Major changes in plastid protein import and the origin of the Chloroplastida

[1]Michael Knopp*, [1]Sriram G. Garg*, [1]Maria Handrich, [1]Sven B. Gould[#]

[1]Institute for Molecular Evolution, HH-University Düsseldorf, 40225 Düsseldorf, Germany

*These authors contributed equally

[#]Corresponding author: gould@hhu.de

Abstract

While core components of plastid protein import (Toc and Tic) and the principle of using N-terminal targeting sequences (NTS) are conserved, lineage-specific differences are known. Rhodophytes and glaucophytes carry a conserved NTS motif, which was lost in the green lineage that also added novel proteins to Toc and Tic. Here we compare the components of plastid protein import and generated RNA-Seq, pigment profile and trans-electron microscopy data based on high-light stress from representatives of the three archaeplastidal groups. In light of plastid protein targeting, we compare the response to high-light stress of archaeplastidal representatives based on RNA-Seq, pigment profile and trans-electron microscopy data. Like land plants, the chlorophyte *Chlamydomonas reinhardtii* displays a broad respond to high-light stress, not observed to the same degree in the glaucophyte *Cyanophora paradoxa* or the rhodophyte *Porphyridium purpureum*. We find that only the green lineage encodes a conserved duplicate of the outer plastid membrane protein channel Oep80, namely Toc75 and suggest that the ability to respond to high-light stress entailed evolutionary changes in protein import, including the departure from phenylalanine-based targeting and the introduction of a green-specific Toc75 next to other import components unique to Chloroplastida. One consequence of relaxed NTS specificity was the origin of dual-targeting of plastid derived proteins to mitochondria and vice versa, using a single ambiguous NTS. Changes in the plastid protein import enabled the green lineage to import proteins at a more efficient rate, including those required for high-light stress response, a prerequisite for the colonization of land.

Keyword: Chloroplast evolution, protein import, Toc75, protein targeting, dual-targeting

High-lights

- Loss of Phe-based N-terminal targeting sequences (NTS) triggered the origin of dual-targeting using a single ambiguous NTS
- The Chloroplastida evolved a green-specific Toc75 for high throughput import, next to a universal and ancient Omp85 present in all Archaeplastida
- A broad response to high-light stress appears unique to Chloroplastida
- Relaxation of functional constraints allowed a broader modification of the green Toc/Tic machinery
- Critical changes in plastid targeting enabled the origin and success of the Chloroplastida and their later conquer of land

Introduction

Mitochondria and plastids are of endosymbiotic origin and compartments surrounded by a double-membrane[1,2]. Most possess their own genomes, but the bulk of their former coding capacity was either lost or integrated into the nuclear genome[3,4]. As a consequence, most of their proteins are post-translationally imported. Guiding of precursor proteins to the matrix or stroma typically relies on N-terminal targeting sequences (NTS)[5–7], although some exceptions are known[8–10]. Archaeplastidal plastids have a monophyletic origin[11–13], which is also evident by the conserved nature plastid import components, a reliable indicator for the monophyly of organelles[14–16].

While sharing a single origin, the plastids of the three algal lineages have evolved considerable differences since their divergence more than a billion years ago[17,18]. These include, but are not limited to: (i) the thickness of a remaining peptidoglycan layer[19,20], (ii) the localisation of starch deposits[21], (iii) the coding capacity of their genomes[3,22], (iv) pigment composition and the types of antenna complexes used[23], (v) the absence or presence of a xanthophyll cycle[24] and (vi) the composition of the protein import machinery[25 26]. It raises the question to what degree the two – critical changes in protein import and changes in plastid biology– are connected, and whether one of the two conditioned or enabled the other. Though most information about plastid protein targeting stems from the green lineage[27], several remarkable differences between the protein import in plastids of the three algal groups (Glaucophyta, Rhodopyhta, and Chloroplastida) are known.

One important difference concerns the NTS that targets proteins to the plastid stroma. Rhodophytes and glaucophytes employ a single amino acid-based motif to target proteins to their plastids[16,27–29]. In most cases this amino acid is a phenylalanine, less frequently other bulky aromatic amino acids[27,30]. The F-based motif is found at the very N-terminus of the NTS (Fig. 1) and even retained in organisms with secondary plastids of red algal origin, such as the cryptophyte *Guillardia theta*, the diatom *Phaeodactylum tricornumutm* and the parasite *Toxoplasma gondii*[31]. It is uncertain why the F-based motif was lost in Chloroplastida, but it came with several changes such as a rise in phosphorylatable serine residues that might help in avoiding erroneous targeting to the mitochondria[32,33].

Despite a tendency towards organelle specificity, eukaryotes also target many proteins simultaneously to two different compartments, a process known as dual-targeting. Dual-

targeting can affect different combinations of compartments[34,35], in plants also mitochondria and plastids. About 100 proteins are dually targeted to the mitochondria and plastids of *Arabidopsis thaliana* after their translation[35,36]. This large number is a consequence of the similarity between the two import mechanisms performed by Tom/Tim (translocator of the outer and inner mitochondrial membrane) and Toc/Tic (translocator of the outer and inner chloroplast membrane)[5,32]. In *A. thaliana*, a duplicate of the Toc64 receptor localizes to the outer mitochondrial membrane and now functions in mitochondrial import[37]. Both *Arabidopsis* organelles also use the same targeting-associated PURPLE ACID PHOSPHATASE2 (*At*PAP2) at their outer membranes[38,39]. The extent of dual-targeting in non-chloroplastidal species remains largely unexplored.

To investigate plastid targeting in a comparative approach across the three main algal lineages, we generated RNA-Seq, pigment profile, and trans-electron microscopy data from three different conditions (with high-light stress as the stimulus) for the chlorophyte *Chlamydomonas reinhardtii*, the rhodophyte *Porphyridium purpureum* and the glaucophyte *Cyanophora paradoxa*. The data were compared and evaluated in light of evolutionary changes regarding protein import. Our analysis connects the loss of F-based targeting and the emergence of new critical import proteins in the ancestor of the green lineage, with a series of critical changes.

Material and Methods

Culturing

Algae were grown in their respective media (see SAG Göttingen or ncma.bigelow.org for recipes) in aerated flasks at 20°C and illuminated with 50µE under a 12/12h day-night cycle. RNA was isolated from cells growing in the exponential phase either at 6h into the day, 6h into the night or after 1h of high-light treatment at 600µE. RNA was isolated, sequenced and assembled exactly as described previously[40] and based on pooled biological triplicates and independently sequenced technical triplicates.

Rapid light curves and pigment profiles

The relative electron transport rates (rETR) of the different algae were measured with use of the FluorCam FC 800MF (Photo Systems Instruments) with modulated red light (emission at 625nm and bandwidth of 40nm) as a source of measuring light (<0.1µmol quanta m^{-2} s^{-1}) and modulated blue light as saturation pulse (> 8000µmol quanta m^{-2} s^{-1}) (Suppl. Fig. 1). The algae samples were dark adapted for 5 min and repeatedly submitted to increasing light intensities (13, 48, 122, 160, 200, 235, 305, 375, 542, 670 µmol quanta m^{-2} s^{-1}) every 11 min. The exported numeric values were fitted according to Eilers & Peeters, 1988. For each pigment extraction the pellet of 50 ml culture was resuspended with 100% acetone, homogenized and kept at -20°C over night. On the next day extracts were centrifuged and supernatant was filtered once through a 200 nm polytetrafluoroethylene membrane and then analyzed by reversed-phage high pressure liquid chromatography (HPLC) with ultraviolet/visible spectroscopy detection (Hitachi/Merck). Pigment concentrations were determined using external pigment standards isolated from spinach thylakoids[56].

Microscopy

For trans-electron microscopy cells were centrifuged at 800 x g and pellet was washed twice with PBS. Afterwards pellets were carefully resuspended with 2,5 % glutaraldehyde in 0,1 M cacodylate buffer and incubated for 2-3 days at 4°C. Fixed cells were then centrifuged and pellets was washed four times with 0,1 M cacodylate buffer with a minimum of 10 min incubation time and centrifugation for 2 min at 13.000 rpm. For contrasting, samples were resuspended in 2% Osmium(VIII)-oxid + 0,8% potassium hexacyanoferrate and incubated for 1 h at room temperature. Then cells were washed again five times and after addition of 3,5 % agarose and resuspension cells were incubated on ice for a minimum of 10 min until agarose became hardened. Tube tips were cut using a guillotine and the solid agar embedded pellet was pulled out and transferred to a small glass container (40 x 19 mm, 5 ml, with plastic lids). Dehydration of cell pellets was achieved using an ascending ethanol washing series starting with 60% ethanol (1 x 10 min), followed by 70% (overnight at 4°C), 80% (2 x 10 min), 90% (2 x 10 min) and 100% (1 x 10 min), finishing with 100% ethanol + molecular strainer (1 x 10 min) and propylenoxid (1 x 15 min). Afterwards epoxide resin/propylenoxid mixtures were added to the samples with increasing epoxide resin concentrations. First, a 1h incubation with epoxide resin/propylenoxid (1:2) was followed by a 1h incubation with epoxid resin/propylenoxid (1:1) and finally an overnight incubation with epoxide resin/propylenoxid (2:1) was performed. Freshly prepared epoxide resin was added the next day and samples incubated for four hours in a vacuum to remove any remaining oxygen within the epoxide resin/cell pellet solution. Finally, pellets were cut in approx. 1 mm slices with a razor blade and placed onto the tip of a notch on a rubber mat and completely covered with epoxide resin. After that epoxide resin filled mats were incubated for 24 h at 40°C followed by 24 h incubation at 60°C for complete polymerization. Probes were then cut using a ultramicroton, placed on monitoring grids and examined using trans-electron microscopy (Zeiss EM902). For the analysis of thylakoid stacks, the distances within 10 cells were counted using Fiji[57]. For each graph 10 cells were analyzed and within each cell 10 different areas counted.

Identification of differentially expressed genes and annotation

Subsequent to the assembly via Trinity[41] (r2013-02-25), edgeR[42] was used to calculate the number of differentially expressed genes. The criteria for the identification were a logarithmic fold change of at least 2 and significance of 0.001 or lower. Since this approach only detected 91 differentially expressed genes for *P. purpureum*, the significance cutoff was lowered to 0.05 as suggested by the EdgeR manual (https://github.com/trinityrnaseq/trinityrnaseq/wiki). The transcripts were ranked according to mean expression values for all three light conditions and each organism. Protein annotation was performed by a BLAST search of all CDS against 112 Refseq plant and algae genomes. All BLAST hits with at least 25% local identity and a maximum E-value of $1x10^{-10}$ were used for annotation. In cases where the hit did not provide enough information (hypothetical proteins, predicted proteins) the next best non-hypothetical hit was selected.

Phylogenomic analysis

The sequence dataset for the phylogenetic analysis of the Toc75/Oep80 homologs consists of 77 amino acid sequences from Chlorophytes, Rhodophytes, Cyanobacteria, Plants and one Glaucophyte. We consulted Inoue and Potter 2004 to obtain 39 amino acid sequences of Toc75 and Oep80 homologs from either the Refseq[43] or GenBank[44] database via their respective gene identifiers (Suppl. table 1)[45]. Additionally, 28 genomes from Chlorophytes, Rhodophytes[46] and one Glaucophyte were downloaded either from the Refseq, GenBank or the JGI Genome Portal[47] (Suppl. table 1). The initial set of sequences was used as query sequences to search for Toc75 and Oep80 homologs via BLASTp (version 2.5.0)[48,49]. All non-redundant hits from each subject genome with at least 25% local identity and a maximum E-value of 0.001 were added to the sequence set. Blast hits of Oep80 and Toc75 sequences of the initial sequence set were named pOep80 and pToc75 respectively.

Multiple protein sequence alignments were constructed using MAFFT (version 7.299b) with the parameters "--maxiterate 1000" and "--localpair"[50]. The initial multiple protein sequence alignment was used to check the quality of identified homologs, resulting in the removal of sequences that differed drastically in overall amino acid composition. The multiple amino acid sequence alignment was then used to construct a phylogenetic tree via RAxML[51] (version 8.2.8) using the substitution model 'PROTCATWAGF' (WAG substitution Matrix and empirical base frequencies) and 100 non-parametric bootstraps. An additional tree was constructed using the new RAxML-NG with the model LG+F+R5 and 1000 bootstraps[52] (Suppl. Fig. 2). The trees were rooted on the split between the monophyletic cyanobacterial sequences and the rest of the taxa.

Sequences of plastid-, mitochondria- and dual-targeted proteins of *A. thaliana* were obtained from Garg and Gould 2016[33]. All proteins were blasted (diamond blastp, identity cutoff: 25%, evalue cutoff: $1x10^{-10}$) against a database of 94 cyanobacterial and 460 alphaproteobacterial proteomes (Suppl. table 2). All hits meeting the cutoffs were plotted against all proteomes in a 2D binary heatmap. The members of each group were sorted according to phylogenetic trees from concatenated alignments, while the order of genes was determined by hierarchical clustering (hclust, method: 'average'). The intracellular localization of the proteins was color coded.

Identification of nuclear encoded, mitochondria- and plastid-targeted genes

Plastid-targeted proteins were identified by blasting known and manually curated plastid-targeted proteins from *A. thaliana*[33] against the genome of *C. reinhardtii*, *C. paradoxa* and *P. purpureum* (identity of at least 50%, query coverage of at least 50%, maximum E-value of $1x10^{-5}$) or extracted from published proteome data when available[53,54]. To identify mitochondria-targeted proteins, we blasted all mitochondria-targeted proteins from human, mouse and rat (according to the IMPI database, marked as "Known mitochondrial") against the genomes of the three algae (identity of at least 50%, query coverage of at least 50%, maximum E-value of $1x10^{-5}$). Sequence logos of the mitochondria- and plastid-targeted proteins were curated manually by aligning the first 20 amino acids following an F, Y, W or L (according to[31]) and plotted using Seq2Logo[55].

Results

Adaptive changes of common photosynthetic pigments upon high-light stress

Plants react in particular to changes in light intensity[58,59]. To analyse the differences that high-light stress has on the three algae, representing the three major groups (Table 1), we set out to perform comparative studies. The algae were adapted to growing at 50 mol photons m^{-2} s^{-1} under a 12/12 day-night cycle and at 20°C. Through rapid light curves we assessed that at 600 mol photons m^{-2} s^{-1}, a saturation of the photosystems was reached in all three species (Suppl. Fig. 2). For the high-light stress treatment, the algae were hence exposed to 600 mol photons m^{-2} s^{-1} for 1h. For comparison we determined the pigment profiles from cultures that were either 6h into the night or 6h into the day phase.

The glaucophyte *C. paradoxa* shows no significant change in pigment concentration or composition, neither at night nor after light stress (Fig. 2a). For the red alga *P. purpureum* we observed only very marginal changes and the concentration of pigments for the samples collected at night was the highest. Pigment concentrations seemed to slowly decrease during the day and even further under high-light stress. This was observed for all three major pigment groups at a similar rate (Fig. 2a). Only in the green alga *C. reinhardtii* the pigment composition changed significantly especially upon high-light stress (Fig. 2a). Here in particular the xanthophyll cycle, i.e. the enzyme-driven and reversible conversion of violaxanthin into zeaxanthin is evident, a component of non-photochemical quenching thought to be absent in glauco- and rhodophytes[24]. Concentrations of chlorophylls and carotenoids actually increase under high-light stress in *C. reinhardtii*, demonstrating their de novo synthesis.

The thylakoid stacks (grana) of land plants relax under high-light stress in order for the repair mechanism of the photosystems to properly function[60]. This concerns in particular the degradation of the D1 protein through the membrane-bound protease FtsH, whose dimerized size is too large for the space where two thylakoid stacks align[59]. Algae form different types of thylakoid stacks[60,61], but no grana-like structures. We performed trans-electron microscopy (TEM)-based analysis of the cells from the three different conditions and determined the distance between neighbouring thylakoid stacks. The differences we observed were in all cases marginal, but only in the case of *C. reinhardtii* did we observe a statistically significant increase in spacing upon high-light stress (Fig. 2b).

The transcriptional response to high-light stress is most pronounced in the chlorophyte

We also generated RNA-Seq data on all samples. They reveal stark differences among the three species in terms of overall transcriptional regulation (Fig. 3). In the chlorophyte, the response to high-light stress was the most pronounced among the three algae, both regarding the number of differentially expressed genes as well as in the number of upregulated genes during high-light conditions. For each condition a clear separation was observed and a specific gene set upregulated in comparison to the average (Fig. 3).

Under high-light conditions the chlorophyte upregulates the expression of photosynthesis machinery components as well as proteins that promote photoprotection. A total of 418 transcripts were found to be differentially expressed, 274 values of which were significant (Suppl. table 3). The upregulated photoprotective proteins include stress-related chlorophyll

binding proteins 1 and 3 involved in energy-dependent quenching to dissipate excess energy[62], members of the early-light inducible protein family (Elip), ancestral homologs of the non-photochemical quenching associated PSBS/LHCSR3 family[65,66], a CPD photolyase class II that reverses the formation of pyrimidine dimers that result from the exposure to strong UV radiation[67], and chlorophyll b reductases and beta-carotene hydroxylases that prevent over-excitation of the photosystem and protects the cells from high-light intensities[68,69]. Next to these photoprotective proteins, photosynthesis house-keeping genes such as PSII Pbs27, Rieske protein, PSII subunit 28, and several proteins of the LHC superfamily were upregulated as well as a few stress-response proteins such as the plastidal homolog of DnaJ and other members of the HSP70 protein family that can form a multichaperone complex together[70] (Suppl. table 3).

In the glaucophyte *C. paradoxa*, most of the 1,463 differentially expressed transcripts were found upregulated during darkness in correspondence to nightly proliferation. The overall difference between day and night was far more pronounced than day versus high-light and the difference between light and high-light conditions smaller than in *Chlamydomonas*. Only 26 transcripts were found to be upregulated during high-light conditions (Suppl. table 3). In comparison to the green alga, fewer proteins involved in photosynthesis regulation and photoprotection were found to be upregulated during high-light stress but also included several Elip proteins.

The identification of differentially expressed genes in *P. purpureum* posed to be more difficult. Only 90 transcripts were initially identified, but by lowering the significance cut-off — according to the standard edgeR protocol — we were able to detect another 980. The expression profile matches that of *C. paradoxa*, although the number of regulated genes is far smaller. For the transition from daylight to high-light, a total of 38 transcripts were identified (Suppl. table 3). The most notable proteins that were upregulated during high-light stress was a high-light inducible protein (Hlip) involved in non-photochemical quenching[71], and several heat shock proteins (HSP70). The response to light stress was far weaker than in the other two algae, but *P. purpureum* might regulate its RNA levels through the extensive use of miRNAs[72] which could contribute to the lower levels of differentially expressed genes identified.

Comparisons of the most highly upregulated proteins of each of the three algae among all conditions revealed additional differences in light-dependent differential gene expression. While *C. reinhardtii* upregulates the synthesis of several photosynthesis and plastid-related proteins during light and high-light conditions, *C. paradoxa* and *P. purpureum* upregulate only a few. In the case of *C. paradoxa*, the biggest notable difference is the focus on protein biosynthesis during darkness/night. The 50 most highly upregulated proteins during the night consist of approx. 90% ribosomal proteins, indicating an increase in overall protein biosynthesis and proliferation (Suppl. table 4). We observe photosynthesis machinery components as well as photoprotection components to be among the most upregulated proteins in combination only in *C. reinhardtii*, illustrating the chlorophyte's more elaborate ability to adapt to differing light conditions compared to the other two screened algae.

The red Toc75 is an Oep80, and Toc75 unique to Chloroplastida

In *Arabidopsis*, most members of its Toc75 family have been characterized. This includes the main import pore of the outer membrane, Toc75[73] (TOC75-III, At3g46740), as well as Oep80 (TOC75-V, At5g19620) whose exact function remains unresolved while the protein is essential for plant viability[74,75], and most recently SP2 (At3g44160) which serves protein export for chloroplast-associated protein degradation[76]. The situation in rhodophytes and glaucophytes differs and they seem not to encode the same number of Toc75 homolog[27,77].

We collected 77 eukaryotic proteins of the Toc75 and Oep80 family from 44 eukaryotic species and routed them against their cyanobacterial homologs for the construction of a phylogenetic tree. The single glaucophyte sequence sits basal to all others, while the rhodophyte sequences form a well-supported group that is sister to all chloroplastidal sequences (Fig. 4). The sequences of green algae and plants fall into two distinct and again well-supported clusters: one comprises a group of proteins including the *At*Oep80, the other a group containing the main import pore *At*Toc75. Within these two groups separating the Oep80 from the Toc75 proteins, the separation between the chloro- and streptophytes is observed, as well as the basal branching of *Chara braunii* – a streptophyte alga related to the ancestor of land plants[78,79] (Fig. 4).

Discussion

If one measures evolutionary success by species diversity, the green lineage is the most successful. About 16,000 green algal, 5,000 rhodophyte and thirteen glaucophyte species have been recognized (with >100,000, 500–1000 and about a dozen that remain to be described, respectively)[80]. Another 400,000 land plant species[81] evolved since the conquering of land some 480 million years ago[82,83]. We argue that the evolutionary origin and success of the green lineage hinges upon early changes in plastid protein targeting.

Algae and plant cells target more than a thousand proteins specifically to each of their two compartments of endosymbiotic origin. Plastid targeting evolved in a cell that had already established mitochondrial targeting, yet both import machineries share similarities and both rely on specific NTSs for matrix and stroma targeting[5]. The origin of the mitochondrial NTS is uncertain, but its positive charge was an early requirement to overcome the bioenergetic inner mitochondrial membrane[10]. The most N-terminal domain carries the charged residues critical for distinguishing between mitochondrial- and plastid targeting (Fig. 1), while the C-terminus is exchangeable[84]. Because the plastid is younger and because the photosynthetic organelle evolved in a eukaryotic cell instead of contributing to its actual origin, we understand more about the origin of the plastid NTS.

On the origin of the N-terminal targeting sequence

It has been speculated that N-terminal targeting sequences evolved from antimicrobial peptides (AMPs)[85], as both share similarities in terms of charged amino acid residues, the ability to form amphiphilic α-helices, and because they are frequently identified in host-endosymbiont relationships[86]. One example regarding the latter is *Paulinella chromatophora*, whose chromatophore origin is independent from that of the Archaeplastida and younger[87]. Two types of NTSs were identified that target nuclear-encoded proteins to the chromatophore, but both are not related to the simultaneously identified AMPs[88], which argues against an AMP-origin

of the NTS in *Paulinella*. The concept is also not compatible with the origin of phenylalanine-based plastid targeting and Toc75.

The components of the Toc and Tic machinery share a mixed pro- and eukaryotic ancestry[89,90]. Toc75, the β-barrel import pore in the outer membrane, is of prokaryotic origin and a member of the Omp85 superfamily[25]. Some bacterial Omp85's recognize their substrates through a C-terminal phenylalanine[91] and evidence is emerging that the POTRA domains of Toc75 act as binding sites for the NTS[92]. If we recall that the phenylalanine-based motif is retained in rhodophytes and glaucophytes[29], we can conclude that the pNTS did not evolve from AMPs but rather adapted in evolution and traces back to a recognition signal for the cyanobacterial Omp85 that evolved into Toc75[93]. The ancestral character of phenylalanine-based plastid targeting was lost with the origin of the Chloroplastida and we suggest simultaneously to the expansion of the Toc75 family – with significant consequences for the green lineage.

Dual-targeting using a single ambiguous signal is the consequence of losing the F-based motif

The use of a F-based motif offered an elegant solution to the archaeplastidal ancestor. It utilized an existing translocons-substrate recognition mechanism and allowed to distinguish cytosolically translated mitochondrial from plastid proteins through a single amino acid-based motif. With the emergence of the green Toc75 and loss of the F-based motif, false targeting likely increased. One counter-measure was the increase in phosphorylation sites in the NTS, which adds negative charge and hampers import of the substrate by mitochondria[10,34,40]. Many proteins, however, remain dually targeted in *Arabidopsis*[36] and we predict this is restricted to the green lineage. Dual-targeting to mitochondrion and plastid does occur in algae with a red plastid, but through alternative transcription/translation initiation and not through the use of a single ambiguous NTSs[94].

Evolution is blind. Dual-targeting evolves from falsely targeted proteins that initially might not offer a direct benefit, but are also not detrimental to the cell's viability. This can re-localise or establish entire new pathways[95] and there is no apparent preference regarding the direction of flow: as much proteins of cyanobacterial origin are targeted to the mitochondria as they are *vice versa* (Fig. 5). Dual-targeted proteins are largely part of the transcription and translation machinery[36]. This might include the plastid-associated polymerases, whose dual-targeting in Chloroplastida might be an ancestral trade of the lineage[79]. Both the mitochondrion and plastid have a genome, and as such information processing proteins suit a dual-targeting route well. A simultaneous control over the transcription and translation of both organelles might allow for a faster and accurate response or simply easier house-keeping. Dual-targeting reinforces the cross-communication between the two organelles of endosymbiotic origin, likely offering an evolutionary advantage to cells carrying dozens of mitochondria and plastids simultaneously such as the cells of land plants but not all algae.

An Oep80 derived Toc75 is unique to the green lineage

One of the earliest descriptions of Toc75 was for a protein isolated from pea[96]. Conserved homologs across all Chloroplastida were quickly identified[15,97], butt required way more effort

368 across the diversity of the Archaeplastida. Through the identification of a Omp85 homolog in
369 algae with secondary red plastids, it became evident that all phototrophic eukaryotes harbour
370 beta-barrel forming proteins of an extended Omp85 family that form the import pore in the
371 outer plastid membrane[98], but with a decisive difference regarding the number of encoded
372 homologs.

373 Our phylogenetic analysis of Toc75 and Oep80 supports previous analyses without the need of
374 any sequence trimming. It demonstrates the clear-cut, likely also functional, separation between
375 the Toc75 and Oep80 proteins of Chloroplastida[25]. The red sequences are closer to their
376 prokaryotic homologs and the green Toc75 is further derived. From the perspective of
377 phylogeny, there is little doubt that Toc75 is unique to the green lineage and originated from
378 the duplication of an ancestral Omp80 that remains conserved in the other two lineages. This
379 suggests a division of labour at the outer chloroplast membrane not found in rhodoplasts or
380 cyanelles, the benefits of which are plenty. Glauco- and rhodophytes work with a single import
381 pore, whereas *Arabidopsis* and its green relatives encode a single full-length Toc75 and a single
382 full-length Oep80. Both of the latter are expressed at high levels in a conserved ratio and in the
383 different tissues according to the gene expression atlas of the TAIR database[99]. Their presence
384 is needed simultaneously and appears synchronized.

385 We speculate that the duplication of Oep80 allows for a more efficient, faster and versatile
386 protein import. It might be a prerequisite for the elaborate response to high-light stress, which
387 our data supports (Fig. 2, Fig. 3). A response to high-light stress is evident in all three lineages
388 (Fig. 3), but differs in quantity and detail. *Chlamydomonas* not only alters its gene expression
389 network the most upon high-light stress, but also focuses more on photosynthesis maintenance
390 and protection, reacts less stressed and rapidly synthetizes pigments de-novo (Fig. 2). The
391 upregulation of Elips that are of cyanobacterial origin occurs in all three lineages, but they were
392 only expanded and diversified in the green lineage[100]. Retrograde signalling (a critical part of
393 the response to high-light stress) is limited by the plastid's import capacity[101], highlighting the
394 direct dependence.

395 If Oep80's main duty is indeed the integration of beta-barrel proteins (and maybe other delicate
396 substrates of unknown nature), then it releases Toc75 from this job. This mirrors the situation
397 in mitochondria, where Tom40 acts as the main import pore while Sam50 incorporates beta-
398 barrel proteins with a complicated topology into the outer plastid membrane[102]. The division of
399 labour appears more effective than the simple increase in number of a single import gateway.
400 This then maybe also allowed the endosymbiotic gene transfer of the small subunit of RubisCo
401 to the nucleus, a trademark of the green lineage[103,104]. The sheer amount of RbcS protein
402 required to be imported might simply overstrain the Oep80 of rhodo- and glaucophytes and its
403 gene transfer from the plastid to the nucleus is hence selected against. These patterns allow to
404 speculate on the sequence of evolutionary events.

405 Initially a duplication of the ancestral import pore Oep80 occurred and both paralogs might
406 have performed the same duty early on. Mutations in one of the two copies led to an
407 independence of F-based targeting, alternative substrate recognition, the emergence of NTS
408 phosphorylation[33], and a cytosolic 14-3-3/Hsp70-based guidance complex[105] that we predict is
409 unique to the green lineage, too. The plastid-encoded Tic214(YCF1)/YCF2/FtsHi complex
410 emerged early in chlorophyte evolution, too, maybe through the duplication of an early Tic20-

like protein[26,106]. The components of this complex are highly diverse, except for a C-terminal motif, and were entirely lost in grasses without impacting protein import[107,108]. Other components were added such as Tic40 that increases import efficiency[109], and which is absent from rhodo- and glaucophytes[15]. Ever more plastid proteins went via the Toc75 route, apart from the slow folding proteins of the outer-membrane that continued to be integrated via Oep80. A more recent extension was the emergence of the CHLORAD pathway (chloroplast-associated protein degradation). Its central component, SP2, is an Omp85 paralog as well but lacks the POTRA domains[76]. It likely emerged in angiosperms and might facilitate the remodelling of plastids (e.g of a chloro- to a chromoplast), a feature unique to higher land plants and their embryoplast[18]. Therefore, the implementation of another plastid protein transport pathway based on an Oep80 duplication coincided with yet another major step in land plant evolution.

Conclusion

Plastid endosymbiosis introduced phototrophs to the eukaryotic tree of life. A critical step was the evolution of a basic Toc/Tic protein import machinery that is conserved across all algae and plants. It is evident that major modifications of the Tic/Toc machinery and changes in the targeting sequences occurred early in the origin of the Chloroplastida. This concerns especially (i) the loss of phenylalanine-based targeting and (ii) the emergence of new import machinery components such as Tic40, a plastid-encoded Tic214, and a Toc75 that evolved from the duplication of the ancestral Omp80. We speculate that the former resulted in the emergence of dual organelle (plastid and mitochondrion) targeting using a single ambiguous targeting sequence and that the latter introduced a “high-throughput” import pathway for nuclear-encoded proteins. The main import pore of the green plastid, Toc75, is released from dealing with slow-folding proteins of the outer membrane and no longer left hamstrung when there is the need for rapid import of proteins required to cope with high-light stress. Whatever the details regarding the substrates imported by Oep80, the Chloroplastida make use of two major import pores, where rhodophytes and glaucophytes need to cope with one. Responses to high-light stress is variegated, but it requires the efficient and immediate import of over a hundred nuclear-encoded plastid proteins simultaneously after retrograde plastid signaling. This was realized by the implementation of an efficient plastid import pathway that enabled the evolutionary success of the Chloroplastida, a pinnacle of which was the conquer of land.

Acknowledgments

We thank Matheus Sanita Lima for discussing dual targeting and Prof. Peter Jahns for providing access to the HPLC and help in analyzing the pigment profiles. This work was supported through the DFG (267205415 – SFB 1208) and the VolkswagenStiftung (Life).

Figures

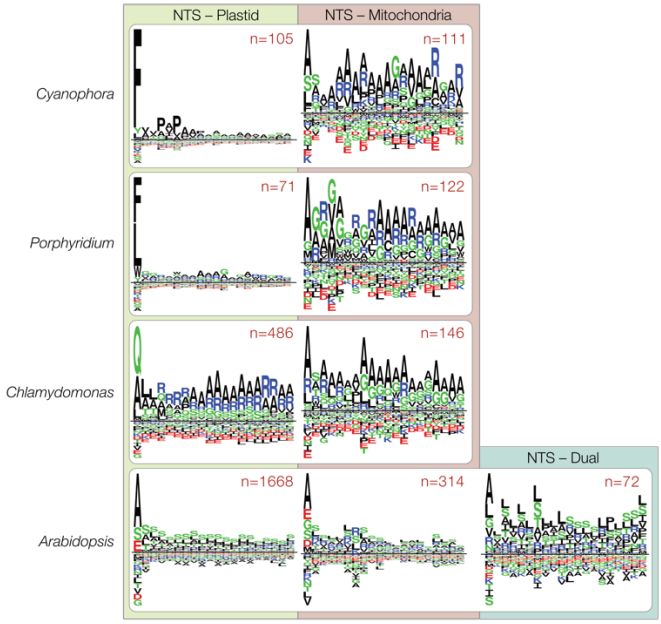


Fig. 1: Targeting motifs and phylogenetic origin of organelle targeted proteins. (a) NTS of plastid- or mitochondria-targeted proteins of *C. paradoxa*, *C. merolae* and *A. thaliana*. The three species showcase the NTS for plastid- or mitochondira-targeting in the *Glaucophyta, Rhodophyta* and *Chlorophyta*, respectively. While an F-based plastid targeting motif is evident in *Glaucophyta* and *Rhodophyta*, it was lost in the green lineage.

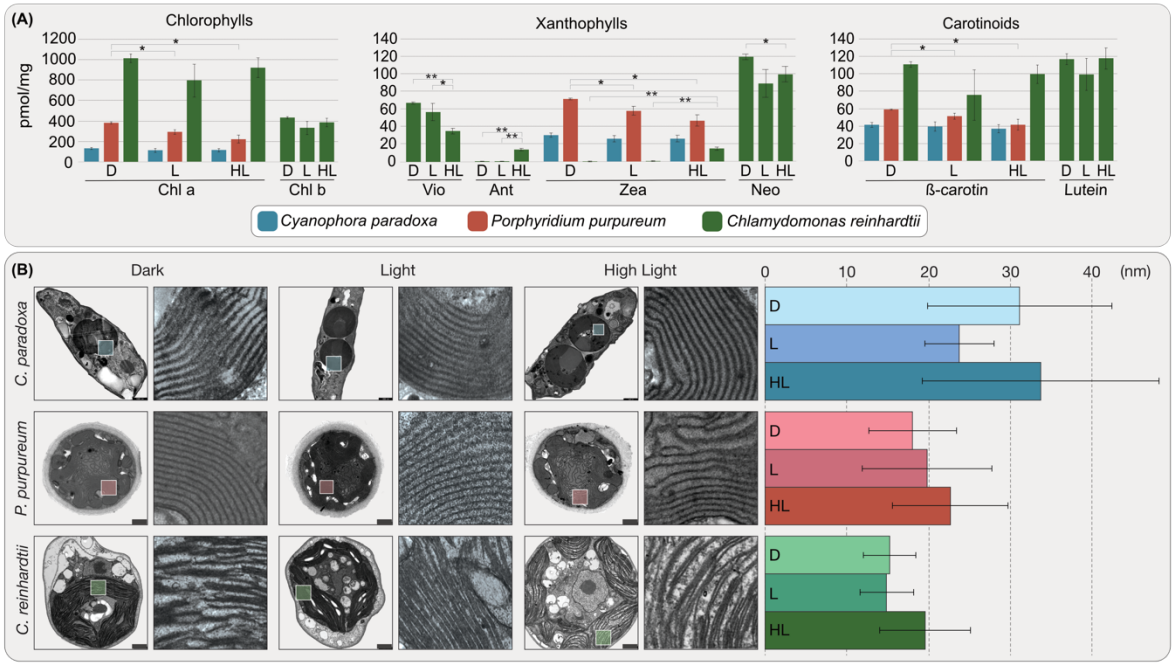
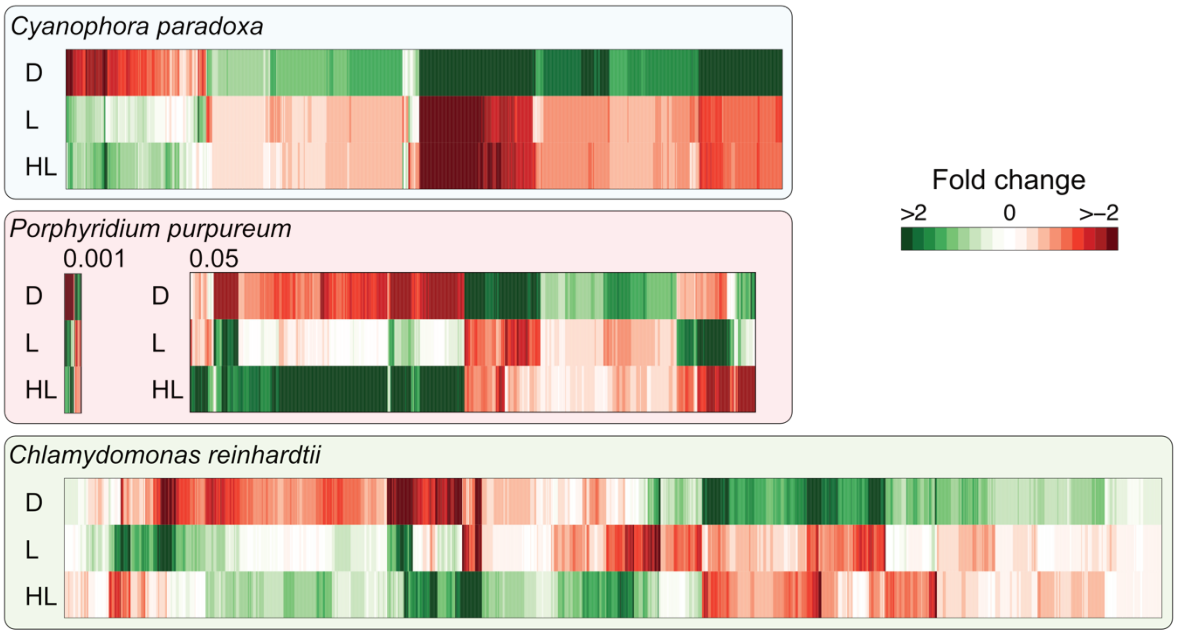


Fig. 2: Pigment profiles and analysis of thylakoid stack distance during high-light stress. (a) Pigments were extracted by homogenization with acetone and their concentrations determined by an HPLC analysis. In both the glaucophyte and rhodophyte the pigment concentrations remain rather stable and only a slight decrease in the overall pigment concentration is observed during the day and even more so during high-light stress. On the

462 contrary, in C. reinhardtii all three types of pigment change their concentration significantly
463 and e.g. the step-wise reduction of violaxanthin (Vio) to antheraxanthin (Ant) and zeaxanthin
464 (zea) is evident. **(b)** Cells from the three different conditions were fixed and analyzed using
465 trans-electron microscopy and distances between the thylakoids were measured using Fiji. An
466 obvious and statistically significant increase in thylakoid distance upon high-light stress is only
467 observed in C. reinhardtii, although a similar but less significant trend is observed in the red
468 alga *P. purpureum*. *<0,05 **<0,001 ***<0,0001



469

470 **Fig. 3: Differentially expressed genes of *C. reinhardtii*, *C. paradoxa* and *P.purpureum*.**
471 Visualization of all differentially expressed genes of *C. reinhardtii*, *C. paradoxa* and *P.*
472 *purpureum*, colored according to the logarithmic fold change of the expression among all tested
473 light conditions. Logarithmic fold changes of gene expression are color coded. For *C.*
474 *reinhardtii* and *C. paradoxa*, the fold change's significance of all visualized transcripts is at
475 least 0.001. For *P. purpureum*, the significance cutoff was lowered to 0.05, since the original
476 cutoff revealed only 90 differentially expressed genes. *C. reinhardtii* shows distinct sets of
477 genes each tailored towards one of the tested light conditions. *C. paradoxa* and *P. purpureum*
478 on the other hand, do not show such an adaptation to altering light conditions, especially not to
479 high-light. *C. paradoxa* does not change much of its gene expression between daylight and
480 high-light conditions, showing its lack of adaptation. Although *P. purpureum* expresses a set
481 of genes only during high-light conditions, their differential expression was only detectable by
482 lowering the significance cutoff. Even if all differentially expressed genes of *P. purpureum* are
483 considered, its transcriptional changes during high-light remain minor.

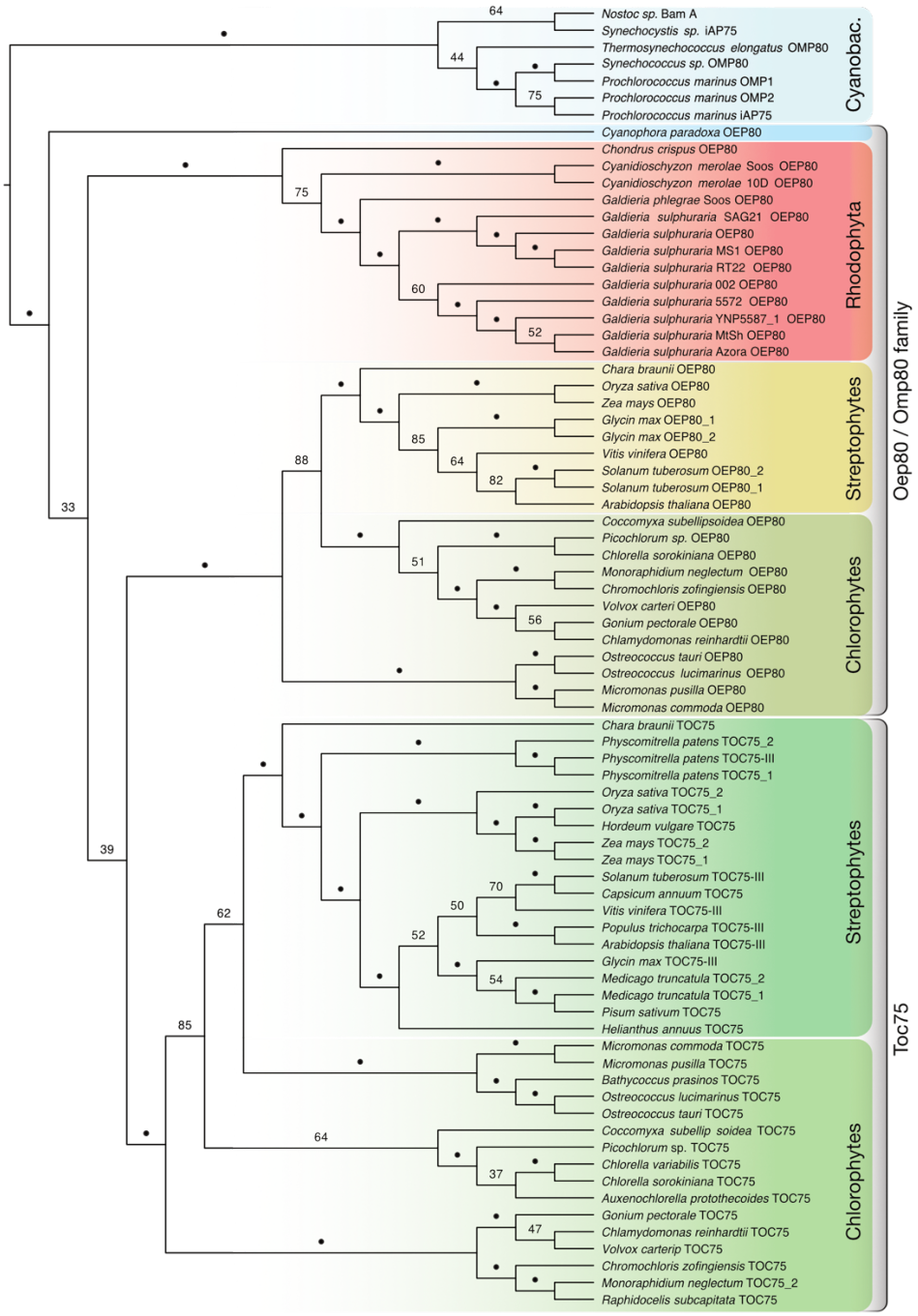


Fig. 4: Phylogenetic analysis of Oep80 and Toc75 homologs. A total of 77 amino acid sequences of Oep80/Toc75 homologs from members of the chlorophytes, rhodophytes and glaucophytes were used for phylogeny reconstruction via RAxML (PROTCATWAGF) with 100 bootstraps. The tree was rooted on the split between the monophyletic cyanobacteria and the eukaryotic sequences. The cyanobacteria as well as all three algal groups form monophyletic groups. Within the green lineage, the Toc75 and Oep80 sequences form separate clusters, indicating the emergence of Toc75 within the green lineage.

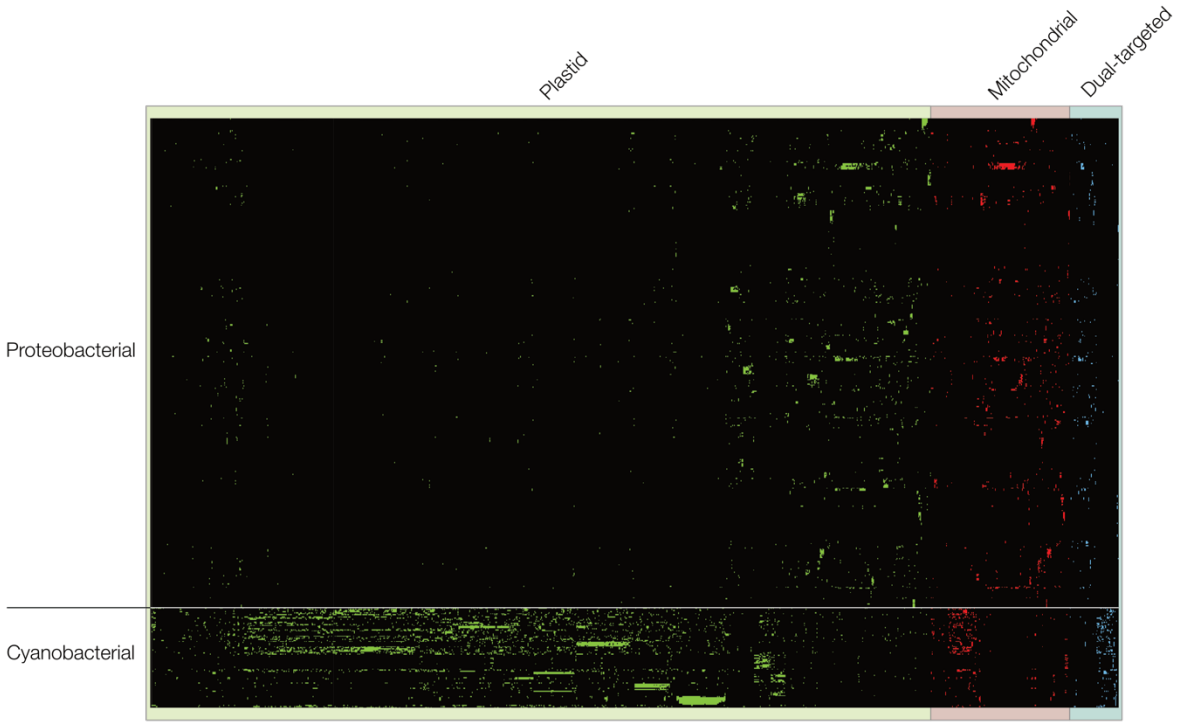


Fig. 5: Phylogenetic oigin of plastid- and mitochondria-targeted proteins of *Arabidopsis*. Binary presence and absence pattern of homologs of plastid- (green), mitochondria- (red) and dual-targeted (blue) proteins of *A. thaliana* within 94 cyanobacterial and 460 alphaproteobacterial proteomes. Organisms are sorted according to previously constructed group-specific phylogenies, while genes are sorted by hierarchical clustering. Most homologs of plastid-targeted genes were identified in cyanobacteria, but for more than one fifth (22%) of the plastid-targeted genes the majority of homologs were identified in alphaproteobacteria. In the case of mitochondria-targeted genes, for almost one third (35%) of the genes most homologs were identified in cyanobacteria instead of alphaproteobacteria. The phylogenetic signal of the dual-targeted genes is more evenly distributed among cyanobacteria and alphaproteobacterial with one half (45%) showing a cyanobacterial origin and the other half (55%) showing an alphaproteobacterial origin.

Table 1: Major differences among the three primary algae lineages and land plants, concerning their coding capacity, composition of the photosynthetic apparatus and carbon storage properties.

Organism	Protein-coding genes			Antenna proteins	Chlorophylls	Antenna pigments	Thylakoid organization	Starch & Storage
	Nucleus	Plastid	Mitochondrion					
Arabidopsis thaliana *(Streptophyte plant)*	35,176	88	122	LHC protein complex	a,b	beta-Carotin, Lutein, Neoxanthin, Violaxanthin, Antheraxanthin, Zeaxanthin	Stacked, Grana	Starch
Chara braunii *(Streptophyte algae)*	23,546	105	46	LHC protein complex	a,b	beta-Carotin, Lutein, Neoxanthin, Violaxanthin, Antheraxanthin, Zeaxanthin	Stacked	Starch
Chlamydomonas reinhardtii *(Chlorophyte algae)*	14,411	69[1]	8[1]	LHC protein complex	a,b	beta-Carotin, Lutein, Neoxanthin, Violaxanthin, Antheraxanthin, Zeaxanthin	Stacked	Starch
Porphyridium purpureum *(Rhodophyte)*	8,355	199[3]	n.d.	Phycobilisomes	a	beta-Carotin, Zeaxanthin	Unstacked, equidistant and single	Glycogen, Floridean starch
Cyanophora paradoxa *(Glaucophyte)*	27,921 (25,831)	136[4]	44[5]	Phycobilisomes	a	beta-Carotin, Zeaxanthin	Unstacked, equidistant and single	Floridean starch

References

1. Zimorski, V., Ku, C., Martin, W. F. & Gould, S. B. Endosymbiotic theory for organelle origins. *Curr. Opin. Microbiol.* **22**, 38–48 (2014).
2. Archibald, J. M. Endosymbiosis and Eukaryotic Cell Evolution. *Curr. Biol.* **25**, R911–R921 (2015).
3. Timmis, J. N., Ayliff, M. A., Huang, C. Y. & Martin, W. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics* **5**, 123–135 (2004).
4. Martin, W. & Herrmann, R. G. Gene Transfer from Organelles to the Nucleus: How Much, What Happens, and Why? *Plant Physiol.* **118**, 9–17 (1998).
5. Schleiff, E. & Becker, T. Common ground for protein translocation: Access control for mitochondria and chloroplasts. *Nature Reviews Molecular Cell Biology* **12**, 48–59 (2011).
6. Dudek, J., Rehling, P. & van der Laan, M. Mitochondrial protein import: Common principles and physiological networks. *Biochimica et Biophysica Acta - Molecular Cell Research* **1833**, 274–285 (2013).
7. Paila, Y. D., Richardson, L. G. L. & Schnell, D. J. New insights into the mechanism of chloroplast protein import and its integration with protein quality control, organelle biogenesis and development. *Journal of Molecular Biology* **427**, 1038–1060 (2015).

8. Goldberg, A. V. *et al.* Localization and functionality of microsporidian iron-sulphur cluster assembly proteins. *Nature* **452**, 624–628 (2008).

9. Hamilton, V., Singha, U. K., Smith, J. T., Weems, E. & Chaudhuri, M. Trypanosome alternative oxidase possesses both an N-terminal and internal mitochondrial targeting signal. *Eukaryot. Cell* **13**, 539–47 (2014).

10. Garg, S. *et al.* Conservation of transit peptide-Independent protein import into the mitochondrial and hydrogenosomal matrix. *Genome Biol. Evol.* **7**, 2716–2726 (2015).

11. Rodríguez-Ezpeleta, N. *et al.* Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. *Curr. Biol.* **15**, 1325–1330 (2005).

12. Jackson, C. J. & Reyes-Prieto, A. The mitochondrial genomes of the glaucophytes gloeochaete wittrockiana and cyanoptyche gloeocystis: Multilocus phylogenetics suggests amonophyletic archaeplastida. *Genome Biol. Evol.* **6**, 2774–2785 (2014).

13. Sánchez-Baracaldo, P., Raven, J. A., Pisani, D. & Knoll, A. H. Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E7737–E7745 (2017).

14. Cavalier-Smith, T. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J. Eukaryot. Microbiol.* **46**, 347–66 (1999).

15. Kalanon, M. & McFadden, G. I. 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 (2008).

16. Gould, S. B., Maier, U. G. & Martin, W. F. Protein Import and the Origin of Red Complex Plastids. *Current Biology* **25**, R515–R521 (2015).

17. Gibson, T. M. *et al.* Precise age of Bangiomorpha pubescens dates the origin of eukaryotic photosynthesis. *Geology* **46**, (2017).

18. de Vries, J., Stanton, A., Archibald, J. M. & Gould, S. B. Streptophyte Terrestrialization in Light of Plastid Evolution. *Trends in Plant Science* **21**, 467–476 (2016).

19. Pfanzagl, B. *et al.* Primary structure of cyanelle peptidoglycan of Cyanophora paradoxa: A prokaryotic cell wall as part of an organelle envelope. *J. Bacteriol.* **178**, 332–339 (1996).

20. Hirano, T. *et al.* Moss chloroplasts are surrounded by a peptidoglycan wall containing D-amino acids. *Plant Cell* **28**, 1521–1532 (2016).

21. Suzuki, E. & Suzuki, R. Variation of Storage Polysaccharides in Phototrophic Microorganisms. *J. Appl. Glycosci.* **60**, 21–27 (2013).

22. Allen, D. K., Laclair, R. W., Ohlrogge, J. B. & Shachar-Hill, Y. Isotope labelling of Rubisco subunits provides in vivo information on subcellular biosynthesis and exchange of amino acids between compartments. *Plant, Cell Environ.* **35**, 1232–1244 (2012).

23. Tomitani, A. *et al.* Chlorophyll b and phycobilins in the common ancestor of cyanobacteria and chloroplasts. *Nature* **400**, 159–162 (1999).

24. Goss, R. & Jakob, T. Regulation and function of xanthophyll cycle-dependent photoprotection in algae. *Photosynthesis Research* **106**, 103–122 (2010).

25. Day, P. M., Potter, D. & Inoue, K. Evolution and targeting of omp85 homologs in the

chloroplast outer envelope membrane. *Front. Plant Sci.* **5**, (2014).

26. Kikuchi, S. *et al.* Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science.* **339**, 571–574 (2013).

27. Köhler, D. *et al.* Identification of protein N-termini in Cyanophora paradoxa cyanelles: Transit peptide composition and sequence determinants for precursor maturation. *Front. Plant Sci.* **6**, 1–11 (2015).

28. Steiner, J. M., Yusa, F., Pompe, J. A. & Löffelhardt, W. Homologous protein import machineries in chloroplasts and cyanelles. *Plant J.* (2005). doi:10.1111/j.1365-313X.2005.02559.x

29. Wunder, T., Martin, R., Löffelhardt, W., Schleiff, E. & Steiner, J. M. The invariant phenylalanine of precursor proteins discloses the importance of Omp85 for protein translocation into cyanelles. *BMC Evol. Biol.* **7**, (2007).

30. Gruber, A. *et al.* Protein targeting into complex diatom plastids: Functional characterisation of a specific targeting motif. *Plant Mol. Biol.* **64**, 519–530 (2007).

31. Patron, N. J. & Waller, R. F. Transit peptide diversity and divergence: A global analysis of plastid targeting signals. *BioEssays* **29**, 1048–1058 (2007).

32. Garg, S. G. & Gould, S. B. The Role of Charge in Protein Targeting Evolution. *Trends in Cell Biology* **26**, 894–905 (2016).

33. Lee, J., O'Neill, R. C., Park, M. W., Gravel, M. & Braun, P. E. Mitochondrial localization of CNP2 is regulated by phosphorylation of the N-terminal targeting signal by PKC: Implications of a mitochondrial function for CNP2 in glial and non-glial cells. *Mol. Cell. Neurosci.* **31**, 446–462 (2006).

34. Karniely, S. & Pines, O. Single translation-dual destination: Mechanisms of dual protein targeting in eukaryotes. *EMBO Reports* **6**, 420–425 (2005).

35. Carrie, C. & Small, I. A reevaluation of dual-targeting of proteins to mitochondria and chloroplasts. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833**, 253–259 (2013).

36. Carrie, C. *et al.* Approaches to defining dual-targeted proteins in Arabidopsis. *Plant J.* **57**, 1128–1139 (2009).

37. Chew, O. *et al.* A plant outer mitochondrial membrane protein with high amino acid sequence identity to a chloroplast protein import receptor. *FEBS Lett.* **557**, 109–14 (2004).

38. Sun, F. *et al.* AtPAP2 is a tail-anchored protein in the outer membrane of chloroplasts and mitochondria. *Plant Signal. Behav.* **7**, (2012).

39. Law, Y.-S. *et al.* Phosphorylation and Dephosphorylation of the Presequence of Precursor MULTIPLE ORGANELLAR RNA EDITING FACTOR3 during Import into Mitochondria from Arabidopsis. *Plant Physiol.* **169**, 1344–55 (2015).

40. Gould, S. B. *et al.* Adaptation to life on land at high O_2 via transition from ferredoxin- to NADH-dependent redox balance. *Proc. R. Soc. B Biol. Sci.* **286**, 20191491 (2019).

41. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).

42. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).

43. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).

44. Clark, K., Karsch-mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res.* **44**, 67–72 (2016).

45. Inoue, K. & Potter, D. The chloroplastic protein translocation channel Toc75 and its paralog OEP80 represent two distinct protein families and are targeted to the chloroplastic outer envelope by different mechanisms. *Plant J.* **39**, 354–365 (2004).

46. Rossoni AW, Price DC, Seger M, Lyska D, Lammers P, Bhattacharya D, Weber AP.. The genomes of polyextremophilic cyanidiales contain 1% horizontally transferred genes with diverse adaptive functions. *eLife Sciences* **8**,1283 (2019).

47. Grigoriev, I. V *et al.* The Genome Portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Res.* **40**, 26–32 (2012).

48. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402 (1997).

49. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

50. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).

51. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

52. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* (2019). doi:10.1093/bioinformatics/btz305

53. Terashima M, Specht M, Naumann B, Hippler M.. Characterizing the anaerobic response of *Chlamydomonas reinhardtii* by quantitative proteomics. *Mol. Cell Proteomics* **9**,1514–1532 (2010).

54. Facchinelli F, Pribil M, Oster U, Ebert NJ, Bhattacharya D, Leister D, Weber APM. 2013. Proteomic analysis of the Cyanophora paradoxa muroplast provides clues on early events in plastid endosymbiosis. Planta 237:637–651.

55. Thomsen, M. C. F. & Nielsen, M. Seq2Logo: A method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res.* **40**, W281–W287 (2012).

56. Färber, A., Young, A. J., Ruban, A. V., Horton, P. & Jahns, P. Dynamics of Xanthophyll-Cycle Activity in Different Antenna Subcomplexes in the Photosynthetic Membranes of Higher Plants (The Relationship between Zeaxanthin Conversion and Nonphotochemical Fluorescence Quenching). *Plant Physiol.* **115**, 1609–1618 (1997).

57. Schindelin, J. *et al.* Fiji: An open-source platform for biological-image analysis. *Nature Methods* **9**, 676–682 (2012).

58. Lichtenthaler, H. K. *et al.* Photosynthetic activity, chloroplast ultrastructure, and leaf characteristics of high-light and low-light plants and of sun and shade leaves. *Photosynth. Res.* **2**, 115–141 (1981).

59. Zhu, J. K. Abiotic Stress Signaling and Responses in Plants. *Cell* **167**, 313–324 (2016).

60. Khatoon, M. *et al.* Quality control of photosystem II: Thylakoid unstacking necessary to avoid further damage to the D1 protein and to facilitate D1 degradation under light stress in spinach thylakoids. *J. Biol. Chem.* **284**, 25343–25352 (2009).

61. Yoshioka-Nishimura, M. & Yamamoto, Y. Quality control of Photosystem II: The molecular basis for the action of FtsH protease and the dynamics of the thylakoid membranes. *Journal of Photochemistry and Photobiology B: Biology* **137**, 100–106 (2014).

62. Bertrand, M. Carotenoid biosynthesis in diatoms. *Photosynthesis Research* **106**, 89–102 (2010).

63. Tsekos, I., Reiss, H. D., Orfanidis, S. & Orologas, N. Ultrastructure and supramolecular organization of photosynthetic membranes of some marine red algae. *New Phytol.* **133**, 543–551 (1996).

64. Bonente, G. *et al.* Analysis of LHcSR3, a protein essential for feedback de-excitation in the green alga Chlamydomonas reinhardtii. *PLoS Biol.* **9**, e1000577 (2011).

65. Hutin, C. *et al.* Early light-induced proteins protect Arabidopsis from photooxidative stress. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 4921–4926 (2003).

66. Engelken, J., Brinkmann, H. & Adamska, I. Taxonomic distribution and origins of the extended LHC (light-harvesting complex) antenna protein superfamily. *BMC Evol. Biol.* **10**, (2010).

67. Carell, T., Burgdorf, L. T., Kundu, L. M. & Cichon, M. The mechanism of action of DNA photolyases. *Current Opinion in Chemical Biology* **5**, 491–498 (2001).

68. Sato, R., Ito, H. & Tanaka, A. Chlorophyll b degradation by chlorophyll b reductase under high-light conditions. *Photosynth. Res.* **126**, 249–59 (2015).

69. Davison, P. A., Hunter, C. N. & Horton, P. Overexpression of β-carotene hydroxylase enhances stress tolerance in Arabidopsis. *Nature* **418**, 203–206 (2002).

70. Willmund, F., Dorn, K. V., Schulz-Raffelt, M. & Schroda, M. The chloroplast DnaJ homolog CDJ1 of Chlamydomonas reinhardtii is part of a multichaperone complex containing HSP70B, CGE1, and HSP90C. *Plant Physiol.* **148**, 2070–2082 (2008).

71. Komenda, J. & Sobotka, R. Cyanobacterial high-light-inducible proteins - Protectors of chlorophyll-protein synthesis and assembly. *Biochim. Biophys. Acta - Bioenerg.* **1857**, 288–295 (2016).

72. Gao, F. *et al.* Identification of conserved and novel microRNAs in Porphyridium purpureum via deep sequencing and bioinformatics. *BMC Genomics* **17**, (2016).

73. Kessler, F., Blobel, G., Patel, H. A. & Schnell, D. J. Identification of two GTP-binding proteins in the chloroplast protein import machinery. *Science (80-.).* **266**, 1035–1039 (1994).

74. Hsu, S.-C., Patel, R., Bédard, J., Jarvis, P. & Inoue, K. Two distinct Omp85 paralogs in the chloroplast outer envelope membrane are essential for embryogenesis in Arabidopsis thaliana. *Plant Signal. Behav.* **3**, 1134–5 (2008).

75. Patel, R., Hsu, S. C., Bédard, J., Inoue, K. & Jarvis, P. The Omp85-related chloroplast outer envelope protein OEP80 is essential for viability in Arabidopsis. *Plant Physiol.* **148**, 235–245 (2008).

76. Ling, Q. *et al.* Ubiquitin-dependent chloroplast-associated protein degradation in

plants. *Science (80-.).* **363**, (2019).

77. Töpel, M., Ling, Q. & Jarvis, P. Neofunctionalization within the Omp85 protein superfamily during chloroplast evolution. *Plant Signal. Behav.* **7**, (2012).

78. Wickett, N. J. *et al.* Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E4859–E4868 (2014).

79. Nishiyama, T. *et al.* The Chara Genome: Secondary Complexity and Implications for Plant Terrestrialization. *Cell* **174**, 448-464.e24 (2018).

80. Andersen, R. A. Diversity of eukaryotic algae. *Biodivers. Conserv.* **1**, 267–292 (1992).

81. Govaerts, R. How many species of seed plants are there? *Taxon* **50**, 1085–1090 (2001).

82. Kenrick, P., Wellman, C. H., Schneider, H. & Edgecombe, G. D. A timeline for terrestrialization: Consequences for the carbon cycle in the Palaeozoic. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 519–536 (2012).

83. Delwiche, C. F. & Cooper, E. D. The evolutionary origin of a terrestrial flora. *Current Biology* **25**, R899–R910 (2015).

84. Lee, D. W. *et al.* Molecular Mechanism of the Specificity of Protein Import into Chloroplasts and Mitochondria in Plant Cells. *Mol. Plant* **12**, 951–966 (2019).

85. Wollman, F. A. An antimicrobial origin of transit peptides accounts for early endosymbiotic events. *Traffic* **17**, 1322–1328 (2016).

86. Mergaert, P., Kikuchi, Y., Shigenobu, S. & Nowack, E. C. M. Metabolic Integration of Bacterial Endosymbionts through Antimicrobial Peptides. *Trends in Microbiology* **25**, 703–712 (2017).

87. Nowack, E. C. M. Paulinella chromatophora - Rethinking the transition from endosymbiont to organelle. *Acta Societatis Botanicorum Poloniae* **83**, 387–397 (2014).

88. Singer, A. *et al.* Massive Protein Import into the Early-Evolutionary-Stage Photosynthetic Organelle of the Amoeba Paulinella chromatophora. *Curr. Biol.* **27**, 2763-2773.e5 (2017).

89. Jarvis, P. & Soll, J. Toc, Tic, and chloroplast protein import. *Biochimica et Biophysica Acta - Molecular Cell Research* **1541**, 64–79 (2001).

90. Day, P. M. & Theg, S. M. Evolution of protein transport to the chloroplast envelope membranes. *Photosynthesis Research* **138**, 315–326 (2018).

91. Robert, V. *et al.* Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. *PLoS Biol.* **4**, 1984–1995 (2006).

92. O'Neil, P. K. *et al.* The POTRA domains of Toc75 exhibit chaperone-like function to facilitate import into chloroplasts. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E4868–E4876 (2017).

93. Sommer, M. S. *et al.* Chloroplast Omp85 proteins change orientation during evolution. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13841–13846 (2011).

94. Gile, G. H., Moog, D., Slamovits, C. H., Maier, U. G. & Archibald, J. M. Dual organellar targeting of aminoacyl-tRNA synthetases in diatoms and cryptophytes. *Genome Biol. Evol.* **7**, 1728–1742 (2015).

95. Martin, W. Evolutionary origins of metabolic compartmentalization in eukaryotes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 847–855 (2010).

748 96. Schnell, D. J., Kessler, F. & Blobel, G. Isolation of components of the chloroplast
749 protein import machinery. *Science (80-.).* **266**, 1007–1012 (1994).

750 97. Shi, L. X. & Theg, S. M. The chloroplast protein import system: From algae to trees.
751 *Biochimica et Biophysica Acta - Molecular Cell Research* **1833**, 314–331 (2013).

752 98. Bullmann, L. *et al.* Filling the gap, evolutionarily conserved Omp85 in plastids of
753 chromalveolates. *J. Biol. Chem.* **285**, 6848–6856 (2010).

754 99. Berardini, T. Z. *et al.* The arabidopsis information resource: Making and mining the
755 "gold standard" annotated reference plant genome. *Genesis* **53**, 474–485 (2015).

756 100. Heddad, M. & Adamska, I. The evolution of light stress proteins in photosynthetic
757 organisms. *Comp. Funct. Genomics Comp Funct Genom* **3**, 504–510 (2002).

758 101. Wu, G. Z. *et al.* Control of retrograde signalling by protein import and cytosolic
759 folding stress. *Nat. Plants* **5**, 525–538 (2019).

760 102. Wiedemann, N. *et al.* Machinery for protein sorting and assembly in the mitochondrial
761 outer membrane. *Nature* **424**, 565–571 (2003).

762 103. Broglie, R., Coruzzi, G., Lamppa, G., Keith, B. & Chua, N. H. Structural Analysis of
763 Nuclear Genes Coding for the Precursor to the Small Subunit of Wheat Ribulose-1,5-
764 Bisphosphate Carboxylase. *Bio/Technology* **1**, 55–61 (1983).

765 104. Coen, D. M., Bedbrook, J. R., Bogorad, L. & Rich, A. Maize chloroplast DNA
766 fragment encoding the large subunit of ribulosebisphosphate carboxylase. *Proc. Natl.*
767 *Acad. Sci.* **74**, 5487–5491 (1977).

768 105. May, T. & Soll, J. 14-3-3 Proteins Form a Guidance Complex with Chloroplast
769 Precursor Proteins in Plants. *Plant Cell* **12**, 53–63 (2000).

770 106. Kikuchi S, Asakura Y, Imai M, Nakahira Y, Kotani Y, Hashiguchi Y, Nakai Y,
771 Takafuji K, Bédard J, Hirabayashi-Ishioka Y, et al.. A Ycf2-FtsHi heteromeric AAA-
772 ATPase complex is required for chloroplast protein import. *Plant Cell* tpc.00357.2018–
773 tpc.00357.2028 (2018).

774 107. de Vries J, Archibald JM, Gould SB. The carboxy terminus of YCF1 contains a motif
775 conserved throughout >500 million years of streptophyte evolution. *Genome Biol Evol.*
776 **9**, 473-479 (2017).

777 108. de Vries J, Sousa FL, Bölter B, Soll J, Gould SB. 2015. YCF1: A Green TIC? *Plant*
778 *Cell* **27**, 1827–1833 (2015).

779 109. Chou M-L, Fitzpatrick LM, Tu SL, Budziszewski G, Potter-Lewis S, Akita M, Levin
780 JZ, Keegstra K, Li H-M. Tic40, a membrane-anchored co-chaperone homolog in the
781 chloroplast protein translocon. EMBO J 22:2970–2980 (2003).
782