A molecular atlas of plastid and mitochondrial evolution from algae to angiosperms

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

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

*Corresponding author
E-mail: gould@hhu.de

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 changes likely accompanied the adaptation to new ecological niches such as the terrestrial habitat. 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. 1906 and 825 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

Introduction

Fewer natural phenomena have been as transformative to planet Earth as the global greening through plants [1,2]. The proliferation of plants on 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, evolving gradually in the common ancestor of land plants (LCA), provided a decisive fitness advantage over the 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 LCA encountered. Improving 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 and maybe WHIRLY [17,18]. In combination, these adaptations were decisive for the success of streptophytes, which is evident in the number of species they have evolved and the sheer biomass they produce [1,19].

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 [20,21]. Marked metabolic and physical interactions of plastids also concern the only other 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 that are frequently in crosstalk with the photosynthetic organelle [22,23]. Like plastids, mitochondria were critical in the evolution and continued adaptation of important physiological traits, which characterize the green lineage. A notable example of preadaptation includes malate decarboxylation in the C4 photosynthetic pathway [24] – a trait of the green lineage [25] that improves plant photosynthetic efficiency in warm and dry habitats [26]. Similarly, some components of mitochondrial retrograde signaling also evolved in the land plants and likely contributed to its ROS and draught tolerance [27].

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 or plant terrestrialization. Previous work has determined genes specific to certain plant clades and that are catalogued by valuable resources such as the "GreenCut" [28]. Such analyses, however, did not focus 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 [11,29–37] and more organelle proteomes and better functional annotations are available. Similarly, and concomitantly, the development of novel and accurate algorithms for orthology inference [38–41], along with advances in experimental biology allow to now identify critical evolutionary changes 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 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, referred from here on to as the 'core GOGs' (Fig. 1B). For only 5% of all GOGs, a functional annotation could be identified (Fig. 1C, Table S1E). For embryophyte-specific GOGs the numbers were comparable, yet they maintained a consistent distribution of identified functions, except for an increased fraction of membrane trafficking proteins (Fig. 1D, Table S1F). Notably, for the core GOGs the number is higher. For 30% functional annotations covering photosynthesis, mitochondrial formation, trafficking, and information processing could be identified (Fig. 1E, Table S1G). The functions for a vast majority of the GOGs remain elusive (Table S1H), numbers that mirror those of previous studies [28], and they hence provide an excellent ground for experimental exploration.

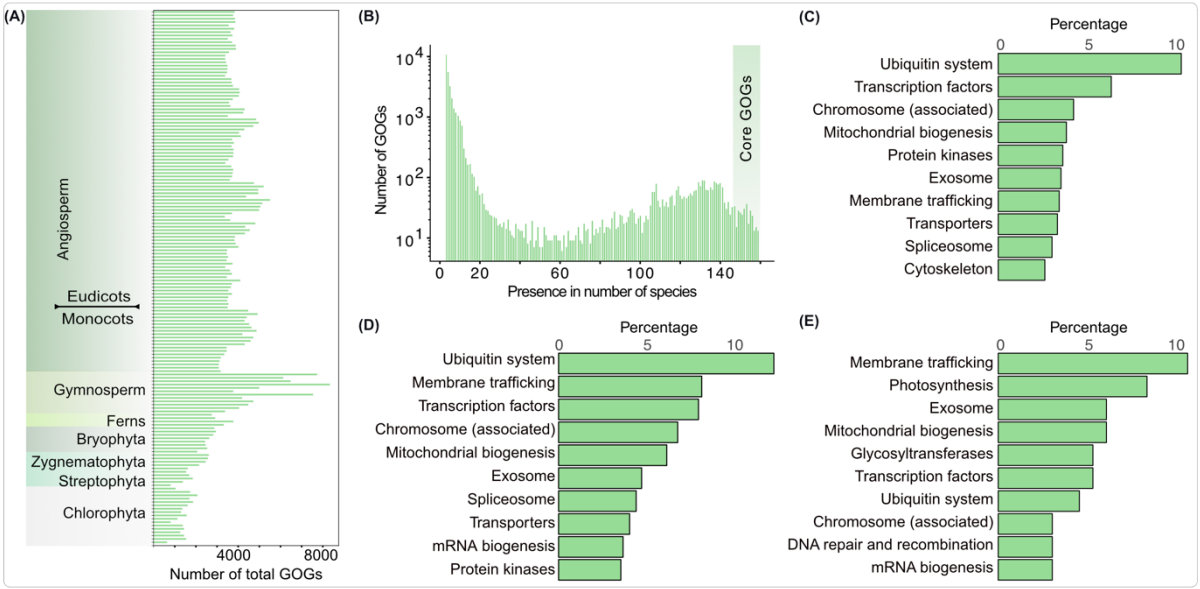

Fig. 1: Distribution and functional annotation of green orthogroups (GOGs). (A) Total number of GOGs present in each species from major Chloroplastida taxa. (B) Number of GOGs as a function of their presence

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).

Mitochondrial and plastid proteomes of the Chloroplastida expanded with the origin and diversification of the green lineage

To investigate changes in the proteomes of plastids and mitochondria, we curated 1906 plastid and 825 mitochondrial orthogroups (POGs and MOGs, respectively) based on published proteome data and homology-based protein clustering of 204 eukaryotes, including that of secondarily photosynthetic eukaryotes (Fig. S1B, Table S2A-D). In comparison to rhodophytes and glaucophytes, the green lineage encodes almost twice as many POGs (Fig. 2A, Table S2E). 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, while the latter underscores functional divergence that followed gene duplications. The distribution of MOGs appears qualitatively similar to that of POGs (Fig. 2B, Table S2F). 60% of the POGs could be functionally annotated, predominantly operating in biosynthetic and other metabolic pathways such as photosynthesis (Fig. 2C, Table S2G). Around 75% of the MOGs could be annotated, containing proteins for mitochondrial biogenesis, membrane trafficking and translation (Fig. 2D, Table S2H). Protein biosynthesis-related proteins are abundant in both, POGs and MOGs, underscoring their biosynthetic activity. Proteins for mitochondrial biogenesis also appear in both. For instance, about 60 POGs are annotated as mitochondrial biogenesis and they encompass numerous PPR and mTERF proteins (crucial for RNA editing and metabolism) and proteins involved in various other information processing activities, probable to function in both organelles. Analysis of the N-terminal 20 amino acids show their charge to range from 0 to 2, indicating they might be dually targeted to plastids and mitochondria [42]. Five of the mTERFs are part of a POG and MOG simultaneously (Fig. S3D). Overall, the trends show that in embryophytes the number of protein families associated with an endosymbiotic organelle function increased.

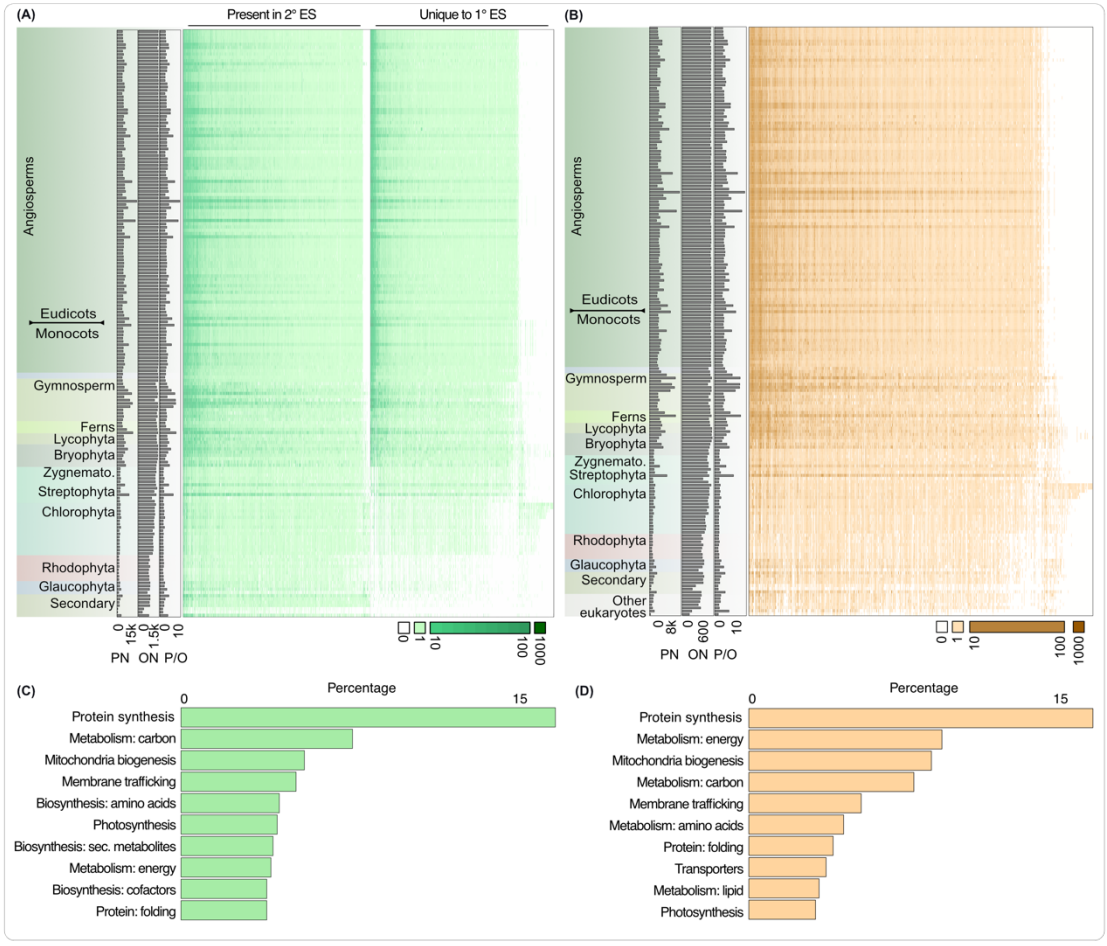

Fig. 2: Mitochondrial and plastid orthogroups across archaeplastidal species. Distribution of plastid (POGs; A) and mitochondrial orthogroups (MOGs; B). The distribution of POGs was determined for plastids of primary (1° ES) and secondary endosymbiotic origin (2° ES). Protein copy numbers within each 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 in the green ancestor, and (b) secondary losses at the origin of rhodophytes [43]. We used ancestral state reconstruction (ASR) to resolve between these two possibilities. The branching order of the archaeplastidal lineages remains challenging [44], as sometimes glaucophytes [45] and sometimes rhodophytes come out as the sister to the other remaining archaeplastidal lineages [4,46]. An inferred eukaryotic tree (with 31 non-Archaeplastida eukaryotes as an outgroup to Archaeplastida) placed the rhodophytes and glaucophytes as sister clades (Fig. S2). This tree and ASR pipeline were validated using rbcS control (Fig. S3A), and further undergirded the main results which are also consistent with varying thresholds of probability of presence and absence in a given ancestor on this eukaryotic tree (Fig. S3B) as well as Archaeplastida only phylogeny manually rooted to have glaucophyte and rhodophyte as an outgroup to chloroplastida (Fig. S4-7).

The result suggests that the plastid proteome of the last common ancestor of Archaeplastida united ca. 1000 POGs (Fig. 3A, Fig. S3B, Fig. S6, Table S3A-C). This inferred proteome witnessed significant

155 gains of protein families at the emergence of the green ancestor (and later speciation). 50% of these
156 newly gained POGs could be functionally annotated (Fig. 3C, Table S4A), showing that at the origin
157 of the green lineage novel photosynthesis- and metabolism-related POGs were recruited, while the
158 transition to land (Z/E and embryophyte ancestors) added metabolism-related, as well as protein
159 synthesis- and ubiquitin-related POGs to the toolkit (Table S4A). Using hidden Markov searches, we
160 verify that more than half of the protein families recruited in embryophyte and Z/E ancestors are absent
161 in non-zygnematophyceae algae (Fig. S3C). The mitochondrial proteome followed a qualitatively
162 similar trend of expansion (Fig. 3B, Fig. S3B, Fig. S7, Table S3D-F). ca. 500 MOGs trace back to the
163 archaeplastidal ancestor, while ca. 700 MOGs were identified at the root of angiosperms (Fig. 3C, Fig.
164 S3B). Around 50% of the newly gained MOGs could be functionally annotated, showing that the
165 chloroplastidal gains contribute to carbon metabolism, protein synthesis and mitochondrial biogenesis.
166 Terrestrialization also witnessed a similar gain of MOGs, most of which function in metabolism as well
167 as mitochondrial biogenesis and membrane trafficking (Fig.3C, Table S4B).

168 In summary, across plant species, plastid and mitochondrial proteomes are predicted to have gained a
169 significant number of protein families reflecting the dynamic nature of organellar proteomes post-
170 endosymbiosis [47,48]. A closer look at the function of the newly gained organelle proteins shows a
171 wide variety, including lipid and carbon metabolism, information processing, development and division
172 of organelles.

173

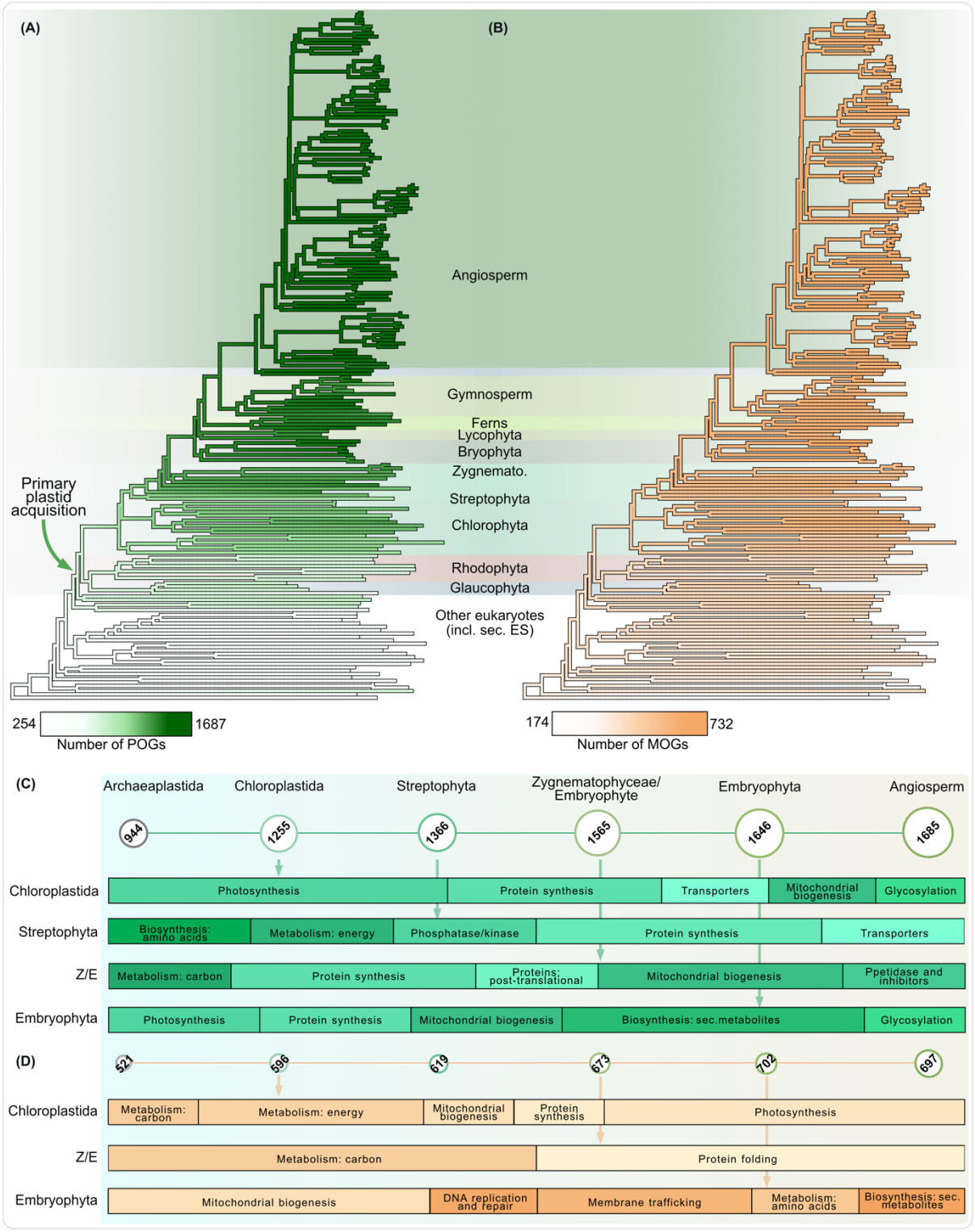

Fig. 3: Evolution of organelle proteomes in Archaeplastida. Gains in plastid (POGs; A) and mitochondrial orthogroups (MOGs; B) across all nodes of archaeplastidal evolution and POGs coinciding with primary and secondary plastid acquisitions. Gains across main nodes of interest in (C), where each circle represents an ancestor, with its predicted number of protein families shown in the circle and whose diameter correlates with the number of OGs. Major gains occurred in the chloroplastidal ancestor, the common ancestor of Zygnematophyceae-embryophytes (Z/E) and that of embryophytes. In (D) the same as in (C), but for mitochondrial OGs. Their functions are shown in the proportionate bar charts below the ancestors.

183

184 **Increased complexity of RNA metabolism and photosynthetic adaptability**

185 RNA metabolisms such as editing intercepts the linear information flow from mRNA to protein and is
186 crucial for organelles to function [49–51]. Two main domains, the PPR and mTERF domain, are
187 associated with RNA editing and metabolism [52,53]. We first screened for organelle orthogroups
188 containing either of these two domains in at least 60% of all proteins within each respective orthogroup
189 (Fig. S1C). Around 50 POGs and 20 MOGs were found. More than 80% of them were restricted to
190 embryophytes, only few were present in some algae (Fig. 4). A closer look revealed that most of the
191 algal homologues lacked PPR and mTERF domains and they are hence unlikely true orthologues. More
192 generally, this shows that any detailed interpretation regarding an inferred orthogroup's function should
193 be supported by screening for functionally relevant domains.

194 True PPR or mTERF domain-containing RNA-editing, splicing (and processing at large) proteins
195 increased significantly in number by recruiting new orthogroups, also through adding the two domains
196 to proteins that did not contain these in their algal ancestor. A presence-absence pattern shows that
197 >90% of proteins containing PPR/mTERF domains are exclusive to land plants, except for
198 *Chara braunii* and *Klebsormidium flaccidum* (Fig. 4B). These proteins include, but are not limited to,
199 OTP51 and SOT5 (present in embryophytes and *Chara*) as well as SOT1, SVR7, THA8, PDM4 (present
200 only in embryophytes; Fig. S9). Target transcripts of these RNA metabolism factors point to the
201 synthesis and assembly of photosynthesis-related proteins and to proteins of the thylakoid membrane
202 (Fig. 6B). Likewise, mTERFs, which are crucial for plastid and leaf development, are also uniquely
203 expanded in the terrestrial clade with examples of protein re-targeting across organelles [54]. The dual
204 targeted (plastid and mitochondrion) mTERF6, unique to the land plants (Fig. S9) and the streptophyte
205 alga *Klebsormidium*, takes part in retrograde signalling to the nucleus via ABA and imparts abiotic
206 stress tolerance [55]. Overall, RNA metabolism across plants has undergone major changes and has a
207 significant impact on photosynthesis, improvement of which was key to thriving on land (Fig. 6B).

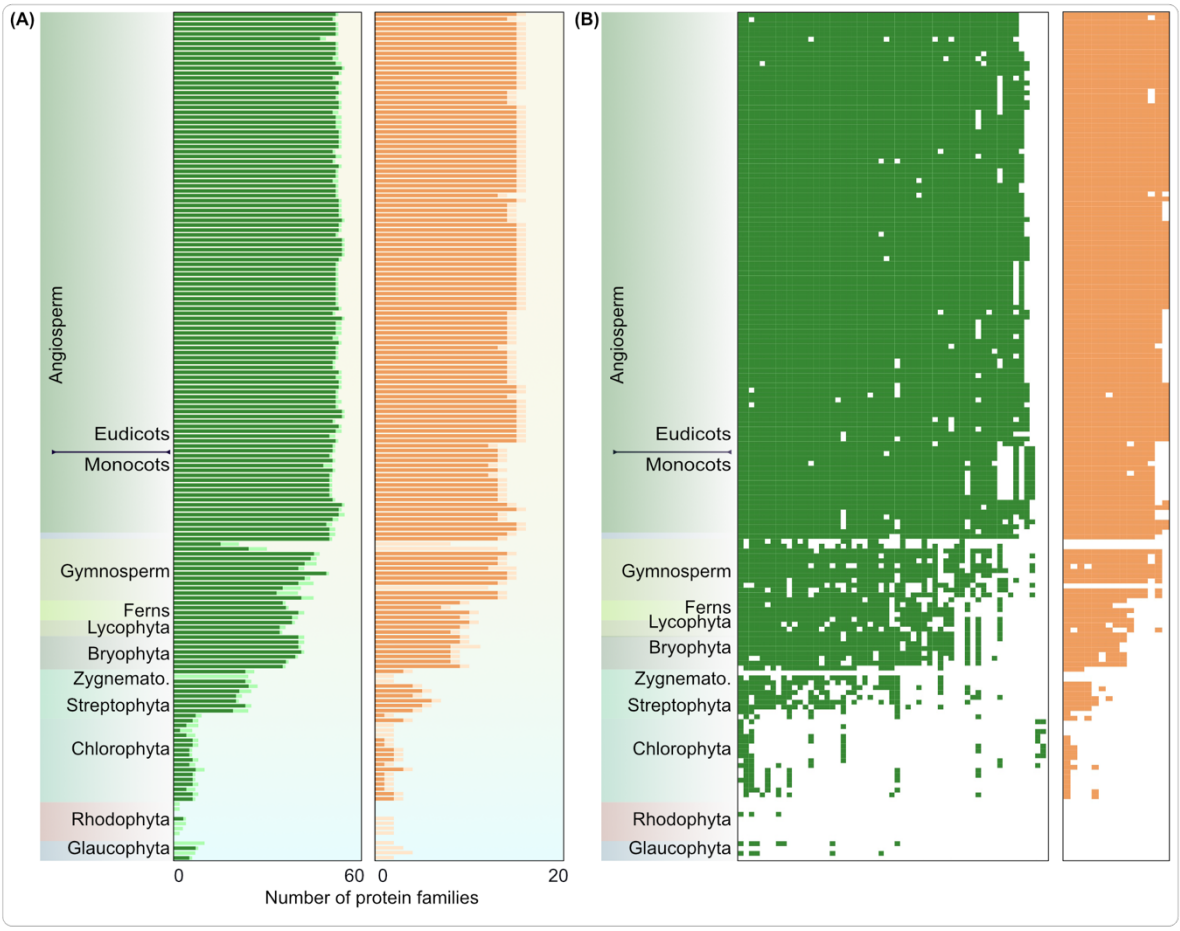

Fig. 4: Recruitment of PPR and mTERF domains in organelle proteins. (A) Number of POGs (left) and MOGs (right) where at least one protein contains a PPR/mTERF domain is shown in bars with dark shades of colors. Total number of orthogroups (regardless of presence or absence of PPR/mTERF domain in that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

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 [56], evolved as one countermeasure [57]. We find that cutin and suberin evolution was enabled by the recruitment of an organelle-specific GPAT (Glycerol-3-phosphate acyltransferases) family in the embryophyte ancestor (Fig. 5), which includes GPAT1 (mitochondrial), GPAT 4,6 and 8 of the endoplasmic reticulum [58,59]. Trafficking of these lipids across organelles was made possible by a dual targeted TGD4 [60] that was recruited in the chloroplastida 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. S9). Duplication and divergence of ACP desaturases in embryo- and spermatophytes played an important role in regulating lipid composition shifts in response to temperature and drought, the regulation of seed oil content and development [61]. Likewise, acyl-carrier proteins also increased in copy number (Fig. S9) and adapted towards a light-induced expression and regulation of the seed fatty acid content [62,63]. 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. S9). AGPases diverged to regulate starch metabolism under osmotic and light stress, as well as the differential regulation of starch synthesis and degradation [64–68]. Another key regulatory enzyme, PGI (phosphoglucose isomerase) evolved a distinct family (PGI1) in Zygnematophyceae (Fig. S9). 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 [69]. PTST3 also emerged around the time of terrestrialization (Fig. S9), which evolved to regulate starch synthesis with significant impact on plastid development [70]. In contrast to the flow of carbon through glycolysis, GSM2 (which originated in streptophytes; Fig. S9), shunts carbon towards the pentose-phosphate pathway and protects plastids from oxidative stress in *Arabidopsis* [71].

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 [12,13,22]. We screened for the origin and diversification of EXECUTOR and SVR4, both components of retrograde signalling. We also screened for WHIRLY, a protein family that acts on RNA splicing and ribosome biogenesis, but also relocates between compartments and remains a candidate for retrograde signalling [18,72–76] . EXECUTOR, key to regulating retrograde signalling, oxygen and light stress regulation [77–79], originated in the ancestor of the Chloroplastida and so did WHIRLY (Fig. 5); the latter underwent copy number expansion in embryophytes and was likely lost in some bryophytes (Fig. S9). 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 [80–82]. These emergence of the Chloroplastida was marked by the two components EXECUTOR and WHIRLY. Additional paralogs evolved, each with a specific function in the main green lineages, and they likely aided in the colonization of the terrestrial habitat by the ancestor of land plants (Fig. 6B).

SVR4, a dual targeted (plastid and nucleus) recruited around terrestrialization (Fig. 5), likely communicates required gene expression changes needed for light-induced plastid development, thylakoid stacking and thermomorphogenesis [83,84]. 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 associated with the the 3′-phosphoadenosine-5′-phosphate family, which facilitated the emergence of stomatal closing in land plants [85]. The recruitment and diversification of the proteins we highlight were quintessential for responding to two major stressors that are more pronounced and more rapidly changing on land than in water: light and temperature (Fig. 6B).

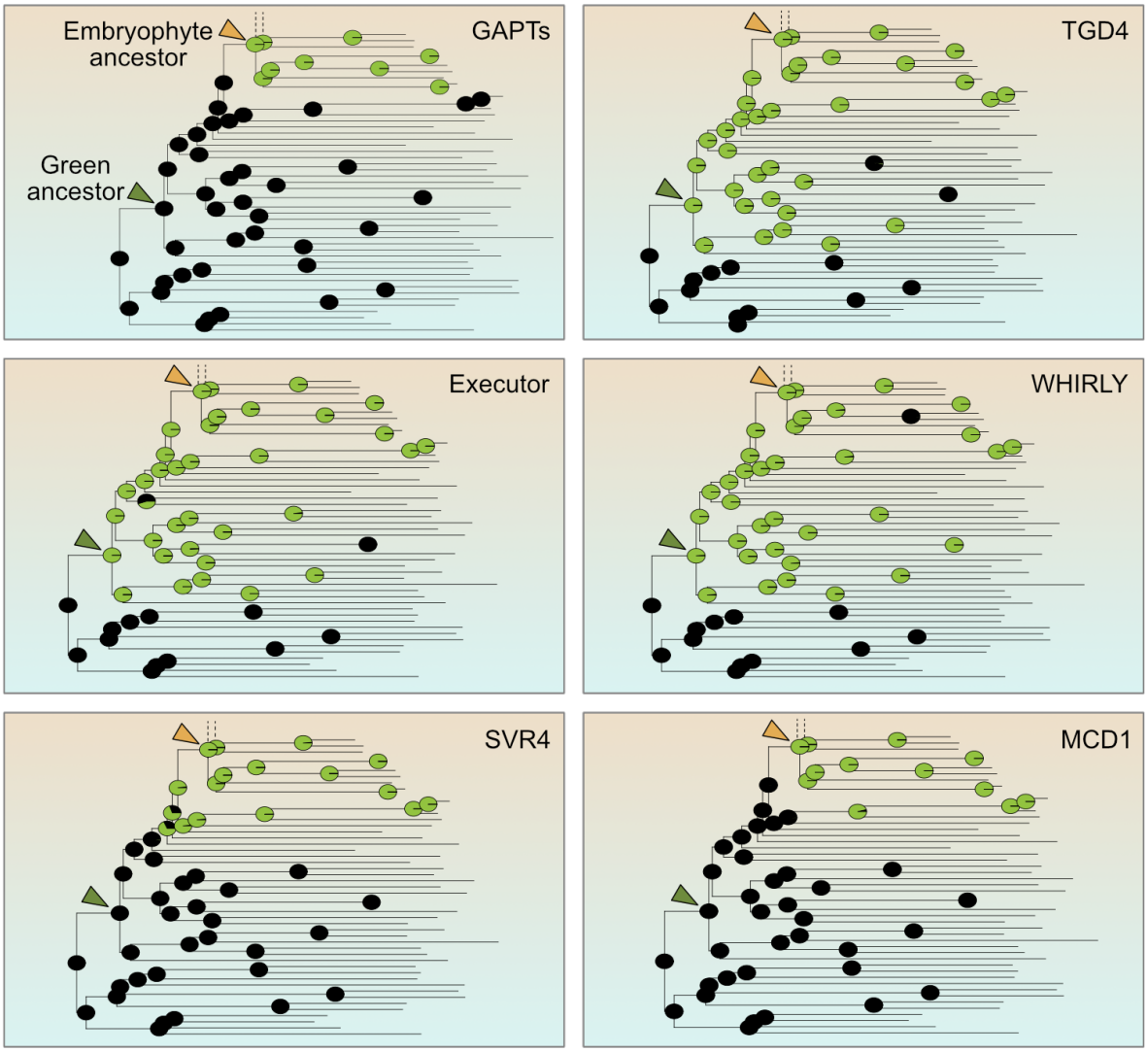

Fig. 5: Origins of key proteins involved in metabolism, communication and development. Ancestor state reconstruction (ASR) for selected lipid metabolism (GAPT and TGD4), retrograde signalling (Executor and Whirly), plastid development (SVR4) and division (MCD1) related proteins. The pie charts at each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

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 [86–88]. Polyplastidy is a trait of land plants and many macroscopic algae such as *Bryopsis* or *Chara* [89] and known molecular determinants include MinD, MinE, ARC3 and the FtsZ proteins [16,87]. Our data supports that MULTIPLE CHLOROPLAST DIVISION SITE 1 (MCD1), a core component of the plastid division machinery [90], originated in the ancestral embryophyte (Fig. 5). The cotyledon chloroplast biogenesis factor CYO1 and the transcriptionally active chromosome factor 7 (TAC7) are important components of thylakoid biogenesis and the plastid translation machinery, respectively. Both originated in the streptophyte ancestor (Fig. S9) and, in *Arabidopsis*, play key roles in chloroplast, cotyledon, thylakoid and leaf development [91–93]. Lastly, CRUMPLED LEAF (CRL), a protein residing in the outer plastid membrane, emerged during terrestrialization, too (Fig. S9), likely for regulating plastid division and securing correct plastid inheritance during embryogenesis [94,95].

294 Crucial for plastid biogenesis, especially in light of an expanding proteome, is the import of proteins.
295 The membrane GTPase TOC159 is essential for chloroplast biogenesis via the selective recognition and
296 import of the photosynthetic proteins [96] and is unique to the green lineage (Fig. S9). The membrane
297 recruitment of this protein requires TOC75, of which a special variant evolved in the green ancestor
298 after the duplication of OEP80 [14,97]. The copy number of TOC159 expanded from the
299 Zygnematophyceae onwards (Fig. S9), hinting at its functional diversification. Unlike in the
300 chlorophyte alga *Chlamydomonas*, land plant TOC159 homologues possess an N-terminal acidic
301 domain that gets phosphorylated to alter substrate specificity [96,98]. Furthermore, TOC159, along with
302 TOC132 and TOC120, play important roles in regulating plastid lipid synthesis and membrane fluidity
303 and in *Arabidopsis* show tissue specific expression (The Arabidopsis Information Resource) [99–101].
304 Further on the course of evolution, the J-domain-containing protein TOC12 [102] was likely recruited
305 in the ancestral embryophyte for supporting the import machinery at the intermembrane space (Fig. S9).
306 The terrestrial habitat demands a highly efficient and fluid import of proteins, for example upon high
307 light and other abiotic stresses [14,103]. The expansion of the TOC/TIC system in the embryophyte
308 ancestor reflects how the organelle dealt with an ever-increasing diversity of substrates that were
309 required to be processed.
310

311 **Discussion**

312 The settling of land by a streptophyte alga and the subsequent evolution and spreading of plants (Fig.
313 6A) was pivotal in the transformation of the terrestrial habitat and it laid the foundation for the
314 concomitant evolution and diversification of animals [1,2]. Throughout the hundreds of millions of
315 years of plant evolution, both organelles of endosymbiotic origin underwent a multitude of molecular
316 adaptations, hereby evolving into the plastid and mitochondrion of modern plants. We identified 31,650
317 protein families unique to the green lineage, approximately 50% of which are unique to embryophytes.
318 It demonstrates an expansion and divergence of protein families at the time of plant terrestrialization
319 and in line with a recent study that identified around 10,000 duplications at the birth of embryophytes
320 [104].

321 Expansion of proteins families is evident in both organellar proteomes at the origin of the green lineage
322 itself and at the water-to-land transition. The gain of protein families at the origin of the Chloroplastida
323 needs to be treated with caution due to the documented genetic bottleneck that characterizes rhodophyte
324 origin [105–109] and the sparse availability of glaucophyte genome data. Some of the newly recruited
325 protein families at the origin of the green lineage might rather be explained by a loss in rhodophytes
326 and a retention in the chloroplastidal ancestor instead of a gain. Regardless, this has little bearing on the
327 biological significance of a given protein family with respect to the overall increase in complexity of
328 organelle biology – both concerning the variety as well as the number of proteins targeted to plastids
329 and mitochondria – throughout streptophyte evolution. It affected the organelles metabolic,
330 informational and developmental complexity, and facilitated the evolutionary successful transition from
331 water to land more than 500 million years ago (Fig. 6).

332 Changes in organelle lipid biochemistry contributed to one of the key adaptations in land plants that is
333 the cuticle. Land plant GPATs (Glycerol-3-phosphate acyltransferases; crucial to lipid synthesis for
334 cutin and suberin) contribute to increased hydrophobicity and water retention in embryophytes [57] and
335 their activity in embryophytes differs from that in algae [110,111]. Our analyses pinpoint the origins of
336 organelle specific GPATs (GPAT 1,4,6, and 8) to the embryophyte ancestor, and of which deleting
337 GPAT4 and GPAT8 distorts cuticles and increases water loss by several fold [58,59]. In parallel, lipid
338 trafficking was mediated by the recruitment or divergence of proteins such as TGD4 and acyl carrier
339 thioesterases, which contributed to wax biosynthesis and deposition on leaves, cuticle development,

340 thylakoid membrane stacking [60], seed development and germination [61]. As for starch metabolism,
341 the archaeplastidal ancestor likely stored starch in the cytosol [112], but the red and green lineage
342 experienced different fates from there on. Rhodophytes continued to store starch in the cytosol in the
343 form of Floridean starch [113], while in the green lineage, particularly in complex plants, more localized
344 control of starch synthesis and degradation was facilitated by changes in regulatory proteins (eg
345 AGPase). Together, organelle metabolism evolved to serve key roles in the synthesis, regulation and
346 trafficking of lipids involved in wax coating to prevent water loss in the land plant ancestor, as well as
347 synthesis and storage of starch (Fig. 6D).

348 RNA processing and editing is a crucial component of information processing and overall functionality
349 of plant organelles [49,50]. Changes in RNA metabolism are evident from the origin of the green lineage
350 itself, where RNAse-P (tRNA maturation) was replaced by protein only RNAse P or PROPs [114,115].
351 Subsequent expansion of PROPs in embryophytes (Fig. S9) led to organelle-localised copies, of which
352 some are essential for maintaining organelle morphology, function and plant viability [116].
353 Components associated with plastid encoded RNA polymerase (PEP associated proteins, PAPs) also
354 show a gradual recruitment from the green ancestor to embryophyte ancestor (Fig. S8). RNA editing
355 of C to U is not found in algae, however, and editing sites in embryophytes are unlike those of any other
356 eukaryote, suggesting they emerged independently [50]. Of the many RNA-metabolism proteins we
357 find that were gained during terrestrialization, known targets are transcripts involved in photosynthesis
358 and stress tolerance-related transcripts, both key to colonising the land (Fig. 6B). For instance, THA8,
359 PDM4, SVR7 and SOT1 associate with transcripts such as ycf2 and ycf3, and contribute to thylakoid
360 development and biogenesis [117], the generation of photosynthetic complex proteins, grana stacking,
361 and embryo and plastid development [117,119,120]. OTP51 and SOT5 splice transcripts related to
362 chlorophyll synthesis, photosynthesis and thylakoid membranes (ycf3, TRNK and RPL2) [121–123],
363 whereas DOG1 is important for high temperature response and chloroplast development [124]. This
364 elaborate RNA processing in organelles, especially plastids, appears to serve photosynthesis (and
365 thylakoid) related transcripts. It is feasible that by benefitting photosynthesis, organelle RNA editing
366 continued to be positively selected for during terrestrialization and was expanded.

367 One evolutionary step towards efficient photosynthesis, where RNA editing also plays a key role, are
368 grana stacks [86]. The evolutionary origin of grana remains elusive, along with the underlying
369 developmental pathways involved in regulating its formation and maintenance [86,125,126]. Highly
370 organized grana stacks are observed in embryophytes and some Zygnematophyceae (e.g. the
371 *Cosmarium* genus) [127], but not chlorophytes such as *Chlamydomonas* [128]. We noticed a patchy
372 distribution of grana morphology associated proteins such as CURT1, RIQ1 and RIQ2 (Fig. S9), with
373 both RIQs being present in all streptophytes and some chlorophytes but excluding *Chlamydomonas*. In
374 light of the many key adaptations in Zygnematophyceae discussed here and elsewhere [11,129], we
375 speculate that a sophisticated stacking of grana originated in streptophytes and was beneficial for
376 thriving on land through photosynthesis optimization, in particular with respect to photosystem repair
377 and the separation of the photosystems and the ATP synthase [130,131].

378 This expansion of an organelle proteome necessitates improving the capacity to import proteins.
379 Changes in some import receptors within the green lineage and in targeting sequences at its origins are
380 known, with phosphorylation likely emerging as a key regulator for sorting the newly expanded
381 proteome differentially to plastid and mitochondria (Fig. 6C) [14,42]. Despite such adaptations, protein
382 sorting is never perfect and some mistargeting might be positively selected for. A regulated distribution
383 of newly recruited proteins (e.g. WHIRLY, TGD4, mTERF6; Fig. 6B) to multiple organelles (with
384 distinct organellar functions) hints at adaptive values of this apparent mis-sorting. How many of newly
385 recruited proteins get 'mis-sorted' owing to biological adaptability versus stochasticity remains to be

386 explored together with obtaining a more comprehensive picture of (regulatory) mechanisms associated
387 with sorting in general.

388 Embryophyte cells target proteins not to a single plastid, but many simultaneously. The presence of
389 multiple plastids per cell, polyplastidy, in the green lineage, evolved in an embryophyte ancestor, maybe
390 the common ancestor of embryo– and charophytes, likely through changes in plastid fission and a
391 decoupling of the latter from the cell cycle [15,16]. We find that MCD1, a core regulator of the plastid
392 division proteins FtsZ2 and ARC3, emerged in the embryophyte ancestor, which corroborates the idea
393 of a mono- to polyplastidy switch during the land transition of green algae [16,89,132,133]. A change
394 in the copy number of plastids also requires a mechanism that maintains a functional organelle to cell
395 volume ratio and resource allocation (Fig. 6C). The REDUCED CHLOROPLAST COVERAGE (REC)
396 protein is involved in such a mechanism in *Arabidopsis* [134] and the phylogenetically related protein
397 FRIENDLY regulates the distribution of mitochondria, also in plants and non-photosynthetic organisms
398 [135,136]. REC and FRIENDLY share almost all of their domains. How they exactly function and
399 differentiate between the two organelles remains elusive. From what we can tell, FRIENDLY emerged
400 during eukaryogenesis and the origin of mitochondria. REC we can trace back to the streptophyte
401 ancestor (Fig. S9) and after a likely duplication event of FRIENDLY. We speculate that the origin of
402 REC helped to cement polyplastidy, which itself supports larger body plans and the diversification of
403 different plastid types [15]. Lastly, an increase in organelle copy number also requires an overall
404 increase in the capacity to synthesize proteins. The largest fraction of organelle proteins operate in
405 tRNA, amino acid and ribosomal biosynthesis and undergird the biosynthetic capacity of organelles, an
406 adaptation strategy akin to their bacterial ancestor [137,138].

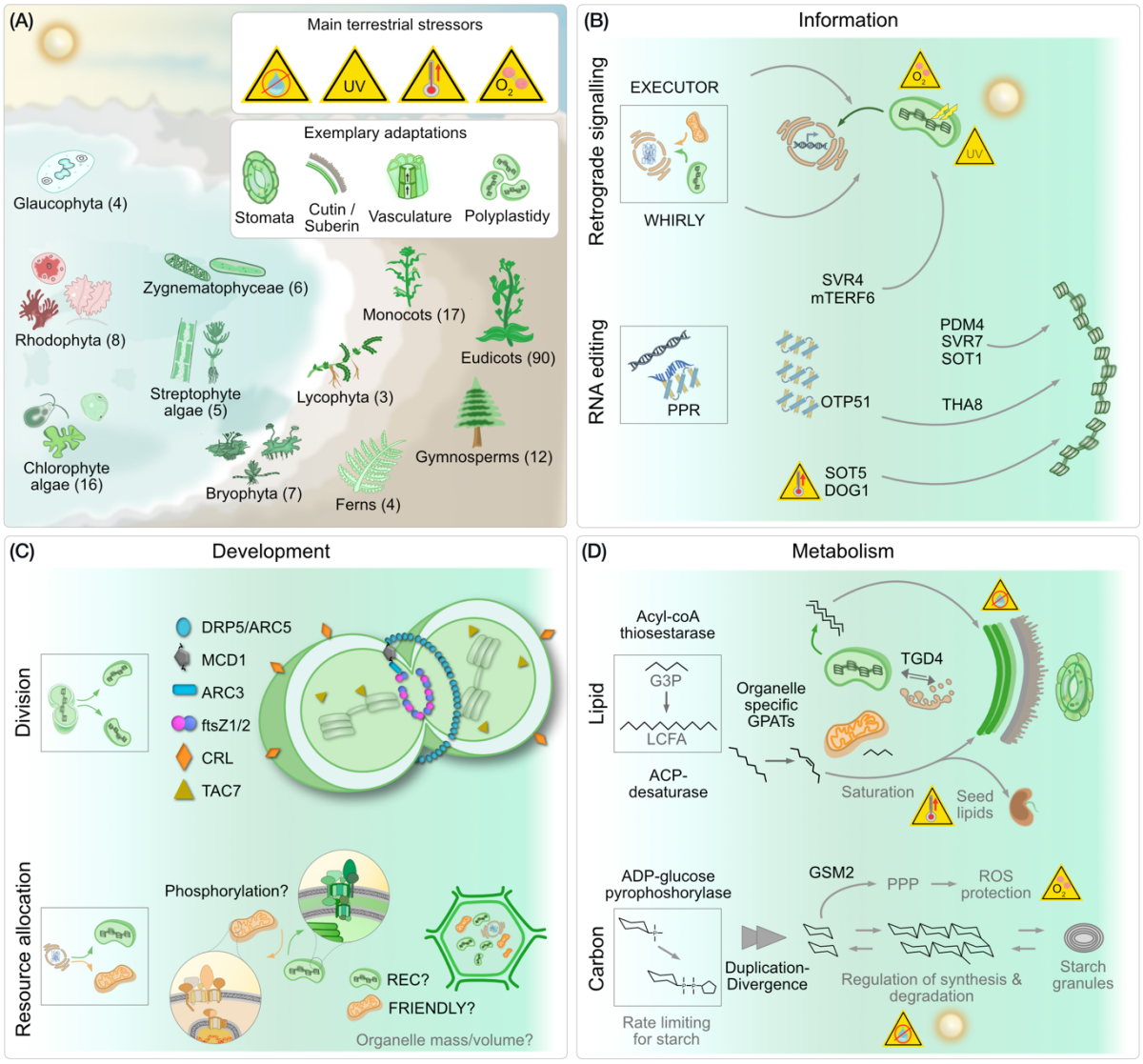

Fig. 6: The global greening and endosymbiotic organelles. (A) After the endosymbiotic origin of the plastid, three aboriginal lineages emerged that form the Archaeplastida: the glaucophytes, rhodophytes and chlorophytes. From the latter, streptophyte algae evolved, including the zygnematophyceae, that 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 that was 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 is associated with the origin of the endomembrane system and necessitated the emergence of eukaryotic traits including mito- and autophagy [139–141]. Our analyses show that the integration of a subsequent endosymbiont, the plastid, coincided with the emergence of proteins that work for the endomembrane system. Salient are changes in the ubiquitin system during terrestrialization, when polyplastidy in the green lineage also emerged (Table S2G). Ubiquitination is key to proteosome-mediated degradation and is performed chiefly by the E3 ubiquitin ligase family, which are important in land plants also for photomorphogenesis[142]. RING (Really interesting new gene) E3 ligases contribute to growth, development and stress response

427 via also mediating protein-protein interactions [143–146]. We trace a number of RING finger and
428 related proteins to terrestrialization (Fig. S9) that include, but are not limited to,
429 *DAL1* and *DAL2* (for *Drosophila* DIAP1 like 1 and 2), KEG (Keep on going), and NIP1 and NIP2.
430 *DAL1* and *DAL2* play a key role in regulation of programmed cell death [147], peroxisome and
431 chloroplast biogenesis [148–150]. KEG contributes to stress mitigation [151,152], while NIP1 and NIP2
432 play a role in plastid development by docking plastid RNA polymerase to the thylakoid membrane
433 [153]. The regulated degradation of plastids and other changes in the endomembrane system are a
434 prerequisite for housing multiple plastids per cell and we find many more recruitments broadly affiliated
435 with the endomembrane system, with poorly characterised functions until now. Exploring the functions
436 of these proteins will add valuable insights into the cell biological changes that endosymbiosis
437 stipulates.

438 In closing, although experimentally reported plant plastid and mitochondrial proteomes are scarce, we
439 were able to generate a first comprehensive molecular atlas of the changes of plastid and mitochondrial
440 protein families in the evolution of the green lineage. Ancestral state reconstruction (ASR) allows to
441 map the organelle transformations that facilitated the major transitions such as terrestrialization and
442 which will improve with every new proteome that is added. By inferring plastid and mitochondrial
443 proteomes for 173 species, we set testable expectations for new proteomes to come and provide a solid
444 database, where origins and across species orthologues of any known (organelle) protein can be
445 searched (Table S2C-D). Additional proteomes, once available, will likely solidify the general pattern
446 observed and uncover more lineage-specific curiosities. We identify numerous mitochondrial protein
447 recruitments, whose physiological roles and adaptive values help to better understand plant
448 mitochondrial biology. For plastid proteins, we infer their functions and physiological importance based
449 on the extensively studied *Arabidopsis* system. Utilizing an advanced orthology search technique [40],
450 we postulate that orthologues of *Arabidopsis* are likely to exhibit similar functions in other species. Our
451 methodologically robust approach maps various changes in evolution, associated in particular with
452 terrestrialization, that can now be experimentally explored across selected models and with a focus on
453 less-well studied streptophyte algal and bryophyte species [154,155].

454

455

456 **Conclusions**

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

469

470

Material and Methods

Curation of green orthogroups (GOGs). Input protein sequences from 686 proteomes (from KEGG [156]and Phytozome [29], Table S1A) were clustered using Orthofinder version 2.5.4 [40], after all vs all blasts were conducted (E-value cutoff 10e-10) using diamond blast version2.011 [38]. 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. (Inhouse python script used for this, and other data processing, are available on Github as repository 'Molecular-Atlas-of-plant-organelle-evolution'.)

Curation of plastid and mitochondria orthogroups (POGs and MOGs). 5,452,977 proteins from 204 eukaryotes (Table S2A) 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* [157], *P. patens* [158] , *Z. mays* [159], *A. thaliana* [159], 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 [160], 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 [161] which places Rhodophyta as the sister to all others. Independently, the same unrooted phylogeny was manually rooted using FigTree (v1.4.4) [162] such that Glaucophyta were at the base. Ancestor state of presence and absence of organelle protein families across nodes, were inferred using Phytool [163] package 0.7.80. Based on character state at the tips of the tree, Phytool inferred Bayesian posterior probabilities under a single rate model [164,165] 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 metabolism POGs and MOGs. Hidden Markov models (HMM) of PPR and mTERF domains were downloaded from pFAM [166] 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 metabolism OG. Origin of such OGs were traced using ASR as described above.

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 (HHU Düsseldorf) for his help. We thank Alice Barkan (University of Oregon) for her useful feedback on our bioXRiv preprint and three anonymous reviewers for their time evaluating and improving our manuscript.

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

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. Foyer CH, Karpinska B, Krupinska K. The functions of WHIRLY1 and REDOX-RESPONSIVE TRANSCRIPTION FACTOR 1 in cross tolerance responses in plants: a hypothesis. Philosophical Transactions of the Royal Society B: Biological Sciences. 2014;369: 20130226. doi:10.1098/rstb.2013.0226

19. 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

20. 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

21. 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

22. 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

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

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

599 25. Peers G, Niyogi KK. Pond scum genomics: the genomes of Chlamydomonas and
600 Ostreococcus. Plant Cell. 2008;20: 502–507. doi:10.1105/tpc.107.056556

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

603 27. Khan K, Van Aken O. The colonization of land was a likely driving force for the evolution of
604 mitochondrial retrograde signalling in plants. Jones M, editor. Journal of Experimental Botany.
605 2022;73: 7182–7197. doi:10.1093/jxb/erac351

606 28. Heinnickel ML, Grossman AR. The GreenCut: re-evaluation of physiological role of
607 previously studied proteins and potential novel protein functions. Photosynthesis Research.
608 2013;116: 427–436. doi:10.1007/s11120-013-9882-6

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

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

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

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

621 33. Lang D, Ullrich KK, Murat F, Fuchs J, Jenkins J, Haas FB, et al. The Physcomitrella patens
622 chromosome-scale assembly reveals moss genome structure and evolution. The Plant Journal.
623 2018;93: 515–533. doi:10.1111/tpj.13801

624 34. Nishiyama T, Sakayama H, de Vries J, Buschmann H, Saint-Marcoux D, Ullrich KK, et al.
625 The Chara Genome: Secondary Complexity and Implications for Plant Terrestrialization. Cell.
626 2018. doi:10.1016/j.cell.2018.06.033

627 35. Li F-W, Nishiyama T, Waller M, Frangedakis E, Keller J, Li Z, et al. Anthoceros genomes
628 illuminate the origin of land plants and the unique biology of hornworts. Nature Plants. 2020;6:
629 259–272. doi:10.1038/s41477-020-0618-2

630 36. Wang S, Li L, Li H, Sahu SK, Wang H, Xu Y, et al. Genomes of early-diverging streptophyte
631 algae shed light on plant terrestrialization. Nature Plants. 2020. doi:10.1038/s41477-019-0560-
632 3

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

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

39. 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

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

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

42. 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

43. 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

44. 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

45. 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

46. 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

47. 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

48. 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

49. 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

50. 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

51. Barkan A. Expression of Plastid Genes: Organelle-Specific Elaborations on a Prokaryotic Scaffold. Plant Physiology. 2011;155: 1520–1532. doi:10.1104/pp.110.171231

52. Barkan A, Small I. Pentatricopeptide Repeat Proteins in Plants. Annu Rev Plant Biol. 2014;65: 415–442. doi:10.1146/annurev-arplant-050213-040159

674 53. Wobbe L. The Molecular Function of Plant mTERFs as Key Regulators of Organellar Gene
675 Expression. Plant and Cell Physiology. 2021;61: 2004–2017. doi:10.1093/pcp/pcaa132

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

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

681 56. Wu L, Zhou Z-Y, Zhang C-G, Chai J, Zhou Q, Wang L, et al. Functional Roles of Three Cutin
682 Biosynthetic Acyltransferases in Cytokinin Responses and Skotomorphogenesis. PLOS ONE.
683 2015;10: e0121943.

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

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

691 59. Fernández-Santos R, Izquierdo Y, López A, Muñiz L, Martínez M, Cascón T, et al. Protein
692 Profiles of Lipid Droplets during the Hypersensitive Defense Response of Arabidopsis against
693 Pseudomonas Infection. Plant and Cell Physiology. 2020;61: 1144–1157.
694 doi:10.1093/pcp/pcaa041

695 60. Xu C, Fan J, Cornish AJ, Benning C. Lipid Trafficking between the Endoplasmic Reticulum
696 and the Plastid in Arabidopsis Requires the Extraplastidic TGD4 Protein. The Plant Cell.
697 2008;20: 2190–2204. doi:10.1105/tpc.108.061176

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

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

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

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

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

714 66. Crevillén P, Ventriglia T, Pinto F, Orea A, Mérida Á, Romero JM. Differential Pattern of
715 Expression and Sugar Regulation of Arabidopsis thaliana ADP-glucose Pyrophosphorylase-

716 encoding Genes*. Journal of Biological Chemistry. 2005;280: 8143–8149.
717 doi:https://doi.org/10.1074/jbc.M411713200

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

721 68. Pourtau N, Jennings R, Pelzer E, Pallas J, Wingler A. Effect of sugar-induced senescence on
722 gene expression and implications for the regulation of senescence in Arabidopsis. Planta.
723 2006;224: 556–568. doi:10.1007/s00425-006-0243-y

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

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

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

734 72. Melonek J, Mulisch M, Schmitz-Linneweber C, Grabowski E, Hensel G, Krupinska K.
735 Whirly1 in chloroplasts associates with intron containing RNAs and rarely co-localizes with
736 nucleoids. Planta. 2010;232: 471–481. doi:10.1007/s00425-010-1183-0

737 73. Lepage E, Zampini E, Brisson N. Plastid Genome Instability Leads to Reactive Oxygen
738 Species Production and Plastid-to-Nucleus Retrograde Signaling in Arabidopsis. PLANT
739 PHYSIOLOGY. 2013;163: 867–881. doi:10.1104/pp.113.223560

740 74. Desveaux D, Subramaniam R, Després C, Mess J-N, Lévesque C, Fobert PR, et al. A "Whirly"
741 Transcription Factor Is Required for Salicylic Acid-Dependent Disease Resistance in
742 Arabidopsis. Developmental Cell. 2004;6: 229–240. doi:10.1016/S1534-5807(04)00028-0

743 75. Comadira G, Rasool B, Kaprinska B, García BM, Morris J, Verrall SR, et al. WHIRLY1
744 Functions in the Control of Responses to Nitrogen Deficiency But Not Aphid Infestation in
745 Barley. Plant Physiol. 2015;168: 1140–1151. doi:10.1104/pp.15.00580

746 76. Ren Y, Li Y, Jiang Y, Wu B, Miao Y. Phosphorylation of WHIRLY1 by CIPK14 Shifts Its
747 Localization and Dual Functions in Arabidopsis. Molecular Plant. 2017;10: 749–763.
748 doi:10.1016/j.molp.2017.03.011

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

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

756 79. Uberegui E, Hall M, Lorenzo Ó, Schröder WP, Balsera M. An Arabidopsis soluble chloroplast
757 proteomic analysis reveals the participation of the Executer pathway in response to increased

758 light conditions. Journal of Experimental Botany. 2015;66: 2067–2077.
759 doi:10.1093/jxb/erv018

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

763 81. Krupinska K, Desel C, Frank S, Hensel G. WHIRLIES Are Multifunctional DNA-Binding
764 Proteins With Impact on Plant Development and Stress Resistance. Frontiers in Plant Science.
765 2022;13. Available: https://www.frontiersin.org/articles/10.3389/fpls.2022.880423

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

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

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

775 85. Wang S-W, Li Y, Zhang X-L, Yang H-Q, Han X-F, Liu Z-H, et al. Lacking chloroplasts in
776 guard cells of crumpled leaf attenuates stomatal opening: both guard cell chloroplasts and
777 mesophyll contribute to guard cell ATP levels. Plant, Cell & Environment. 2014;37: 2201–
778 2210. doi:10.1111/pce.12297

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

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

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

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

788 90. Chen L, Sun B, Gao W, Zhang Q, Yuan H, Zhang M. MCD1 Associates with FtsZ Filaments
789 via the Membrane-Tethering Protein ARC6 to Guide Chloroplast Division. The Plant Cell.
790 2018;30: 1807–1823. doi:10.1105/tpc.18.00189

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

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

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

803 94. Asano T, Yoshioka Y, Kurei S, Sakamoto W, Machida Y. A mutation of the CRUMPLED
804 LEAF gene that encodes a protein localized in the outer envelope membrane of plastids affects
805 the pattern of cell division, cell differentiation, and plastid division in Arabidopsis. Plant J.
806 2004;38: 448–459. doi:10.1111/j.1365-313X.2004.02057.x

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

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

813 97. Day PM, Potter D, Inoue K. Evolution and targeting of Omp85 homologs in the chloroplast
814 outer envelope membrane. Front Plant Sci. 2014;5: 535. doi:10.3389/fpls.2014.00535

815 98. Agne B, Andrès C, Montandon C, Christ B, Ertan A, Jung F, et al. The Acidic A-Domain of
816 Arabidopsis Toc159 Occurs as a Hyperphosphorylated Protein. Plant Physiology. 2010;153:
817 1016–1030. doi:10.1104/pp.110.158048

818 99. Afitlhile M, Fry M, Workman S. The TOC159 mutant of Arabidopsis thaliana accumulates
819 altered levels of saturated and polyunsaturated fatty acids. Plant Physiology and Biochemistry.
820 2015;87: 61–72. doi:10.1016/j.plaphy.2014.12.018

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

824 101. Afitlhile M, Worthington R, Baldric J. The toc132toc120 heterozygote mutant of Arabidopsis
825 thaliana accumulates decreased levels of the major chloroplast lipids. Phytochemistry.
826 2021;184: 112652. doi:10.1016/j.phytochem.2020.112652

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

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

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

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

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

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

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

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

852 110. Yang W, Pollard M, Li-Beisson Y, Beisson F, Feig M, Ohlrogge J. A distinct type of glycerol-
853 3-phosphate acyltransferase with sn-2 preference and phosphatase activity producing 2-
854 monoacylglycerol. Proceedings of the National Academy of Sciences of the United States of
855 America. 2010;107: 12040–12045. doi:10.1073/pnas.0914149107

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

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

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

864 114. Lechner M, Rossmanith W, Hartmann RK, Thölken C, Gutmann B, Giegé P, et al. Distribution
865 of ribonucleoprotein and protein-only RNase P in Eukarya. Molecular Biology and Evolution.
866 2015;32: 3186–3193. doi:10.1093/molbev/msv187

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

870 116. Gobert A, Gutmann B, Taschner A, Gössringer M, Holzmann J, Hartmann RK, et al. A single
871 Arabidopsis organellar protein has RNase P activity. Nature structural & molecular biology.
872 2010;17: 740–744. doi:10.1038/nsmb.1812

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

876 118. Wang X, Zhao L, Man Y, Li X, Wang L, Xiao J. PDM4, a Pentatricopeptide Repeat Protein,
877 Affects Chloroplast Gene Expression and Chloroplast Development in Arabidopsis thaliana.
878 Frontiers in Plant Science. 2020;11. Available:
879 https://www.frontiersin.org/article/10.3389/fpls.2020.01198

880 119. Zoschke R, Qu Y, Zubo YO, Börner T, Schmitz-Linneweber C. Mutation of the
881 pentatricopeptide repeat-SMR protein SVR7 impairs accumulation and translation of

882 chloroplast ATP synthase subunits in Arabidopsis thaliana. Journal of Plant Research.
883 2013;126: 403–414. doi:10.1007/s10265-012-0527-1

884 120. Wu W, Liu S, Ruwe H, Zhang D, Melonek J, Zhu Y, et al. SOT1, a pentatricopeptide repeat
885 protein with a small MutS-related domain, is required for correct processing of plastid 23S–
886 4.5S rRNA precursors in Arabidopsis thaliana. The Plant Journal. 2016;85: 607–621.
887 doi:10.1111/tpj.13126

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

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

895 123. Huang W, Zhu Y, Wu W, Li X, Zhang D, Yin P, et al. The Pentatricopeptide Repeat Protein
896 SOT5/EMB2279 Is Required for Plastid rpl2 and trnK Intron Splicing. Plant Physiology.
897 2018;177: 684–697. doi:10.1104/pp.18.00406

898 124. Cyrek M, Fedak H, Ciesielski A, Guo Y, Sliwa A, Brzezniak L, et al. Seed Dormancy in
899 Arabidopsis Is Controlled by Alternative Polyadenylation of DOG1. Plant Physiology.
900 2016;170: 947–955. doi:10.1104/pp.15.01483

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

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

905 127. Stamenković M, Woelken E, Hanelt D. Ultrastructure of Cosmarium strains
906 (Zygnematophyceae, Streptophyta) collected from various geographic locations shows species-
907 specific differences both at optimal and stress temperatures. Protoplasma. 2014;251: 1491–
908 1509. doi:10.1007/s00709-014-0652-x

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

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

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

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

921 132. Nakanishi H, Suzuki K, Kabeya Y, Miyagishima S. Plant-Specific Protein MCD1 Determines
922 the Site of Chloroplast Division in Concert with Bacteria-Derived MinD. Current Biology.
923 2009;19: 151–156. doi:10.1016/j.cub.2008.12.018

133. 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

134. 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

135. 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

136. 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

137. 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

138. 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

139. 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

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

141. 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

142. 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

143. 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

144. 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

145. 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

146. 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

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

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

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

973 150. Oikawa K, Matsunaga S, Mano S, Kondo M, Yamada K, Hayashi M, et al. Physical interaction
974 between peroxisomes and chloroplasts elucidated by in situ laser analysis. Nature Plants.
975 2015;1: 15035. doi:10.1038/nplants.2015.35

976 151. Liu H, Stone SL. Abscisic Acid Increases Arabidopsis ABI5 Transcription Factor Levels by
977 Promoting KEG E3 Ligase Self-Ubiquitination and Proteasomal Degradation. The Plant Cell.
978 2010;22: 2630–2641. doi:10.1105/tpc.110.076075

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

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

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

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

992 156. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on
993 genomes, pathways, diseases and drugs. Nucleic Acids Research. 2017;45: D353–D361.
994 doi:10.1093/nar/gkw1092

995 157. Terashima M, Specht M, Hippler M. The chloroplast proteome: a survey from the
996 Chlamydomonas reinhardtii perspective with a focus on distinctive features. Current Genetics.
997 2011;57: 151–168. doi:10.1007/s00294-011-0339-1

998 158. Mueller SJ, Lang D, Hoernstein SNW, Lang EGE, Schuessele C, Schmidt A, et al.
999 Quantitative Analysis of the Mitochondrial and Plastid Proteomes of the Moss Physcomitrella
1000 patens Reveals Protein Macrocompartmentation and Microcompartmentation. Plant
1001 Physiology. 2014;164: 2081–2095. doi:10.1104/pp.114.235754

1002 159. Sun Q, Zybailov B, Majeran W, Friso G, Olinares PDB, van Wijk KJ. PPDB, the Plant
1003 Proteomics Database at Cornell. Nucleic Acids Research. 2009;37: D969–D974.
1004 doi:10.1093/nar/gkn654

1005 160. Emms DM, Kelly S. STAG: Species Tree Inference from All Genes. bioRxiv. 2018; 267914–
1006 267914. doi:10.1101/267914

1007 161. Tria FDK, Landan G, Dagan T. Phylogenetic rooting using minimal ancestor deviation. Nature
1008 Ecology & Evolution. 2017;1: 0193. doi:10.1038/s41559-017-0193

1009 162. Rambaut A, Drummond A. FigTree Version 1.4.4. 2018.

1010 163. Revell LJ. phytools: An R package for phylogenetic comparative biology (and other things).
1011 Methods in Ecology and Evolution. 2012. doi:10.1111/j.2041-210X.2011.00169.x

1012 164. Mooers AØ, Schluter D. Reconstructing Ancestor States with Maximum Likelihood: Support
1013 for One-and Two-Rate Models. Systematic Biology. 1999;48: 623–633.
1014 doi:10.1080/106351599260193

1015 165. Schluter D, Price T, Mooers AØ, Ludwig D. Likelihood of ancestor states in adaptive
1016 radiation. Evolution. 1997;51: 1699–1711. doi:10.1111/j.1558-5646.1997.tb05095.x

1017 166. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam:
1018 The protein families database in 2021. Nucleic Acids Res. 2021;49: D412–D419.
1019 doi:10.1093/nar/gkaa913

1020

1021