.20.00913

51. Chen X, Truksa M, Snyder CL, El-Mezawy A, Shah S, Weselake RJ. Three Homologous Genes Encoding sn-Glycerol-3-Phosphate Acyltransferase 4 Exhibit Different Expression Patterns and Functional Divergence in Brassica napus. Plant Physiology. 2011;155: 851–865. doi:10.1104/pp.110.169482

52. Fernández-Santos R, Izquierdo Y, López A, Muñiz L, Martínez M, Cascón T, et al. Protein Profiles of Lipid Droplets during the Hypersensitive Defense Response of Arabidopsis against

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Pseudomonas Infection. Plant and Cell Physiology. 2020;61: 1144–1157. doi:10.1093/pcp/pcaa041

53. Xu C, Fan J, Cornish AJ, Benning C. Lipid Trafficking between the Endoplasmic Reticulum and the Plastid in Arabidopsis Requires the Extraplastidic TGD4 Protein. The Plant Cell. 2008;20: 2190–2204. doi:10.1105/tpc.108.061176

54. Yonghua Li-Beisson, Basil Shorrosh, Fred Beisson, Mats X. Andersson, Vincent Arondel, Philip D. Bates, et al. Acyl-Lipid Metabolism. The Arabidopsis Book. 2013;2013. doi:10.1199/tab.0161

55. Huang J, Xue C, Wang H, Wang L, Schmidt W, Shen R, et al. Genes of ACYL CARRIER PROTEIN Family Show Different Expression Profiles and Overexpression of ACYL CARRIER PROTEIN 5 Modulates Fatty Acid Composition and Enhances Salt Stress Tolerance in Arabidopsis. Front Plant Sci. 2017;8: 987. doi:10.3389/fpls.2017.00987

56. Bonaventure G, Ohlrogge JB. Differential regulation of mRNA levels of acyl carrier protein isoforms in Arabidopsis. Plant Physiol. 2002;128: 223–235.

57. Figueroa CM, Asencion Diez MD, Ballicora MA, Iglesias AA. Structure, function, and evolution of plant ADP-glucose pyrophosphorylase. Plant Molecular Biology. 2022;108: 307–323. doi:10.1007/s11103-021-01235-8

58. Liu K, Zou W, Gao X, Wang X, Yu Q, Ge L. Young seedlings adapt to stress by retaining starch and retarding growth through ABA-Dependent and -independent pathways in Arabidopsis. Biochemical and Biophysical Research Communications. 2019;515: 699–705. doi:https://doi.org/10.1016/j.bbrc.2019.06.023

59. Crevillén P, Ventriglia T, Pinto F, Orea A, Mérida Á, Romero JM. Differential Pattern of Expression and Sugar Regulation of Arabidopsis thaliana ADP-glucose Pyrophosphorylase-encoding Genes*. Journal of Biological Chemistry. 2005;280: 8143–8149. doi:https://doi.org/10.1074/jbc.M411713200

60. Eliyahu E, Rog I, Inbal D, Danon A. ACHT4-driven oxidation of APS1 attenuates starch synthesis under low light intensity in *Arabidopsis* plants. Proceedings of the National Academy of Sciences. 2015;112: 12876–12881. doi:10.1073/pnas.1515513112

61. Pourtau N, Jennings R, Pelzer E, Pallas J, Wingler A. Effect of sugar-induced senescence on gene expression and implications for the regulation of senescence in Arabidopsis. Planta. 2006;224: 556–568. doi:10.1007/s00425-006-0243-y

62. Bahaji A, Almagro G, Ezquer I, Gámez-Arcas S, Sánchez-López ÁM, Muñoz FJ, et al. Plastidial Phosphoglucose Isomerase Is an Important Determinant of Seed Yield through Its Involvement in Gibberellin-Mediated Reproductive Development and Storage Reserve Biosynthesis in Arabidopsis. The Plant Cell. 2018;30: 2082–2098. doi:10.1105/tpc.18.00312

63. Seung D, Boudet J, Monroe J, Schreier TB, David LC, Abt M, et al. Homologs of PROTEIN TARGETING TO STARCH Control Starch Granule Initiation in Arabidopsis Leaves. The Plant Cell. 2017;29: 1657–1677. doi:10.1105/tpc.17.00222

64. Zheng M, Zhu C, Yang T, Qian J, Hsu Y-F. GSM2, a transaldolase, contributes to reactive oxygen species homeostasis in Arabidopsis. Plant Mol Biol. 2020;104: 39–53. doi:10.1007/s11103-020-01022-x

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

65. Lee KP, Kim C, Landgraf F, Apel K. EXECUTER1- and EXECUTER2-dependent transfer of stress-related signals from the plastid to the nucleus of Arabidopsis thaliana. Proceedings of the National Academy of Sciences. 2007;104: 10270–10275. doi:10.1073/pnas.0702061104

66. Kim C, Meskauskiene R, Zhang S, Lee KP, Lakshmanan Ashok M, Blajecka K, et al. Chloroplasts of Arabidopsis Are the Source and a Primary Target of a Plant-Specific Programmed Cell Death Signaling Pathway. The Plant Cell. 2012;24: 3026–3039. doi:10.1105/tpc.112.100479

67. Uberegui E, Hall M, Lorenzo Ó, Schröder WP, Balsera M. An Arabidopsis soluble chloroplast proteomic analysis reveals the participation of the Executer pathway in response to increased light conditions. Journal of Experimental Botany. 2015;66: 2067–2077. doi:10.1093/jxb/erv018

68. Maréchal A, Parent J-S, Véronneau-Lafortune F, Joyeux A, Lang BF, Brisson N. Whirly proteins maintain plastid genome stability in Arabidopsis. Proceedings of the National Academy of Sciences. 2009;106: 14693–14698. doi:10.1073/pnas.0901710106

69. Krupinska K, Desel C, Frank S, Hensel G. WHIRLIES Are Multifunctional DNA-Binding Proteins With Impact on Plant Development and Stress Resistance. Frontiers in Plant Science. 2022;13. Available: https://www.frontiersin.org/articles/10.3389/fpls.2022.880423

70. Taylor RE, West CE, Foyer CH. WHIRLY protein functions in plants. Food and Energy Security. 2023;12: e379. doi:10.1002/fes3.379

71. Yu F, Park S-S, Liu X, Foudree A, Fu A, Powikrowska M, et al. SUPPRESSOR OF VARIEGATION4, a New var2 Suppressor Locus, Encodes a Pioneer Protein that Is Required for Chloroplast Biogenesis. Molecular Plant. 2011;4: 229–240. doi:10.1093/mp/ssq074

72. Powikrowska M, Khrouchtchova A, Martens HJ, Zygadlo-Nielsen A, Melonek J, Schulz A, et al. SVR4 (suppressor of variegation 4) and SVR4-like: two proteins with a role in proper organization of the chloroplast genetic machinery. Physiologia Plantarum. 2014;150: 477–492. doi:10.1111/ppl.12108

73. Wang S-W, Li Y, Zhang X-L, Yang H-Q, Han X-F, Liu Z-H, et al. Lacking chloroplasts in guard cells of crumpled leaf attenuates stomatal opening: both guard cell chloroplasts and mesophyll contribute to guard cell ATP levels. Plant, Cell & Environment. 2014;37: 2201–2210. doi:10.1111/pce.12297

74. Jarvis P, López-Juez E. Biogenesis and homeostasis of chloroplasts and other plastids. Nature Reviews Molecular Cell Biology. 2013. doi:10.1038/nrm3702

75. Osteryoung KW, Pyke KA. Division and dynamic morphology of plastids. Annual Review of Plant Biology. 2014. doi:10.1146/annurev-arplant-050213-035748

76. Richardson LGL, Schnell DJ. Origins, function, and regulation of the TOC–TIC general protein import machinery of plastids. Journal of Experimental Botany. 2020;71: 1226–1238. doi:10.1093/jxb/erz517

77. de Vries J, Gould SB. The monoplastidic bottleneck in algae and plant evolution. Journal of Cell Science. 2018. doi:10.1242/jcs.203414

78. Chen L, Sun B, Gao W, Zhang Q, Yuan H, Zhang M. MCD1 Associates with FtsZ Filaments via the Membrane-Tethering Protein ARC6 to Guide Chloroplast Division. The Plant Cell. 2018;30: 1807–1823. doi:10.1105/tpc.18.00189

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

79. Muranaka A, Watanabe S, Sakamoto A, Shimada H. Arabidopsis cotyledon chloroplast biogenesis factor CYO1 uses glutathione as an electron donor and interacts with PSI (A1 and A2) and PSII (CP43 and CP47) subunits. Journal of Plant Physiology. 2012;169: 1212–1215. doi:10.1016/j.jplph.2012.04.001

80. Tominaga J, Mizutani H, Horikawa D, Nakahara Y, Takami T, Sakamoto W, et al. Rice CYO1, an ortholog of Arabidopsis thaliana cotyledon chloroplast biogenesis factor AtCYO1, is expressed in leaves and involved in photosynthetic performance. Journal of Plant Physiology. 2016;207: 78–83. doi:10.1016/j.jplph.2016.10.005

81. Yu Q-B, Lu Y, Ma Q, Zhao T-T, Huang C, Zhao H-F, et al. TAC7, an essential component of the plastid transcriptionally active chromosome complex, interacts with FLN1, TAC10, TAC12 and TAC14 to regulate chloroplast gene expression in Arabidopsis thaliana. Physiologia Plantarum. 2013;148: 408–421. doi:10.1111/j.1399-3054.2012.01718.x

82. Asano T, Yoshioka Y, Kurei S, Sakamoto W, Machida Y. A mutation of the CRUMPLED LEAF gene that encodes a protein localized in the outer envelope membrane of plastids affects the pattern of cell division, cell differentiation, and plastid division in Arabidopsis. Plant J. 2004;38: 448–459. doi:10.1111/j.1365-313X.2004.02057.x

83. Chen Y, Asano T, Fujiwara MT, Yoshida S, Machida Y, Yoshioka Y. Plant Cells Without Detectable Plastids are Generated in the crumpled leaf Mutant of Arabidopsis thaliana. Plant and Cell Physiology. 2009;50: 956–969. doi:10.1093/pcp/pcp047

84. Smith MD, Rounds CM, Wang F, Chen K, Afitlhile M, Schnell DJ. atToc159 is a selective transit peptide receptor for the import of nucleus-encoded chloroplast proteins. Journal of Cell Biology. 2004;165: 323–334. doi:10.1083/jcb.200311074

85. Agne B, Andrès C, Montandon C, Christ B, Ertan A, Jung F, et al. The Acidic A-Domain of Arabidopsis Toc159 Occurs as a Hyperphosphorylated Protein. Plant Physiology. 2010;153: 1016–1030. doi:10.1104/pp.110.158048

86. Afitlhile M, Fry M, Workman S. The TOC159 mutant of Arabidopsis thaliana accumulates altered levels of saturated and polyunsaturated fatty acids. Plant Physiology and Biochemistry. 2015;87: 61–72. doi:10.1016/j.plaphy.2014.12.018

87. Afitlhile M, Worthington R, Heda G, Brown L. The TOC159 null mutant of Arabidopsis thaliana is impaired in the accumulation of plastid lipids and phosphatidylcholine. Plant Physiology and Biochemistry. 2021;159: 148–159. doi:10.1016/j.plaphy.2020.12.011

88. Afitlhile M, Worthington R, Baldric J. The toc132toc120 heterozygote mutant of Arabidopsis thaliana accumulates decreased levels of the major chloroplast lipids. Phytochemistry. 2021;184: 112652. doi:10.1016/j.phytochem.2020.112652

89. Chiu C-C, Chen L-J, Li H. Pea Chloroplast DnaJ-J8 and Toc12 Are Encoded by the Same Gene and Localized in the Stroma. Plant Physiology. 2010;154: 1172–1182. doi:10.1104/pp.110.161224

90. Ling Q, Jarvis P. Regulation of Chloroplast Protein Import by the Ubiquitin E3 Ligase SP1 Is Important for Stress Tolerance in Plants. Curr Biol. 2015;25: 2527–2534. doi:10.1016/j.cub.2015.08.015

91. Harris BJ, Clark JW, Schrempf D, Szöllősi GJ, Donoghue PCJ, Hetherington AM, et al. Divergent evolutionary trajectories of bryophytes and tracheophytes from a complex common ancestor of land plants. Nature Ecology & Evolution. 2022. doi:10.1038/s41559-022-01885-x

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

92. Brawley SH, Blouin NA, Ficko-Blean E, Wheeler GL, Lohr M, Goodson HV, et al. Insights into the red algae and eukaryotic evolution from the genome of Porphyra umbilicalis (Bangiophyceae, Rhodophyta). Proceedings of the National Academy of Sciences. 2017;114: E6361–E6370. doi:10.1073/pnas.1703088114

93. Bhattacharya D, Price DC, Chan CX, Qiu H, Rose N, Ball S, et al. Genome of the red alga Porphyridium purpureum. Nature Communications. 2013;4: 1941. doi:10.1038/ncomms2931

94. Collén J, Porcel B, Carré W, Ball SG, Chaparro C, Tonon T, et al. Genome structure and metabolic features in the red seaweed Chondrus crispus shed light on evolution of the Archaeplastida. Proceedings of the National Academy of Sciences. 2013;110: 5247–5252. doi:10.1073/pnas.1221259110

95. Qiu H, Price DC, Yang EC, Yoon HS, Bhattacharya D. Evidence of ancient genome reduction in red algae (Rhodophyta). Journal of Phycology. 2015;51: 624–636. doi:10.1111/jpy.12294

96. Lee J, Yang EC, Graf L, Yang JH, Qiu H, Zelzion U, et al. Analysis of the Draft Genome of the Red Seaweed Gracilariopsis chorda Provides Insights into Genome Size Evolution in Rhodophyta. Molecular Biology and Evolution. 2018;35: 1869–1886. doi:10.1093/molbev/msy081

97. Yang W, Pollard M, Li-Beisson Y, Beisson F, Feig M, Ohlrogge J. A distinct type of glycerol-3-phosphate acyltransferase with sn-2 preference and phosphatase activity producing 2-monoacylglycerol. Proceedings of the National Academy of Sciences of the United States of America. 2010;107: 12040–12045. doi:10.1073/pnas.0914149107

98. Yang W, Simpson JP, Li-Beisson Y, Beisson F, Pollard M, Ohlrogge JB. A land-plant-specific glycerol-3-phosphate acyltransferase family in arabidopsis: Substrate specificity, sn-2 preference, and evolution. Plant Physiology. 2012;160: 638–652. doi:10.1104/pp.112.201996

99. Ball S, Colleoni C, Cenci U, Raj JN, Tirtiaux C. The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. J Exp Bot. 2011;62: 1775–1801. doi:10.1093/jxb/erq411

100. Viola R, Nyvall P, Pedersén M. The unique features of starch metabolism in red algae. Proc Biol Sci. 2001;268: 1417–1422. doi:10.1098/rspb.2001.1644

101. Lechner M, Rossmanith W, Hartmann RK, Thölken C, Gutmann B, Giegé P, et al. Distribution of ribonucleoprotein and protein-only RNase P in Eukarya. Molecular Biology and Evolution. 2015;32: 3186–3193. doi:10.1093/molbev/msv187

102. Gutmann B, Gobert A, Giegé P. PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. Genes and Development. 2012;26: 1022–1027. doi:10.1101/gad.189514.112

103. Gobert A, Gutmann B, Taschner A, Gössringer M, Holzmann J, Hartmann RK, et al. A single Arabidopsis organellar protein has RNase P activity. Nature structural & molecular biology. 2010;17: 740–744. doi:10.1038/nsmb.1812

104. Ban T, Ke J, Chen R, Gu X, Tan MHE, Zhou XE, et al. Structure of a PLS-class Pentatricopeptide Repeat Protein Provides Insights into Mechanism of RNA Recognition. Journal of Biological Chemistry. 2013;288: 31540–31548. doi:10.1074/jbc.M113.496828

105. Wang X, Zhao L, Man Y, Li X, Wang L, Xiao J. PDM4, a Pentatricopeptide Repeat Protein, Affects Chloroplast Gene Expression and Chloroplast Development in Arabidopsis thaliana.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Frontiers in Plant Science. 2020;11. Available: https://www.frontiersin.org/article/10.3389/fpls.2020.01198

106. Zoschke R, Qu Y, Zubo YO, Börner T, Schmitz-Linneweber C. Mutation of the pentatricopeptide repeat-SMR protein SVR7 impairs accumulation and translation of chloroplast ATP synthase subunits in Arabidopsis thaliana. Journal of Plant Research. 2013;126: 403–414. doi:10.1007/s10265-012-0527-1

107. Wu W, Liu S, Ruwe H, Zhang D, Melonek J, Zhu Y, et al. SOT1, a pentatricopeptide repeat protein with a small MutS-related domain, is required for correct processing of plastid 23S–4.5S rRNA precursors in Arabidopsis thaliana. The Plant Journal. 2016;85: 607–621. doi:10.1111/tpj.13126

108. De Longevialle AF, Hendrickson L, Taylor NL, Delannoy E, Lurin C, Badger M, et al. The pentatricopeptide repeat gene OTP51 with two LAGLIDADG motifs is required for the cis-splicing of plastid ycf3 intron 2 in Arabidopsis thaliana. The Plant Journal. 2008;56: 157–168. doi:10.1111/j.1365-313X.2008.03581.x

109. Ye J-W, Gong Z-Y, Chen C-G, Mi H-L, Chen G-Y. A Mutation of OSOTP 51 Leads to Impairment of Photosystem I Complex Assembly and Serious Photo-damage in Rice. Journal of Integrative Plant Biology. 2012;54: 87–98. doi:10.1111/j.1744-7909.2012.01094.x

110. Huang W, Zhu Y, Wu W, Li X, Zhang D, Yin P, et al. The Pentatricopeptide Repeat Protein SOT5/EMB2279 Is Required for Plastid rpl2 and trnK Intron Splicing. Plant Physiology. 2018;177: 684–697. doi:10.1104/pp.18.00406

111. Cyrek M, Fedak H, Ciesielski A, Guo Y, Sliwa A, Brzezniak L, et al. Seed Dormancy in Arabidopsis Is Controlled by Alternative Polyadenylation of DOG1. Plant Physiology. 2016;170: 947–955. doi:10.1104/pp.15.01483

112. Mullineaux CW. Function and evolution of grana. Trends in Plant Science. 2005;10: 521–525. doi:10.1016/j.tplants.2005.09.001

113. Xu X, Ouyang M, Lu D, Zheng C, Zhang L. Protein Sorting within Chloroplasts. Trends in Cell Biology. 2021;31: 9–16. doi:10.1016/j.tcb.2020.09.011

114. Stamenković M, Woelken E, Hanelt D. Ultrastructure of Cosmarium strains (Zygnematophyceae, Streptophyta) collected from various geographic locations shows species-specific differences both at optimal and stress temperatures. Protoplasma. 2014;251: 1491–1509. doi:10.1007/s00709-014-0652-x

115. Wietrzynski W, Schaffer M, Tegunov D, Albert S, Kanazawa A, Plitzko JM, et al. Charting the native architecture of chlamydomonas thylakoid membranes with single-molecule precision. eLife. 2020. doi:10.7554/eLife.53740

116. Jiao C, Sørensen I, Sun X, Sun H, Behar H, Alseekh S, et al. The Penium margaritaceum Genome: Hallmarks of the Origins of Land Plants. Cell. 2020. doi:10.1016/j.cell.2020.04.019

117. Khatoon M, Inagawa K, Pospíšil P, Yamashita A, Yoshioka M, Lundin B, et al. Quality Control of Photosystem II: Thylakoid unstacking is necessary to avoid further damage to the D1 protein and to facilitate D1 degradation under light stress in spinach thylakoids. Journal of Biological Chemistry. 2009;284: 25343–25352. doi:https://doi.org/10.1074/jbc.M109.007740

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

118. Kirchhoff H, Hall C, Wood M, Herbstová M, Tsabari O, Nevo R, et al. Dynamic control of protein diffusion within the granal thylakoid lumen. Proceedings of the National Academy of Sciences. 2011;108: 20248–20253. doi:10.1073/pnas.1104141109

119. Garg SG, Gould SB. The Role of Charge in Protein Targeting Evolution. Trends in Cell Biology. 2016;26: 894–905. doi:10.1016/j.tcb.2016.07.001

120. Nakanishi H, Suzuki K, Kabeya Y, Miyagishima S. Plant-Specific Protein MCD1 Determines the Site of Chloroplast Division in Concert with Bacteria-Derived MinD. Current Biology. 2009;19: 151–156. doi:10.1016/j.cub.2008.12.018

121. de Vries J, Stanton A, Archibald JM, Gould SB. Streptophyte Terrestrialization in Light of Plastid Evolution. Trends in Plant Science. 2016. doi:10.1016/j.tplants.2016.01.021

122. Larkin RM, Stefano G, Ruckle ME, Stavoe AK, Sinkler CA, Brandizzi F, et al. REDUCED CHLOROPLAST COVERAGE genes from Arabidopsis thaliana help to establish the size of the chloroplast compartment. Proceedings of the National Academy of Sciences. 2016;113: E1116–E1125. doi:10.1073/pnas.1515741113

123. El Zawily AM, Schwarzländer M, Finkemeier I, Johnston IG, Benamar A, Cao Y, et al. FRIENDLY Regulates Mitochondrial Distribution, Fusion, and Quality Control in Arabidopsis. Plant Physiology. 2014;166: 808–828. doi:10.1104/pp.114.243824

124. Zhu Q, Hulen D, Liu T, Clarke M. The cluA- mutant of Dictyostelium identifies a novel class of proteins required for dispersion of mitochondria. Proc Natl Acad Sci U S A. 1997;94: 7308–7313. doi:10.1073/pnas.94.14.7308

125. Raval PK, Ngan WY, Gallie J, Agashe D. The layered costs and benefits of translational redundancy. Pilpel Y, Barkai N, Schirman D, editors. eLife. 2023;12: e81005. doi:10.7554/eLife.81005

126. Roller BRK, Stoddard SF, Schmidt TM. Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. Nature Microbiology. 2016;1: 16160. doi:10.1038/nmicrobiol.2016.160

127. Raval PK, Garg SG, Gould SB. Endosymbiotic selective pressure at the origin of eukaryotic cell biology. Chacinska A, Perry GH, Chacinska A, Perry GH, editors. eLife. 2022;11: e81033. doi:10.7554/eLife.81033

128. Raval PK, Martin WF, Gould SB. Mitochondrial evolution: Gene shuffling, endosymbiosis, and signaling. Science Advances. 2023;9: eadj4493. doi:10.1126/sciadv.adj4493

129. Gould SB, Garg SG, Martin WF. Bacterial Vesicle Secretion and the Evolutionary Origin of the Eukaryotic Endomembrane System. Trends in Microbiology. 2016;24: 525–534. doi:https://doi.org/10.1016/j.tim.2016.03.005

130. Hoecker U. The activities of the E3 ubiquitin ligase COP1/SPA, a key repressor in light signaling. Current Opinion in Plant Biology. 2017;37: 63–69. doi:https://doi.org/10.1016/j.pbi.2017.03.015

131. Borden KLB. RING domains: master builders of molecular scaffolds?11Edited by P. E. Wright. Journal of Molecular Biology. 2000;295: 1103–1112. doi:10.1006/jmbi.1999.3429

132. Stone SL. The role of ubiquitin and the 26S proteasome in plant abiotic stress signaling. Front Plant Sci. 2014;5. doi:10.3389/fpls.2014.00135

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

133. Jiménez-López D, Muñóz-Belman F, González-Prieto JM, Aguilar-Hernández V, Guzmán P. Repertoire of plant RING E3 ubiquitin ligases revisited: New groups counting gene families and single genes. PLOS ONE. 2018;13: e0203442. doi:10.1371/journal.pone.0203442

134. Thayale Purayil F, Sudalaimuthuasari N, Li L, Aljneibi R, Al Shamsi AM, David N, et al. Transcriptome Profiling and Functional Validation of RING-Type E3 Ligases in Halophyte Sesuvium verrucosum under Salinity Stress. International Journal of Molecular Sciences. 2022;23. doi:10.3390/ijms23052821

135. Basnayake BMVS, Li D, Zhang H, Li G, Virk N, Song F. Arabidopsis DAL1 and DAL2, two RING finger proteins homologous to Drosophila DIAP1, are involved in regulation of programmed cell death. Plant Cell Reports. 2011;30: 37–48. doi:10.1007/s00299-010-0941-6

136. Ling Q, Huang W, Baldwin A, Jarvis P. Chloroplast Biogenesis Is Regulated by Direct Action of the Ubiquitin-Proteasome System. Science. 2012;338: 655–659. doi:10.1126/science.1225053

137. Pan R, Satkovich J, Hu J. E3 ubiquitin ligase SP1 regulates peroxisome biogenesis in Arabidopsis. Proceedings of the National Academy of Sciences. 2016;113: E7307–E7316. doi:10.1073/pnas.1613530113

138. Oikawa K, Matsunaga S, Mano S, Kondo M, Yamada K, Hayashi M, et al. Physical interaction between peroxisomes and chloroplasts elucidated by in situ laser analysis. Nature Plants. 2015;1: 15035. doi:10.1038/nplants.2015.35

139. Liu H, Stone SL. Abscisic Acid Increases Arabidopsis ABI5 Transcription Factor Levels by Promoting KEG E3 Ligase Self-Ubiquitination and Proteasomal Degradation. The Plant Cell. 2010;22: 2630–2641. doi:10.1105/tpc.110.076075

140. Liu H, Stone SL. Cytoplasmic Degradation of the Arabidopsis Transcription Factor ABSCISIC ACID INSENSITIVE 5 Is Mediated by the RING-type E3 Ligase KEEP ON GOING*. Journal of Biological Chemistry. 2013;288: 20267–20279. doi:10.1074/jbc.M113.465369

141. Azevedo J, Courtois F, Hakimi M-A, Demarsy E, Lagrange T, Alcaraz J-P, et al. Intraplastidial trafficking of a phage-type RNA polymerase is mediated by a thylakoid RING-H2 protein. Proceedings of the National Academy of Sciences. 2008;105: 9123–9128. doi:10.1073/pnas.0800909105

142. Naramoto S, Hata Y, Fujita T, Kyozuka J. The bryophytes Physcomitrium patens and Marchantia polymorpha as model systems for studying evolutionary cell and developmental biology in plants. The Plant Cell. 2021;34: 228–246. doi:10.1093/plcell/koab218

143. Szövényi P, Frangedakis E, Ricca M, Quandt D, Wicke S, Langdale JA. Establishment of Anthoceros agrestis as a model species for studying the biology of hornworts. BMC Plant Biology. 2015;15: 98. doi:10.1186/s12870-015-0481-x

144. Terashima M, Specht M, Hippler M. The chloroplast proteome: a survey from the Chlamydomonas reinhardtii perspective with a focus on distinctive features. Current Genetics. 2011;57: 151–168. doi:10.1007/s00294-011-0339-1

145. Mueller SJ, Lang D, Hoernstein SNW, Lang EGE, Schuessele C, Schmidt A, et al. Quantitative Analysis of the Mitochondrial and Plastid Proteomes of the Moss Physcomitrella patens Reveals Protein Macrocompartmentation and Microcompartmentation. Plant Physiology. 2014;164: 2081–2095. doi:10.1104/pp.114.235754

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.01.555919; this version posted September 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

146. Sun Q, Zybailov B, Majeran W, Friso G, Olinares PDB, van Wijk KJ. PPDB, the Plant Proteomics Database at Cornell. Nucleic Acids Research. 2009;37: D969–D974. doi:10.1093/nar/gkn654

147. Emms DM, Kelly S. STAG: Species Tree Inference from All Genes. bioRxiv. 2018; 267914–267914. doi:10.1101/267914

148. Tria FDK, Landan G, Dagan T. Phylogenetic rooting using minimal ancestor deviation. Nature Ecology & Evolution. 2017;1: 0193. doi:10.1038/s41559-017-0193

149. Rambaut A, Drummond A. FigTree Version 1.4.4. 2018.

150. Revell LJ. phytools: An R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution. 2012. doi:10.1111/j.2041-210X.2011.00169.x

151. Mooers AØ, Schluter D. Reconstructing Ancestor States with Maximum Likelihood: Support for One-and Two-Rate Models. Systematic Biology. 1999;48: 623–633. doi:10.1080/106351599260193

152. Schluter D, Price T, Mooers AØ, Ludwig D. Likelihood of ancestor states in adaptive radiation. Evolution. 1997;51: 1699–1711. doi:10.1111/j.1558-5646.1997.tb05095.x

153. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49: D412–D419. doi:10.1093/nar/gkaa913

9 Acknowledgements

I would like to express my deepest gratitude to the following individuals for their invaluable support throughout my academic journey.

Firstly, I want to thank Sven for his unwavering help, mentorship, and brilliant discussions. His knowledge and insights were instrumental in shaping my work, testing my hypotheses, and developing my understanding of molecular evolution and evolutionary cell biology. Without your help, support, and inspiration to be a better scientist, I would not have been able to pursue my own research goals. I could not have wishes for a better PhD supervisor.

I am also immensely grateful to Bill for his generous funding, which provided me with the necessary resources to carry out my research in plastid evolution and attend various conferences and workshops.

My sincere appreciation also goes to Parth, Carolina, and Maria for being amazing colleagues and for their tireless assistance in all aspects of my work. Their contributions, support and friendship were truly invaluable and made my research experience enjoyable and rewarding. One could not wish for better fellow team members.

I would also like to extend my gratitude to the entire Molecular Evolution team for their assistance and support. In particular, Falk, Michi, and Nils. Falk, for his deep philosophical discussions; Michi, for his help with BLAST scripts and OrthoFinder; and Nils, for his expertise in server commands and sequence alignment. These were all crucial to the success of my research.

I would like to express my gratitude to the developers of ChatGPT whose software was an essential tool in ensuring that my thesis was grammatically correct.

I would also like to express my thanks to my family for their unwavering personal, financial, and emotional support. Their love and encouragement have been instrumental in enabling me to pursue my PhD, and I am forever grateful for their unwavering belief in me.

Finally, and most importantly, I would like to thank Vy for her love, motivation, and delicious Vietnamese food. Her unwavering support and inspiration were essential to my success, and I am grateful for everything she has done for me.

Thank you all for your invaluable contributions, unwavering support, and encouragement throughout my academic journey. I could not have achieved this milestone without you, and I am forever grateful.

10 References

Altenhoff, A. M., and Dessimoz, C. (2009). Phylogenetic and Functional Assessment of Orthologs Inference Projects and Methods. *PLOS Comput. Biol.* 5, e1000262. doi: 10.1371/journal.pcbi.1000262.

Arneth, A., Sitch, S., Pongratz, J., Stocker, B. D., Ciais, P., Poulter, B., et al. (2017). Historical carbon dioxide emissions caused by land-use changes are possibly larger than assumed. *Nat. Geosci.* 10, 79–84. doi: 10.1038/ngeo2882.

Bachy, C., Wittmers, F., Muschiol, J., Hamilton, M., Henrissat, B., and Worden, A. Z. (2022). The Land–Sea Connection: Insights Into the Plant Lineage from a Green Algal Perspective. *Annu. Rev. Plant Biol.* 73, 585–616. doi: 10.1146/annurev-arplant-071921-100530.

Berg, J., M., Tymoczko, J. L., Gatto, G. J., and Stryer, L. (2015). *Biochemistry 8th Edition*.

Bowles, A. M. C., Williamson, C. J., Williams, T. A., Lenton, T. M., and Donoghue, P. C. J. (2023). The origin and early evolution of plants. *Trends Plant Sci.* 28, 312–329. doi: 10.1016/j.tplants.2022.09.009.

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176.

Carroll, S. (2005). Endless Forms Most Beautiful: The New Science of Evo Devo and the Making of the Animal Kingdom.

Cheng, S., Xian, W., Fu, Y., Marin, B., Keller, J., Wu, T., et al. (2019). Genomes of Subaerial Zygnematophyceae Provide Insights into Land Plant Evolution. *Cell*. doi: 10.1016/j.cell.2019.10.019.

Choi, H., Yi, T., and Ha, S.-H. (2021). Diversity of Plastid Types and Their Interconversions. *Front. Plant Sci.* 12. Available at: https://www.frontiersin.org/article/10.3389/fpls.2021.692024.

Dahl, T. W., and Arens, S. K. M. (2020). The impacts of land plant evolution on Earth's climate and oxygenation state – An interdisciplinary review. *Chem. Geol.* 547, 119665–119665. doi: 10.1016/j.chemgeo.2020.119665.

de Vries, J., and Archibald, J. M. (2018). Plant evolution: landmarks on the path to terrestrial life. *New Phytol.* doi: 10.1111/nph.14975.

de Vries, J., and Gould, S. B. (2018). The monoplastidic bottleneck in algae and plant evolution. *J. Cell Sci.* doi: 10.1242/jcs.203414.

de Vries, J., Stanton, A., Archibald, J. M., and Gould, S. B. (2016). Streptophyte Terrestrialization in Light of Plastid Evolution. *Trends Plant Sci.* doi: 10.1016/j.tplants.2016.01.021.

Delaux, P.-M., Hetherington, A. J., Coudert, Y., Delwiche, C., Dunand, C., Gould, S., et al. (2019). Reconstructing trait evolution in plant evo–devo studies. *Curr. Biol.* 29, R1110–R1118. doi: 10.1016/j.cub.2019.09.044.

Delwiche, C. F., and Timme, R. E. (2011). Plants. *Curr. Biol.* 21. doi: 10.1016/j.cub.2011.04.021.

Dowson, A. J., Lloyd, A. J., Cuming, A. C., Roper, D. I., Frigerio, L., and Dowson, C. G. (2022). Plant peptidoglycan precursor biosynthesis: Conservation between moss chloroplasts and Gram negative bacteria. *Plant Physiol.*, kiac176–kiac176. doi: 10.1093/plphys/kiac176.

Duckett, J. G., and Ligrone, R. (1993). Plastid-dividing Rings in Ferns. *Ann. Bot.* 72, 619–627. doi: 10.1006/anbo.1993.1153.

Edwards, Y., and Keller, P. (2004). Handbook of Comparative Genomics: Principles and Methodology. *Brief. Funct. Genomics* 2. doi: 10.1093/bfgp/2.4.366.

Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16, 157–157. doi: 10.1186/s13059-015-0721-2.

Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238–238. doi: 10.1186/s13059-019-1832-y.

Emms, D. M., and Kelly, S. (2022). SHOOT: phylogenetic gene search and ortholog inference. *Genome Biol.* 23, 85–85. doi: 10.1186/s13059-022-02652-8.

Figueroa-Martinez, F., Nedelcu, A. M., Smith, D. R., and Reyes-Prieto, A. (2017). The plastid genome of polytoma uvella is the largest known among colorless algae and plants and reflects contrasting evolutionary paths to nonphotosynthetic lifestyles. *Plant Physiol.* doi: 10.1104/pp.16.01628.

Flori, S., Jouneau, P.-H., Bailleul, B., Gallet, B., Estrozi, L. F., Moriscot, C., et al. (2017). Plastid thylakoid architecture optimizes photosynthesis in diatoms. *Nat. Commun.* 8, 15885–15885. doi: 10.1038/ncomms15885.

Frain, K. M., Gangl, D., Jones, A., Zedler, J. A. Z., and Robinson, C. (2016). Protein translocation and thylakoid biogenesis in cyanobacteria. *Biochim. Biophys. Acta - Bioenerg.* 1857, 266–273. doi: 10.1016/j.bbabio.2015.08.010.

Frangedakis, E., Marron, A. O., Waller, M., Neubauer, A., Tse, S. W., Yue, Y., et al. (2023). What can hornworts teach us? *Front. Plant Sci.* 14. Available at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1108027.

Frangedakis, E., Shimamura, M., Villarreal, J. C., Li, F. W., Tomaselli, M., Waller, M., et al. (2020). The hornworts: morphology, evolution and development. *New Phytol.* doi: 10.1111/nph.16874.

Frangedakis, E., Waller, M., Nishiyama, T., Tsukaya, H., Xu, X., Yue, Y., et al. (2021). An Agrobacterium-mediated stable transformation technique for the hornwort model Anthoceros agrestis. *New Phytol.* 232, 1488–1505. doi: 10.1111/nph.17524.

Fricke, E. C., Ordonez, A., Rogers, H. S., and Svenning, J.-C. (2022). The effects of defaunation on plants' capacity to track climate change. *Science* 375, 210–214. doi: 10.1126/science.abk3510.

Gao, H., Kadirjan-Kalbach, D., Froehlicht, J. E., and Osteryoung, K. W. (2003). ARC5, a cytosolic dynamin-like protein from plants, is part of the chloroplast division machinery. *Proc. Natl. Acad. Sci. U. S. A.* doi: 10.1073/pnas.0530206100.

Garg, S. G., and Gould, S. B. (2016). The Role of Charge in Protein Targeting Evolution. *Trends Cell Biol.* 26, 894–905. doi: 10.1016/j.tcb.2016.07.001.

Gawryluk, R. M. R., Tikhonenkov, D. V., Hehenberger, E., Husnik, F., Mylnikov, A. P., and Keeling, P. J. (2019). Non-photosynthetic predators are sister to red algae. *Nature*. doi: 10.1038/s41586-019-1398-6.

Glynn, J. M., Froehlich, J. E., and Osteryoung, K. W. (2008). Arabidopsis ARC6 coordinates the division machineries of the inner and outer chloroplast membranes through interaction with PDV2 in the intermembrane Space. *Plant Cell*. doi: 10.1105/tpc.108.061440.

Glynn, J. M., Yang, Y., Vitha, S., Schmitz, A. J., Hemmes, M., Miyagishima, S. Y., et al. (2009). PARC6, a novel chloroplast division factor, influences FtsZ assembly and is required for recruitment of PDV1 during chloroplast division in Arabidopsis. *Plant J.* doi: 10.1111/j.1365-313X.2009.03905.x.

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944.

Grosche, C., and Rensing, S. A. (2017). Three rings for the evolution of plastid shape: a tale of land plant FtsZ. *Protoplasma*. doi: 10.1007/s00709-017-1096-x.

Harris, B. J., Clark, J. W., Schrempf, D., Szöllősi, G. J., Donoghue, P. C. J., Hetherington, A. M., et al. (2022). Divergent evolutionary trajectories of bryophytes and tracheophytes from a complex common ancestor of land plants. *Nat. Ecol. Evol.* doi: 10.1038/s41559-022-01885-x.

Hirano, T., Tanidokoro, K., Shimizu, Y., Kawarabayasi, Y., Ohshima, T., Sato, M., et al. (2016). Moss chloroplasts are surrounded by a peptidoglycan wall containing D-amino acids. *Plant Cell*. doi: 10.1105/tpc.16.00104.

Holland, B. R., Ketelaar-Jones, S., O'Mara, A. R., Woodhams, M. D., and Jordan, G. J. (2020). Accuracy of ancestral state reconstruction for non-neutral traits. *Sci. Rep.* 10, 7644–7644. doi: 10.1038/s41598-020-64647-4.

Homi, S., Takechi, K., Tanidokoro, K., Sato, H., Takio, S., and Takano, H. (2009). The peptidoglycan biosynthesis genes MurA and mray are related to chloroplast division in the moss physcomitrella patens. *Plant Cell Physiol.* doi: 10.1093/pcp/pcp158.

Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., et al. (2014). Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.* doi: 10.1038/ncomms4978.

Izumi, Y., Kuroki, J., Nagafuji, H., Lin, X., and Takano, H. (2008). Effects of Antibiotics That Inhibit Bacterial Peptidoglycan Synthesis on Plastid Division in Pteridophytes. *Cytologia (Tokyo)* 73, 393–400. doi: 10.1508/cytologia.73.393.

Jarvis, P., and López-Juez, E. (2013). Biogenesis and homeostasis of chloroplasts and other plastids. *Nat. Rev. Mol. Cell Biol.* doi: 10.1038/nrm3702.

Joy, J. B., Liang, R. H., McCloskey, R. M., Nguyen, T., and Poon, A. F. Y. (2016). Ancestral Reconstruction. *PLOS Comput. Biol.* 12, e1004763.

Katayama, N., Takano, H., Sugiyama, M., Takio, S., Sakai, A., Tanaka, K., et al. (2003). Effects of Antibiotics that Inhibit the Bacterial Peptidoglycan Synthesis Pathway on Moss Chloroplast Division. *Plant Cell Physiol.* 44, 776–781. doi: 10.1093/pcp/pcg096.

Keeling, P. J. (2010). The endosymbiotic origin, diversification and fate of plastids. *Philos. Trans. R. Soc. B Biol. Sci.* doi: 10.1098/rstb.2009.0103.

Keeling, P. J. (2013). The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu. Rev. Plant Biol.* doi: 10.1146/annurev-arplant-050312-120144.

Knopp, M., Garg, S. G., Handrich, M., and Gould, S. B. (2020). Major Changes in Plastid Protein Import and the Origin of the Chloroplastida. *iScience*. doi: 10.1016/j.isci.2020.100896.

Lang, D., Ullrich, K. K., Murat, F., Fuchs, J., Jenkins, J., Haas, F. B., et al. (2018). The Physcomitrella patens chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* 93, 515–533. doi: 10.1111/tpj.13801.

Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M. A., Graham, S. W., et al. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574, 679–685. doi: 10.1038/s41586-019-1693-2.

Leliaert, F., Smith, D. R., Moreau, H., Herron, M. D., Verbruggen, H., Delwiche, C. F., et al. (2012). Phylogeny and Molecular Evolution of the Green Algae. *Crit. Rev. Plant Sci.* doi: 10.1080/07352689.2011.615705.

Lenton, T. M., Dahl, T. W., Daines, S. J., Mills, B. J. W., Ozaki, K., Saltzman, M. R., et al. (2016). Earliest land plants created modern levels of atmospheric oxygen. *Proc. Natl. Acad. Sci. U. S. A.* doi: 10.1073/pnas.1604787113.

Li, F.-W., Nishiyama, T., Waller, M., Frangedakis, E., Keller, J., Li, Z., et al. (2020a). Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. *Nat. Plants* 6, 259–272. doi: 10.1038/s41477-020-0618-2.

Li, F.-W., Villarreal A., J., and Szövényi, P. (2017). Hornworts: An Overlooked Window into Carbon-Concentrating Mechanisms. *Trends Plant Sci.* 22. doi: 10.1016/j.tplants.2017.02.002.

Li, L., Wang, S., Wang, H., Sahu, S. K., Marin, B., Li, H., et al. (2020b). The genome of Prasinoderma coloniale unveils the existence of a third phylum within green plants. *Nat. Ecol. Evol.* 4, 1220–1231. doi: 10.1038/s41559-020-1221-7.

Li, X., Hou, Z., Xu, C., Shi, X., Yang, L., Lewis, L. A., et al. (2021). Large Phylogenomic Data sets Reveal Deep Relationships and Trait Evolution in Chlorophyte Green Algae. *Genome Biol. Evol.* 13, evab101–evab101. doi: 10.1093/gbe/evab101.

Liechti, G. W., Kuru, E., Hall, E., Kalinda, A., Brun, Y. V., VanNieuwenhze, M., et al. (2014). A new metabolic cell-wall labelling method reveals peptidoglycan in Chlamydia trachomatis. *Nature* 506, 507–510. doi: 10.1038/nature12892.

Liu, J.-W., Li, S.-F., Wu, C.-T., Valdespino, I. A., Ho, J.-F., Wu, Y.-H., et al. (2020). Gigantic chloroplasts, including bizonoplasts, are common in shade-adapted species of the ancient vascular plant family Selaginellaceae. *Am. J. Bot.* 107, 562–576. doi: 10.1002/ajb2.1455.

Machida, M., Takechi, K., Sato, H., Chung, S. J., Kuroiwa, H., Takio, S., et al. (2006). Genes for the peptidoglycan synthesis pathway are essential for chloroplast division in moss. *Proc. Natl. Acad. Sci. U. S. A.* doi: 10.1073/pnas.0510693103.

MacLeod, A. I., Knopp, M. R., and Gould, S. B. (2023). A mysterious cloak: the peptidoglycan layer of algal and plant plastids. *Protoplasma*. doi: 10.1007/s00709-023-01886-y.

MacLeod, A. I., Raval, P. K., Stockhorst, S., Knopp, M. R., Frangedakis, E., and Gould, S. B. (2022). Loss of Plastid Developmental Genes Coincides With a Reversion to Monoplastidy in Hornworts. *Front. Plant Sci.* 13. Available at: https://www.frontiersin.org/article/10.3389/fpls.2022.863076.

Martin, A., Lang, D., Hanke, S. T., Mueller, S. J. X., Sarnighausen, E., Vervliet-Scheebaum, M., et al. (2009). Targeted gene knockouts reveal overlapping functions of the five physcomitrella patens ftsz isoforms in chloroplast division, chloroplast shaping, cell patterning, plant development, and gravity sensing. *Mol. Plant*. doi: 10.1093/mp/ssp076.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., et al. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015.

Miyagishima, S. Y., Froehlich, J. E., and Osteryoung, K. W. (2006). PDV1 and PDV2 mediate recruitment of the dynamin-related protein ARC5 to the plastid division site. *Plant Cell*. doi: 10.1105/tpc.106.045484.

Morris, J. L., Puttick, M. N., Clark, J. W., Edwards, D., Kenrick, P., Pressel, S., et al. (2018). The timescale of early land plant evolution. *Proc. Natl. Acad. Sci. U. S. A.*, 201719588–201719588. doi: 10.1073/pnas.1719588115.

Neubauer, A., Ruaud, S., Waller, M., Frangedakis, E., Li, F.-W., Nötzold, S. I., et al. (2022). Step-by-step protocol for the isolation and transient transformation of hornwort protoplasts. *Appl. Plant Sci.* 10, e11456–e11456. doi: 10.1002/aps3.11456.

Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K. K., et al. (2018). The Chara Genome: Secondary Complexity and Implications for Plant Terrestrialization. *Cell*. doi: 10.1016/j.cell.2018.06.033.

Osteryoung, K. W., and Pyke, K. A. (2014). Division and dynamic morphology of plastids. *Annu. Rev. Plant Biol.* doi: 10.1146/annurev-arplant-050213-035748.

Ouyang, M., Li, X., Zhang, J., Feng, P., Pu, H., Kong, L., et al. (2020). Liquid-Liquid Phase Transition Drives Intra-chloroplast Cargo Sorting. *Cell* 180, 1144-1159.e20. doi: 10.1016/j.cell.2020.02.045.

Park, E., Ménétret, J.-F., Gumbart, J. C., Ludtke, S. J., Li, W., Whynot, A., et al. (2014). Structure of the SecY channel during initiation of protein translocation. *Nature* 506, 102–106. doi: 10.1038/nature12720.

Park, E., Nedo, A., Caplan, J. L., and Dinesh-Kumar, S. P. (2018). Plant–microbe interactions: organelles and the cytoskeleton in action. *New Phytol.* doi: 10.1111/nph.14959.

PITTERMANN, J. (2010). The evolution of water transport in plants: an integrated approach. *Geobiology* 8, 112–139. doi: 10.1111/j.1472-4669.2010.00232.x.

Pudelski, B., Schock, A., Hoth, S., Radchuk, R., Weber, H., Hofmann, J., et al. (2012). The plastid outer envelope protein OEP16 affects metabolic fluxes during ABA-controlled seed development and germination. *J. Exp. Bot.* doi: 10.1093/jxb/err375.

Pyke, K. A., and Leech, R. M. (1992). Chloroplast division and expansion is radically altered by nuclear mutations in Arabidopsis thaliana. *Plant Physiol.* doi: 10.1104/pp.99.3.1005.

Rahmatpour, N., Hauser, D. A., Nelson, J. M., Chen, P. Y., Villarreal A., J. C., Ho, M.-Y., et al. (2021). A novel thylakoid-less isolate fills a billion-year gap in the evolution of Cyanobacteria. *Curr. Biol.* 31, 2857-2867.e4. doi: 10.1016/j.cub.2021.04.042.

Rambaut, A., and Drummond, A. (2018). FigTree Version 1.4.4.

Rensing, S. A. (2018). Great moments in evolution: the conquest of land by plants. *Curr. Opin. Plant Biol.* doi: 10.1016/j.pbi.2018.02.006.

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* doi: 10.1111/j.2041-210X.2011.00169.x.

Richardson, L. G. L., and Schnell, D. J. (2020). Origins, function, and regulation of the TOC–TIC general protein import machinery of plastids. *J. Exp. Bot.* 71, 1226–1238. doi: 10.1093/jxb/erz517.

Rochaix, J.-D., and Bassi, R. (2019). LHC-like proteins involved in stress responses and biogenesis/repair of the photosynthetic apparatus. *Biochem. J.* 476, 581–593. doi: 10.1042/BCJ20180718.

Schön, M. E., Zlatogursky, V. V., Singh, R. P., Poirier, C., Wilken, S., Mathur, V., et al. (2021). Single cell genomics reveals plastid-lacking Picozoa are close relatives of red algae. *Nat. Commun.* 12, 6651–6651. doi: 10.1038/s41467-021-26918-0.

Schreiber, M., Rensing, S. A., and Gould, S. B. (2022). The greening ashore. *Trends Plant Sci.* doi: 10.1016/j.tplants.2022.05.005.

Shimada, H., Koizumi, M., Kuroki, K., Mochizuki, M., Fujimoto, H., Ohta, H., et al. (2004). ARC3, a chloroplast division factor, is a chimera of prokaryotic FtsZ and part of eukaryotic phosphatidylinositol-4-phosphate 5-kinase. *Plant Cell Physiol.* doi: 10.1093/pcp/pch130.

Su, D., Yang, L., Shi, X., Ma, X., Zhou, X., Hedges, S. B., et al. (2021). Large-scale phylogenomic analyses reveal the monophyly of bryophytes and Neoproterozoic origin of land plants. *Mol. Biol. Evol.* doi: 10.1093/molbev/msab106.

Sun, B., Zhang, Q. Y., Yuan, H., Gao, W., Han, B., and Zhang, M. (2020). PDv1 and PDV2 differentially affect remodeling and assembly of the chloroplast DRp5B RIng1. *Plant Physiol.* doi: 10.1104/PP.19.01490.

Thomas, R. (2001). Molecular Evolution and Phylogenetics. *Heredity* 86, 385–385. doi: 10.1046/j.1365-2540.2001.0923a.x.

Tiessen, H., Cuevas, E., and Chacon, P. (1994). The role of soil organic matter in sustaining soil fertility. *Nature* 371, 783–785. doi: 10.1038/371783a0.

Trösch, R., Töpel, M., Flores-Pérez, Ú., and Jarvis, P. (2015). Genetic and Physical Interaction Studies Reveal Functional Similarities between ALBINO3 and ALBINO4 in Arabidopsis. *Plant Physiol.* 169, 1292–1306. doi: 10.1104/pp.15.00376.

Truszkowski, J., and Goldman, N. (2016). Maximum Likelihood Phylogenetic Inference is Consistent on Multiple Sequence Alignments, with or without Gaps. *Syst. Biol.* 65, 328–333. doi: 10.1093/sysbio/syv089.

van Baren, M. J., Bachy, C., Reistetter, E. N., Purvine, S. O., Grimwood, J., Sudek, S., et al. (2016). Evidence-based green algal genomics reveals marine diversity and ancestral characteristics of land plants. *BMC Genomics* 17, 267–267. doi: 10.1186/s12864-016-2585-6.

Villarreal, J. C., and Renner, S. S. (2012). Hornwort pyrenoids, carbon-concentrating structures, evolved and were lost at least five times during the last 100 million years. *Proc. Natl. Acad. Sci.* 109, 18873 LP – 18878. doi: 10.1073/pnas.1213498109.

Vitha, S., Froehlich, J. E., Koksharova, O., Pyke, K. A., Van Erp, H., and Osteryoung, K. W. (2003). ARC6 is a J-domain plastid division protein and an evolutionary descendant of the cyanobacterial cell division protein Ftn2. *Plant Cell*. doi: 10.1105/tpc.013292.

Wang, S., Li, L., Li, H., Sahu, S. K., Wang, H., Xu, Y., et al. (2020). Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nat. Plants*. doi: 10.1038/s41477-019-0560-3.

Wietrzynski, W., Schaffer, M., Tegunov, D., Albert, S., Kanazawa, A., Plitzko, J. M., et al. (2020). Charting the native architecture of Chlamydomonas thylakoid membranes with single-molecule precision. *eLife* 9, e53740–e53740. doi: 10.7554/eLife.53740.

Wood, W. H. J., Barnett, S. F. H., Flannery, S., Hunter, C. N., and Johnson, M. P. (2019). Dynamic Thylakoid Stacking Is Regulated by LHCII Phosphorylation but Not Its interaction with PSI. *Plant Physiol.* 180, 2152–2166. doi: 10.1104/pp.19.00503.

Wood, W. H. J., MacGregor-Chatwin, C., Barnett, S. F. H., Mayneord, G. E., Huang, X., Hobbs, J. K., et al. (2018). Dynamic thylakoid stacking regulates the balance between linear and cyclic photosynthetic electron transfer. *Nat. Plants* 4, 116–127. doi: 10.1038/s41477-017-0092-7.

Xia, X. (2011). Comparative Genomics. *Handb. Stat. Bioinforma.*, 567–600.

Xu, X., Ouyang, M., Lu, D., Zheng, C., and Zhang, L. (2021). Protein Sorting within Chloroplasts. *Trends Cell Biol.* 31, 9–16. doi: 10.1016/j.tcb.2020.09.011.

Yoshida, Y., Mogi, Y., TerBush, A. D., and Osteryoung, K. W. (2016). Chloroplast FtsZ assembles into a contractible ring via tubulin-like heteropolymerization. *Nat. Plants*. doi: 10.1038/NPLANTS.2016.95.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T.-Y. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8, 28–36. doi: 10.1111/2041-210X.12628.

Žárský, J., Žárský, V., Hanáček, M., and Žárský, V. (2022). Cryogenian Glacial Habitats as a Plant Terrestrialisation Cradle – The Origin of the Anydrophytes and Zygnematophyceae Split. *Front. Plant Sci.* 12. Available at: https://www.frontiersin.org/article/10.3389/fpls.2021.735020.

Zhang, J., Fu, X. X., Li, R. Q., Zhao, X., Liu, Y., Li, M. H., et al. (2020). The hornwort genome and early land plant evolution. *Nat. Plants*. doi: 10.1038/s41477-019-0588-4.

Zhang, M., Schmitz, A. J., Kadirjan-Kalbach, D. K., TerBush, A. D., and Osteryoung, K. W. (2013). Chloroplast division protein ARC3 regulates chloroplast FtsZ-ring assembly and positioning in Arabidopsis through interaction with FtsZ2. *Plant Cell*. doi: 10.1105/tpc.113.111047.