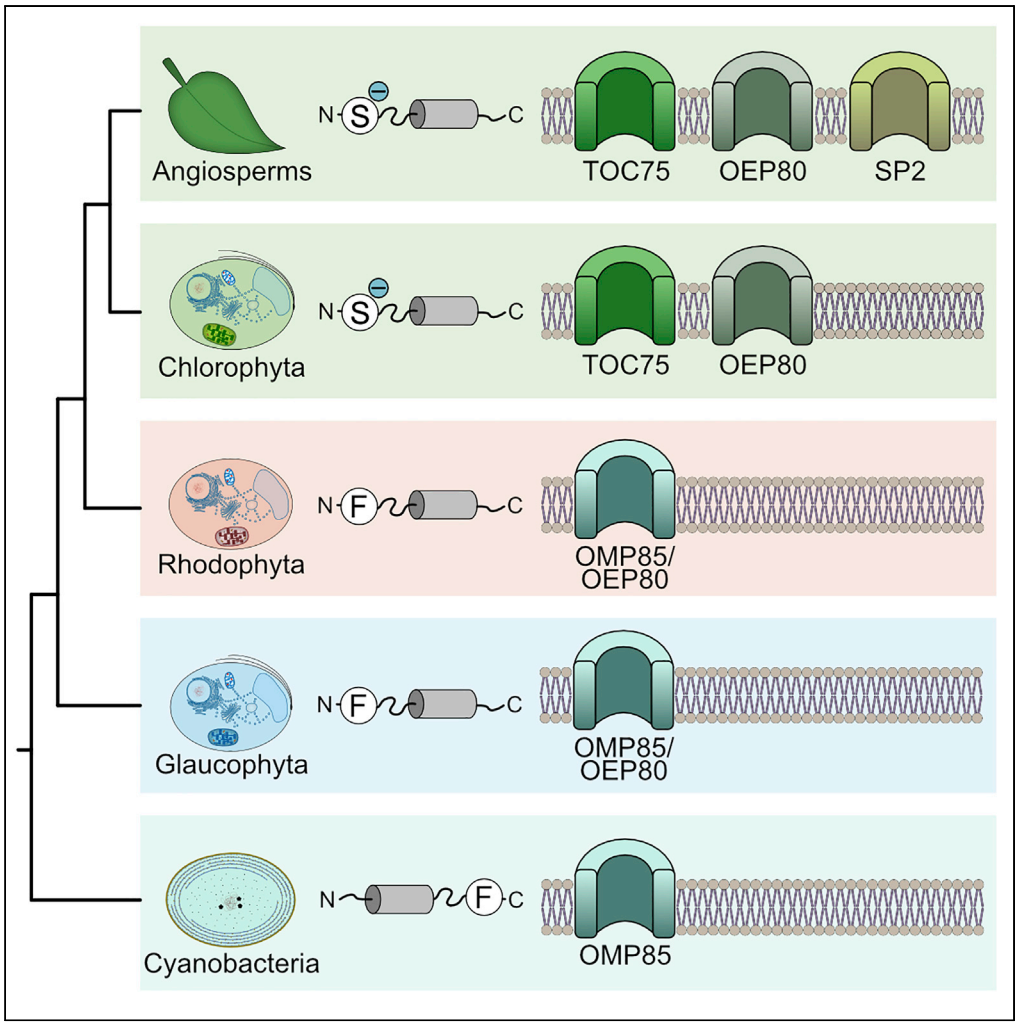
iScience

CellPress

Article

Major Changes in Plastid Protein Import and the Origin of the Chloroplastida



Michael Knopp, Sriram G. Garg, Maria Handrich, Sven B. Gould

gould@hhu.de

HIGHLIGHTS

Chloroplastida evolved a dual system, Toc75/Oep80, for high throughput protein import

Loss of F-based targeting led to dual organelle targeting using a single ambiguous NTS

Relaxation of functional constraints allowed a wider Toc/Tic modification

A broad response to high-light stress appears unique to Chloroplastida

Knopp et al., iScience 23, 100896
March 27, 2020
https://doi.org/10.1016/j.isci.2020.100896

iScience

CellPress

Article

Major Changes in Plastid Protein Import and the Origin of the Chloroplastida

Michael Knopp,[1,2] Sriram G. Garg,[1,2] Maria Handrich,[1] and Sven B. Gould[1,3,*]

SUMMARY

Core components of plastid protein import and the principle of using N-terminal targeting sequences are conserved across the Archaeplastida, but lineage-specific differences exist. Here we compare, in light of plastid protein import, the response to high-light stress from representatives of the three archaeplastidal groups. Similar to land plants, *Chlamydomonas reinhardtii* displays a broad response 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 Chloroplastida encode both Toc75 and Oep80 in parallel and suggest that elaborate high-light stress response is supported by changes in plastid protein import. We propose the origin of a phenylalanine-independent import pathway via Toc75 allowed higher import rates to rapidly service high-light stress, but with the cost of reduced specificity. Changes in plastid protein import define the origin of the green lineage, whose greatest evolutionary success was arguably the colonization of land.

INTRODUCTION

Mitochondria and plastids are of endosymbiotic origin and compartments surrounded by a double membrane (Zimorski et al., 2014; Archibald, 2015). Most possess their own genomes, but the bulk of their former coding capacity was either lost or integrated into the nuclear genome (Timmis et al., 2004; Martin and Herrmann, 1998). As a consequence, most of their proteins are post-translationally imported. Guiding of precursor proteins to the mitochondrial matrix or plastid stroma typically relies on N-terminal targeting sequences (NTS) (Schleiff and Becker, 2011; Dudek et al., 2013; Paila et al., 2015), although some exceptions are known (Goldberg et al., 2008; Hamilton et al., 2014; Garg et al., 2015). Archaeplastidal plastids have a monophyletic origin (Rodríguez-Ezpeleta et al., 2005; Jackson and Reyes-Prieto, 2014; Sánchez-Baracaldo et al., 2017), which is also evident from the conserved nature of plastid import components, a reliable indicator for the monophyly of organelles (Cavalier-Smith, 1999; Kalanon and McFadden, 2008; Gould et al., 2015).

Although they share a single origin, the plastids of the three algal lineages have evolved considerable differences since their divergence more than a billion years ago (Gibson et al., 2017; de Vries et al., 2016). These include, but are not limited to, (1) the thickness of a remaining peptidoglycan layer (Pfanzagl et al., 1996; Hirano et al., 2016), (2) the localization of starch deposits (Suzuki and Suzuki, 2013), (3) the coding capacity of their genomes (Timmis et al., 2004; Allen et al., 2012), (4) pigment composition and the types of antenna complexes used (Tomitani et al., 1999), (5) the absence or presence of a xanthophyll cycle (Goss and Jakob, 2010), and (6) the composition of the protein import machinery (Day et al., 2014; Kikuchi et al., 2013). 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. Although most information about plastid protein targeting stems from the green lineage (Köhler et al., 2015), 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 stroma. Rhodophytes and glaucophytes employ a single amino-acid-based motif to target proteins to their plastids (Gould et al., 2015; Köhler et al., 2015; Steiner et al., 2005; Wunder et al., 2007). In most cases this amino acid is a phenylalanine, less frequently other bulky aromatic amino acids (Köhler et al., 2015; Gruber et al., 2007). The F-based motif is found at the very N-terminus of the NTS (Figure 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* (Patron and Waller, 2007). It is uncertain why the F-based motif was lost in Chloroplastida, but it came with several changes such as a rise in

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

[2]These authors contributed equally

[3]Lead Contact

*Correspondence: gould@hhu.de

https://doi.org/10.1016/j.isci.2020.100896

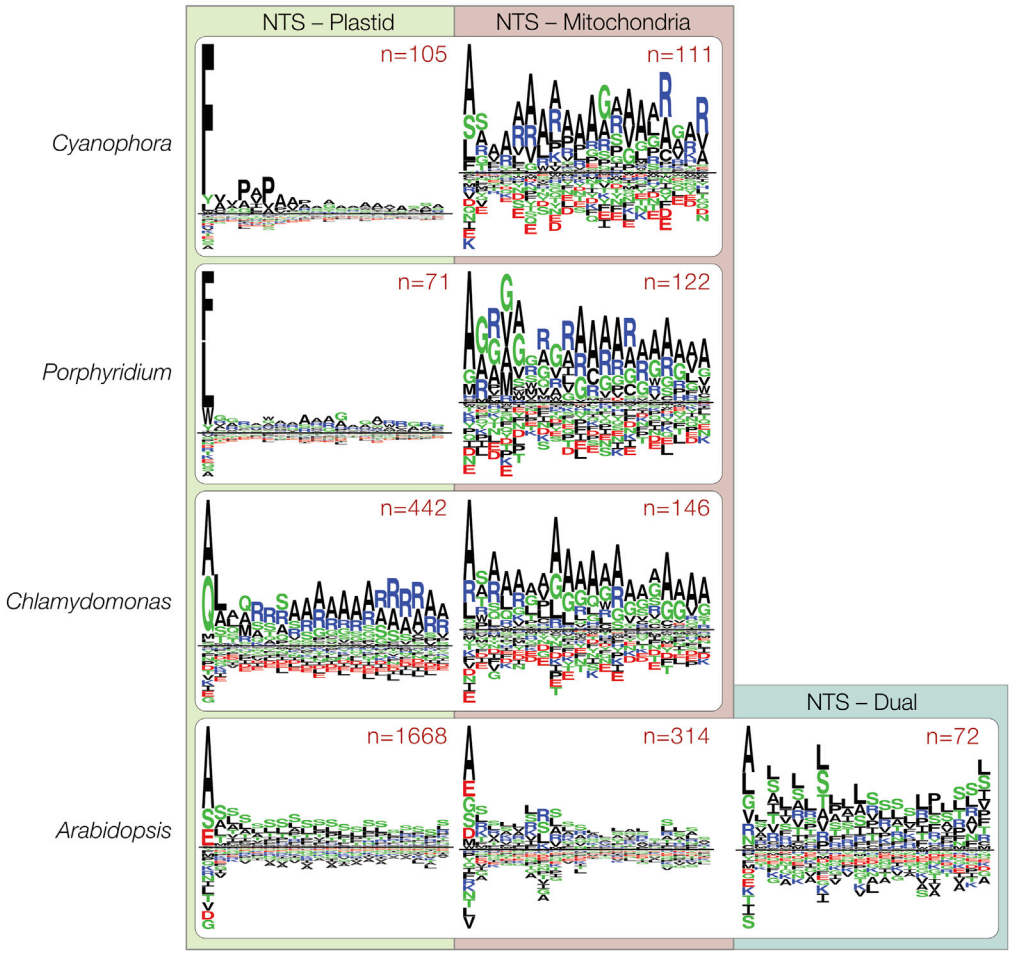


Figure 1. Targeting Motifs of Organelle Targeted Proteins

NTSs of plastid- or mitochondria-targeted proteins of *C. paradoxa*, *P. purpureum*, *C. reinhardtii* and *A. thaliana*. The four species showcase the NTS for plastid- or mitochondria-targeting in the *Glaucophyta*, *Rhodophyta*, *Chlorophyta* and *Streptophyta*, respectively. Although an F-based plastid targeting motif is evident in *Glaucophyta* and *Rhodophyta*, it was lost in the green lineage.

phosphorylatable serine residues that might help in avoiding erroneous targeting to the mitochondria (Garg and Gould, 2016; Lee et al., 2006).

Despite a tendency toward 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 (Karniely and Pines, 2005; Carrie and Small, 2013), in plants also the two organelles of endosymbiotic origin. About 100 proteins are dually targeted to the mitochondria and plastids of *Arabidopsis thaliana* after their translation (Carrie and Small, 2013; Carrie et al., 2009). 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) (Schleiff and Becker, 2011; Garg and Gould, 2016). In *A. thaliana*, a duplicate of the Toc64 receptor localizes to the outer mitochondrial membrane and now functions in mitochondrial import (Chew et al., 2004). Both *Arabidopsis* organelles also use the same targeting-associated PURPLE ACID PHOSPHATASE2 (*At*PAP2) at their outer membranes (Sun et al., 2012; Law et al., 2015). 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

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	8	LHC protein complex	a,b	Beta-Carotin, lutein, neoxanthin, violaxanthin, antheraxanthin, zeaxanthin	Stacked	Starch
Porphyridium purpureum (Rhodophyte)	8,355	199	ND	Phycobilisomes	a	Beta-Carotin, zeaxanthin	Unstacked, equidistant and single	Glycogen, floridean starch
Cyanophora paradoxa (Glaucophyte)	27,921 (25,831)	136	44	Phycobilisomes	a	Beta-Carotin, zeaxanthin	Unstacked, equidistant and single	Floridean starch

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

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 Chloroplastida, with a series of major changes connected to the origin of the green lineage.

RESULTS

Adaptive Changes of Common Photosynthetic Pigments upon High-Light Stress

Plants react in particular to changes in light intensity (Lichtenthaler et al., 1981; Zhu, 2016). To analyze 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 μmol photons $m^{-2}s^{-1}$ under a 12/12-h day-night cycle and at 20°C. Through rapid light curves we assessed that at 600 μmol photons $m^{-2}s^{-1}$, a saturation of the photosystems was reached in all three species (Figure S1). For the high-light stress treatment, the algae were therefore exposed to 600 μmol photons $m^{-2}s^{-1}$ for 1 h. For comparison we determined the pigment profiles from cultures that were either 6 h into the night or 6 h into the day phase.

The glaucophyte *C. paradoxa* showed no significant change in pigment concentration or composition, neither at night nor after light stress (Figure 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 (Figure 2A). Only in the green alga *C. reinhardtii* the pigment composition changed significantly, especially upon high-light stress (Figure 2A). Here in particular the xanthophyll cycle—the enzyme-driven and reversible conversion of violaxanthin into zeaxanthin—was evident, a component of non-photochemical quenching thought to be absent in

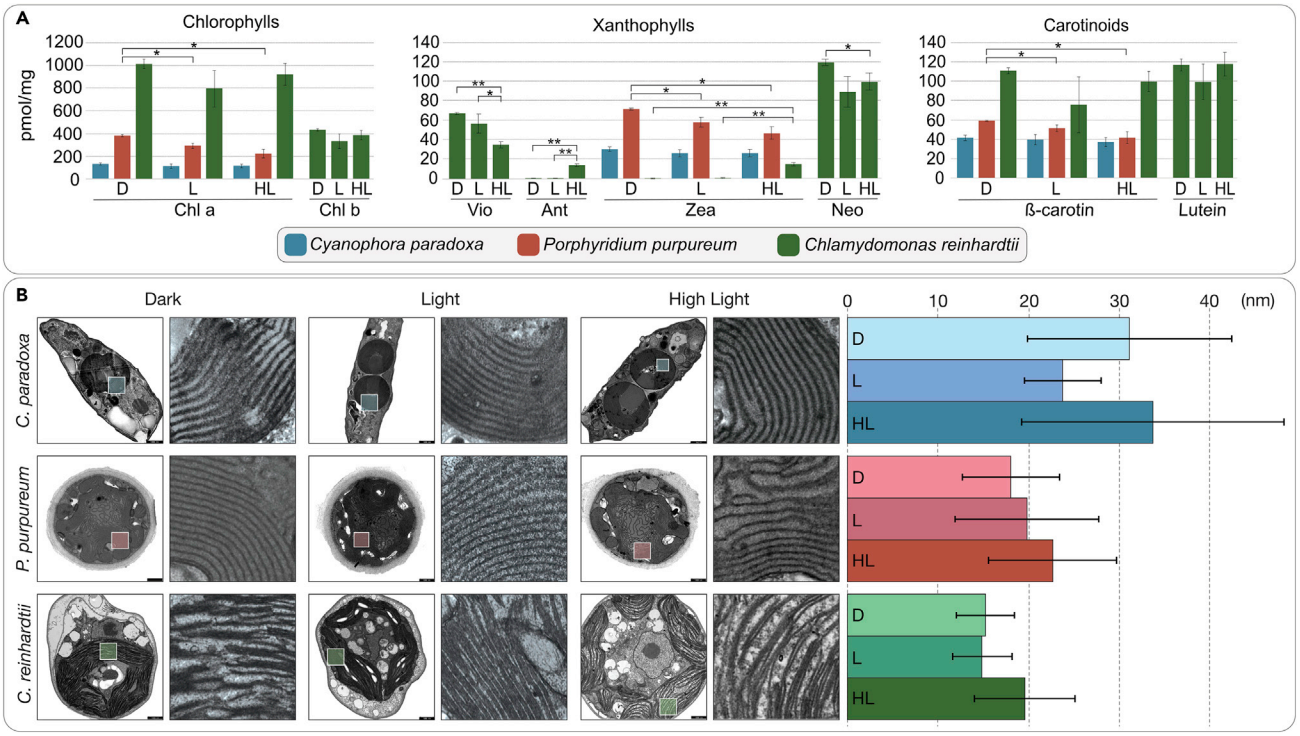


Figure 2. Pigment Profiles and Analysis of Thylakoid Stack Distance during High-Light Stress

(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 contrary, in *C. reinhardtii* all three types of pigment change their concentration significantly and e.g., the stepwise reduction of violaxanthin (Vio) to antheraxanthin (Ant) and zeaxanthin (zea) is evident.

(B) Cells from the three different conditions were fixed and analyzed using transmission electron microscopy and distances between the thylakoids were measured using Fiji. A statistically significant increase in thylakoid distance upon high-light stress is only observed in *C. reinhardtii*, although a similar but less significant trend is observed in the red alga *P. purpureum*. (*<0.05 ** < 0.001 *** < 0.0001). pmol/mg, picomol/milligram of dry weight. Scale bars equal 1000 nm.

glauco- and rhodophytes (Goss and Jakob, 2010). Concentrations of chlorophylls and carotenoids actually increased under high-light stress in *C. reinhardtii*, demonstrating their rapid 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 (Khatoon et al., 2009). 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 (Yoshioka-Nishimura and Yamamoto, 2014). Algae form different types of thylakoid stacks (Bertrand, 2010; Tsekos et al., 1996) 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 neighboring 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 (Figure 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 (Figure 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 the number of upregulated genes during high-light conditions. For each condition a clear separation was observed, and a specific gene set was found to be upregulated relative to the average expression of each gene over all conditions (Figure 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

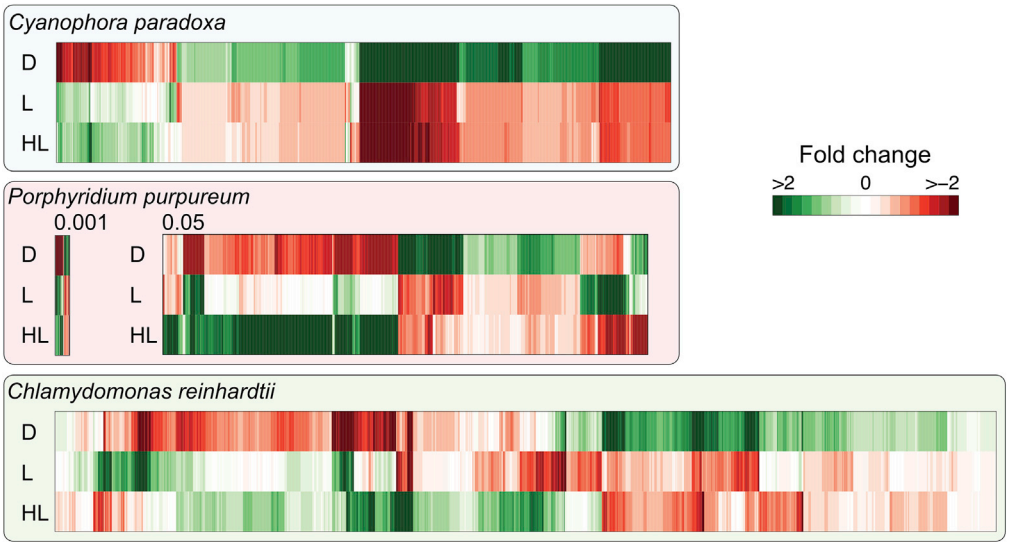


Figure 3. Differentially Expressed Genes of *C. reinhardtii*, *C. paradoxa*, and *P. purpureum*

Visualization of all differentially expressed genes of *C. reinhardtii*, *C. paradoxa*, and *P. purpureum*, colored according to the logarithmic fold change relative to the average expression across all conditions, mean centered and color coded in log2 space. For *C. reinhardtii* and *C. paradoxa*, the fold change's significance of all visualized transcripts is at least 0.001. For *P. purpureum*, the significance cutoff was lowered to 0.05, because the original cutoff revealed only 90 differentially expressed genes. *C. reinhardtii* shows distinct sets of genes, each tailored toward one of the tested light conditions. *C. paradoxa* and *P. purpureum*, on the other hand, do not show such an adaptation to altering light conditions, especially not to high-light ($p = 0.001$). *C. paradoxa* does not change much of its gene expression between daylight and high-light conditions, showing its lack of adaptation. Although *P. purpureum* expresses a set of genes only during high-light conditions, their differential expression was only detectable by lowering the significance cutoff. Even if all differentially expressed genes of *P. purpureum* are considered, its transcriptional changes during high-light remain minor.

differentially expressed (of a total of 2,810 differential expressed transcripts between any conditions), 274 values of which were significantly upregulated compared with daylight conditions (Table S3). The upregulated photoprotective proteins include stress-related chlorophyll-binding proteins 1 and 3 involved in energy-dependent quenching to dissipate excess energy (Bonente et al., 2011), members of the early-light inducible protein family (Elip), and ancestral homologs of the non-photochemical quenching-associated PSBS/LHCSR3 family (Hutin et al., 2003; Engelken et al., 2010), a CPD photolyase class II that reverses the formation of pyrimidine dimers that result from the exposure to strong UV radiation (Carell et al., 2001), and chlorophyll b reductases and beta-carotene hydroxylases that prevent over-excitation of the photosystem and protect the cells from high-light intensities (Sato et al., 2015; Davison et al., 2002). 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 together form a multichaperone complex (Willmund et al., 2008) (Table S3).

In the glaucophyte *C. paradoxa*, most of the 1,463 differentially expressed transcripts were found upregulated during darkness in correspondence to nightly proliferation (Figure 3). 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* (Table S3). In comparison to the green alga, fewer proteins involved in photosynthesis regulation and photoprotection were found to be upregulated during high-light stress, but they also included several Elip proteins.

The identification of differentially expressed genes in *P. purpureum* proved 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 (when compared with the same cutoff of $p = 0.001$). For the transition from daylight to high-light, a total of 38 transcripts were identified (Table S3). The most notable genes that were upregulated during high-light stress were a high-light inducible protein (Hlip) involved in

non-photochemical quenching (Komenda and Sobotka, 2016) and several heat shock proteins (HSP70). The response to light stress was far weaker than in the other two algae ($p = 0.001$), but *P. purpureum* might regulate its RNA levels through the extensive use of miRNAs (Gao et al., 2016), 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. Although *C. reinhardtii* upregulated the synthesis of several photosynthesis and plastid-related proteins during light and high-light conditions, *C. paradoxa* and *P. purpureum* upregulated only a few. In the case of *C. paradoxa*, the biggest notable difference was the focus on protein biosynthesis during darkness/night. The 50 most highly upregulated proteins during the night were ribosomal proteins (approximately 90%), indicating an increase in overall protein biosynthesis and proliferation (Table S4). We observed photosynthesis machinery as well as photoprotection components to be among the most upregulated proteins in combination only in *C. reinhardtii*, illustrating the chlorophyte's more elaborate and immediate ability to adapt to changing light conditions compared with the other two screened algae.

The Red Toc75 Is an Oep80 and Toc75 Unique to Chloroplastida

Most members of the *Arabidopsis*Toc75 family have been characterized. This includes the main import channel of the outer membrane, Toc75 (Kessler et al., 1994) (TOC75-III, TAIR: At3g46740), as well as Oep80 (TOC75-V, TAIR: At5g19620) whose exact function remains unresolved while the protein is essential for plant viability (Hsu et al., 2008; Patel et al., 2008), and the most recently characterized SP2 (TAIR: At3g44160), which serves protein export for chloroplast-associated protein degradation (Ling et al., 2019). The situation in rhodophytes and glaucophytes differs. They do not seem to encode the same number of Toc75 homologs (Day et al., 2014; Töpel et al., 2012).

We collected 77 eukaryotic proteins of the Toc75 and Oep80 family from 44 eukaryotic species and rooted them against their cyanobacterial homologs for the construction of a phylogenetic tree. The single glaucophyte sequence sits basal to all others, whereas the rhodophyte sequences form a well-supported group that is sister to all chloroplastidal sequences (Figure 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 the streptophyte alga *Chara braunii* related to the ancestor of land plants (Wickett et al., 2014; Nishiyama et al., 2018) (Figure 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 13 glaucophyte species have been recognized (with >100,000, 500–1,000 and about a dozen that remain to be described, respectively) (Andersen, 1992). Another 400,000 land plant species (Govaerts, 2001) evolved since the conquering of land some 480 million years ago (Kenrick et al., 2012; Delwiche and Cooper, 2015). 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 (Schleiff and Becker, 2011). The origin of the mitochondrial NTS is uncertain, but its positive charge was an early requirement to overcome the bioenergetic inner mitochondrial membrane (Garg et al., 2015). The most N-terminal domain carries the charged residues critical for distinguishing between mitochondrial- and plastid-targeting (Figure 1), whereas the C-terminus is exchangeable (Lee et al., 2019). 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) (Wollman, 2016), as both share similarities in terms of charged amino acid residues, the ability to form

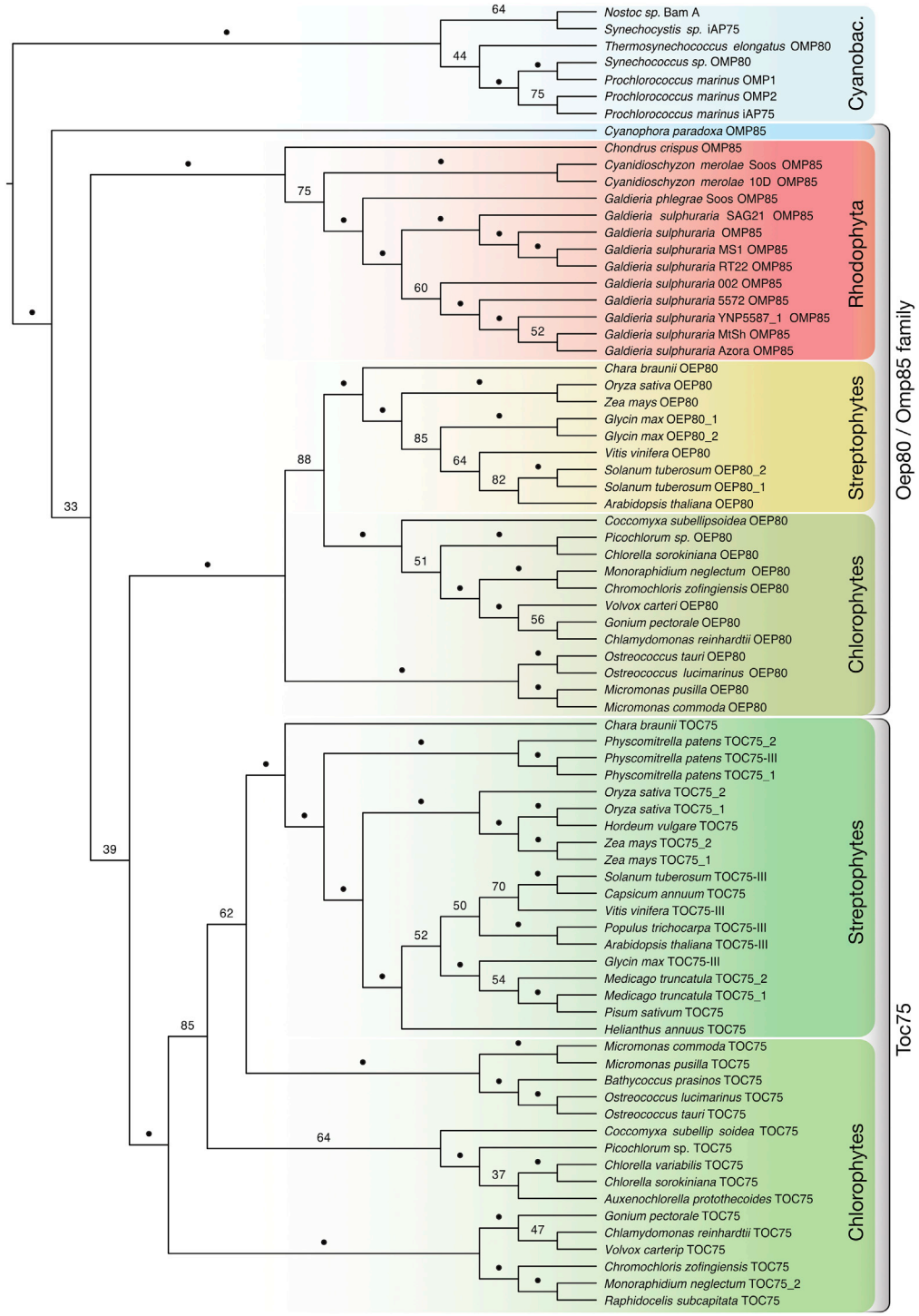


Figure 4. Phylogenetic Analysis of Oep80 and Toc75 Homologs

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. Bootstrap values > 90 are represented by dots.

amphiphilic α-helices, and because they are frequently identified in host-endosymbiont relationships (Mergaert et al., 2017). One example regarding the latter is *Paulinella chromatophora*, whose chromatophore origin is independent from that of the Archaeplastida and much younger (Nowack, 2014). Two types of NTSs were identified that target nuclear-encoded proteins to the chromatophore, but both are not related to the simultaneously identified AMPs (Singer et al., 2017), 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 (Jarvis and Soll, 2001; Day and Theg, 2018). Toc75, the β-barrel import pore in the outer membrane, is of prokaryotic origin and a member of the Omp85 superfamily (Day et al., 2014). Some bacterial Omp85's recognize their substrates through a C-terminal phenylalanine (Robert et al., 2006) and evidence is emerging that the POTRA domains of Toc75 act as binding sites for the NTS (O'Neil et al., 2017). If we recall that the phenylalanine-based motif is retained in rhodophytes and glaucophytes (Wunder et al., 2007), 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 (Sommer et al., 2011). 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 as a Result of Losing the F-based Motif

The use of an 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 (Garg et al., 2015; Lee et al., 2006; Law et al., 2015). Many proteins, however, remain dual targeted in *Arabidopsis* (Carrie and Small, 2013), 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 NTS (Gile et al., 2015).

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-localize or establish entire new pathways (Martin, 2010). Our phylogenetic analysis (Figure 5) shows there is no apparent preference regarding the evolutionary origin (cyanobacterial vs. proteobacterial) and flow of dual-targeted proteins: as much proteins of cyanobacterial origin are targeted to the mitochondria as there are *vice versa* (Figure 5).

Dual-targeted proteins are largely part of the transcription and translation machinery (Carrie and Small, 2013), which might include the plastid-associated polymerases whose dual-targeting in Chloroplastida might be an ancestral trait of the lineage (Nishiyama et al., 2018). 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 simultaneous addressing of the two organelles of endosymbiotic origin, which we speculate offers 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 (Schnell et al., 1994). Conserved homologs across all Chloroplastida were quickly identified (Kalanon and McFadden, 2008; Shi and Theg, 2013) but required more effort across the diversity of the Archaeplastida. Through the identification of an Omp85 homolog in algae with secondary red plastids, it became evident that all phototrophic eukaryotes harbor beta-barrel-forming proteins of an extended Omp85 family that form the import pore in the outer plastid membrane (Bullmann et al., 2010), but with a decisive difference regarding the number of encoded homologs.

Our phylogenetic analysis of Toc75 and Oep80 supports previous analyses without the need of any sequence trimming. It demonstrates the clear-cut, likely also functional, separation between the Toc75

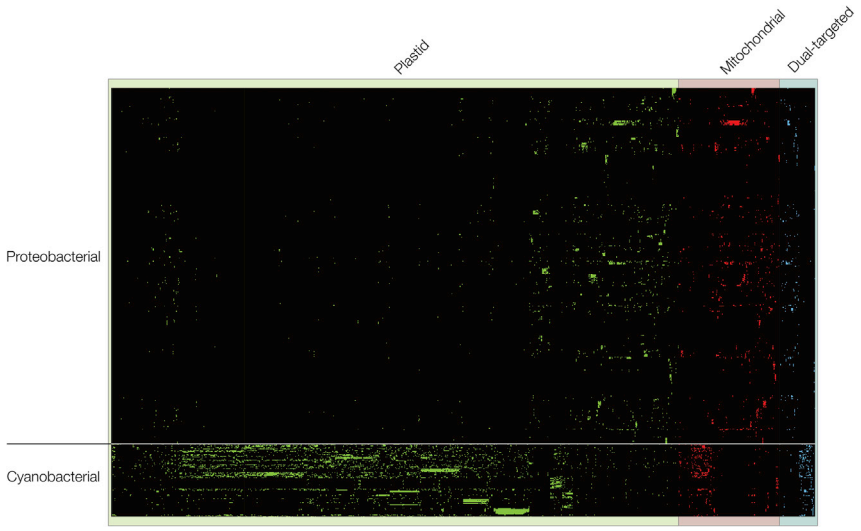


Figure 5. Phylogenetic Origin of Plastid- and Mitochondria-Targeted Proteins of *Arabidopsis*

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, whereas 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 alphaproteobacteria, with one-half (45%) showing a cyanobacterial origin and the other half (55%) showing an alphaproteobacterial origin.

and Oep80 proteins of Chloroplastida (Day et al., 2014). The red sequences are closer to their prokaryotic homologs, and the green Toc75 is further derived. From the perspective of phylogeny, there is little doubt that Toc75 is unique to the green lineage and originated from the duplication of a green lineage-specific Oep80/Omp85. This suggests a division of labor at the outer chloroplast membrane not found in rhodoplasts or cyanelles, the benefits of which are plenty. Glauco- and rhodophytes work with a single import pore, whereas *Arabidopsis* and its green relatives encode a single full-length Toc75 and a single full-length Oep80. Both of the latter are expressed at high levels in a conserved ratio and in the different tissues according to the gene expression atlas of the TAIR database (Berardini et al., 2015). Their presence is needed simultaneously and appears synchronized.

We speculate that the duplication of Omp85/Oep80 allows for a more efficient, faster, and versatile protein import. It might be a prerequisite for the elaborate response to high-light stress, which our data support (Figures 2 and 3). A response to high-light stress is evident in all three lineages (Figure 3), but differs in quantity and detail. *Chlamydomonas* not only alters its gene expression network the most upon high-light stress but also focuses more on photosynthesis maintenance and protection (Table S3), reacts less stressed, and rapidly synthesizes pigments denovo (Figure 2). The upregulation of Elips that are of cyanobacterial origin occurs in all three lineages, but they were only expanded and diversified in the green lineage (Heddad and Adamska, 2002). Retrograde signaling (a critical part of the response to high-light stress) is limited by the plastid's import capacity (Wu et al., 2019), highlighting the direct dependence.

If Oep80's main duty is indeed the integration of beta-barrel proteins (and maybe other delicate substrates of unknown nature), then it releases Toc75 from this job. This would then mirror the situation in mitochondria. Here, Tom40 acts as the main import channel, whereas Sam50 receives beta-barrel proteins with a complicated topology from Tom40 for their integration into the outer mitochondrial membrane (Wiedemann et al., 2003). The division of labor appears more effective than the simple increase in number of a single import gateway. This could have allowed the endosymbiotic gene transfer of the small subunit of RubisCo to the nucleus, a trademark of the green lineage (Broglie et al., 1983; Coen et al., 1977). The sheer amount of RbcS protein required to be imported might simply overstrain the Omp85/Oep80 of rhodo- and

glaucophytes, and its gene transfer from the plastid to the nucleus is hence selected against. These patterns allow to speculate on the sequence of evolutionary events.

Initially a duplication of the ancestral green Omp85 occurred and both paralogs might have performed the same duty early on. Mutations in one of the two copies led to an independence of F-based targeting, alternative substrate recognition, the emergence of NTS phosphorylation (Garg and Gould, 2016), and a cytosolic 14-3-3/Hsp70-based guidance complex (May and Soll, 2000) that we predict is unique to the green lineage, too. The plastid-encoded Tic214(YCF1)/YCF2/FtsHi complex emerged early in chlorophyte evolution, too, maybe through the duplication of an early Tic20-like protein (Wunder et al., 2007; Kikuchi et al., 2018). The components of this complex are highly diverse, except for a C-terminal motif, and were entirely lost in grasses without impacting protein import (de Vries et al., 2015, 2017). Other components were added such as Tic40 that further increases import efficiency (Chou et al., 2003) and which is absent from rhodo- and glaucophytes (Kalanon and McFadden, 2008). Ever more plastid proteins went via the Toc75 route, apart from maybe some slow folding proteins of the outer-membrane that continued to be integrated via Oep80. Noteworthy, however, both import pathways remain linked in their function (Day et al., 2019). A more recent extension was the emergence of the CHLORAD pathway (chloroplast-associated protein degradation). Its central component, SP2, is an Omp85 paralog, too, which however lacks the POTRA domains (Ling et al., 2019). It likely emerged in angiosperms and might facilitate the remodeling of plastids (e.g., of a chloroplast to a chromoplast), a feature that in its entirety is unique to the evolutionary more recent land plants and their embryoplast (de Vries et al., 2016). Therefore, the implementation of yet another plastid protein transport pathway based on an Omp85/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 basicToc/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 (1) the loss of phenylalanine-based targeting and (2) the emergence of new import machinery components including Tic40, a plastid-encoded Tic214, and a Toc75 that evolved from the duplication of the ancestral green Omp85. We suggest 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 an import pathway for nuclear-encoded proteins that permitted higher import rates at the expense of targeting specificity. The main import pore of the green plastid, Toc75, was freed from dealing with slow-folding proteins of the outer membrane and permitting 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 Omp85 homologs, Oep80 and Toc75, where rhodophytes and glaucophytes use only one. The transition to life on land was a transition to high-light stress. Responses to high-light require the efficient and immediate import of over a hundred nuclear-encoded plastid proteins simultaneously after retrograde plastid signaling. This, we speculate, was realized by the implementation of an efficient plastid import pathway that enabled the evolutionary success of the Chloroplastida.

Limitations of the Study

Transcriptome assemblies profit immensely from reference genomes during assembly. Since current reference genomes of *P. purpureum* and *C. paradoxa* still lack quality, we assembled the trancriptomes *de novo*. By updating reference genomes and improving their quality, we might get a better insight into light-stress-induced gene regulation within the three major algal lineages. Many components of protein targeting to the mitochondrion and/or plastid have been identified but for a few their exact function remains unclear. Furthermore, direct protein import kinetics via proteomic studies are still limited by reliable plastid isolation protocols for all three species.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

DATA AND CODE AVAILABILITY

Transcriptomes are available via the Sequence Read Archive of NCBI (https://www.ncbi.nlm.nih.gov/sra) with the accession number PRJNA509798.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.100896.

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).

AUTHOR CONTRIBUTIONS

SBG together with SGG designed the study and drafted the manuscript. MH cultured the algae, did microscopy, light stress experiments, and early stages of transcriptome assembly. MRK did the phylogenomic analysis of Omp85/Oep80 homologs and dual-, mitochondria-, or plastid-targeted proteins as well as the identification of differentially expressed genes during the three tested light conditions.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 31, 2019
Revised: January 9, 2020
Accepted: February 4, 2020
Published: March 27, 2020

REFERENCES

Allen, D.K., Laclair, R.W., Ohlrogge, J.B., and Shachar-Hill, Y. (2012). 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.

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

Archibald, J.M. (2015). Endosymbiosis and eukaryotic cell evolution. Curr. Biol. *25*, R911–R921.

Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., and Huala, E. (2015). The arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. Genesis *53*, 474–485.

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

Bonente, G., Ballottari, M., Truong, T.B., Morosinotto, T., Ahn, T.K., Fleming, G.R., Niyogi, K.K., and Bassi, R. (2011). Analysis of LHcSR3, a protein essential for feedback de-excitation in the green alga Chlamydomonas reinhardtii. PLoS Biol. *9*, e1000577.

Broglie, R., Coruzzi, G., Lamppa, G., Keith, B., and Chua, N.H. (1983). Structural analysis of nuclear genes coding for the precursor to the small subunit of wheat ribulose-1,5-bisphosphate carboxylase. Biotechnology *1*, 55–61.

Bullmann, L., Haarmann, R., Mirus, O., Bredemeier, R., Hempel, F., Maier, U.G., and Schleiff, E. (2010). Filling the gap, evolutionarily conserved Omp85 in plastids of chromalveolates. J. Biol. Chem. *285*, 6848–6856.

Carell, T., Burgdorf, L.T., Kundu, L.M., and Cichon, M. (2001). The mechanism of action of DNA photolyases. Curr. Opin. Chem. Biol. *5*, 491–498.

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

Carrie, C., Kühn, K., Murcha, M.W., Duncan, O., Small, I.D., O'Toole, N., and Whelan, J. (2009). Approaches to defining dual-targeted proteins in Arabidopsis. Plant J. *57*, 1128–1139.

Cavalier-Smith, T. (1999). Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. J. Eukaryot. Microbiol. *46*, 347–366.

Chew, O., Lister, R., Qbadou, S., Heazlewood, J.L., Soll, J., Schleiff, E., Millar, A.H., and Whelan, J. (2004). A plant outer mitochondrial membrane protein with high amino acid sequence identity to a chloroplast protein import receptor. FEBS Lett. *557*, 109–114.

Chou, M.L., Fitzpatrick, L.M., Tu, S.L., Budziszewski, G., Potter-Lewis, S., Akita, M., Levin, J.Z., Keegstra, K., and Li, H.M. (2003). Tic40, a membrane-anchored co-chaperone homolog in the chloroplast protein translocon. EMBO J. *22*, 2970–2980.

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

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

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

Day, P.M., Inoue, K., and Theg, S.M. (2019). Chloroplast outer membrane β-barrel proteins use components of the general import Apparatus. Plant Cell *31*, 1845–1855.

Day, P.M., Potter, D., and Inoue, K. (2014). Evolution and targeting of omp85 homologs in the chloroplast outer envelope membrane. Front. Plant Sci. *5*, 1–18.

Delwiche, C.F., and Cooper, E.D. (2015). .The evolutionary origin of a terrestrial flora. Curr. Biol. *25*, R899–R910.

Dudek, J., Rehling, P., and van der Laan, M. (2013). Mitochondrial protein import: common principles and physiological networks. Biochim. Biophys. Acta *1833*, 274–285.

Engelken, J., Brinkmann, H., and Adamska, I. (2010). .Taxonomic distribution and origins of the extended LHC (light-harvesting complex) antenna protein superfamily. BMC Evol. Biol. *10*, 233.

Gao, F., Nan, F., Feng, J., Lv, J., Liu, Q., and Xie, S. (2016). Identification of conserved and novel microRNAs in Porphyridium purpureum via deep sequencing and bioinformatics. BMC Genomics *17*, 612.

Garg, S.G., and Gould, S.B. (2016). Therole of charge in protein targeting evolution. Trends Cell Biol. *26*, 894–905.

Garg, S., Stölting, J., Zimorski, V., Rada, P., Tachezy, J., Martin, W.F., and Gould, S.B. (2015). Conservation of transit peptide-Independent protein import into the mitochondrial and

hydrogenosomal matrix. Genome Biol. Evol. *7*, 2716–2726.

Gibson, T.M., Shih, P.M., Cumming, V.M., Fischer, W.W., Crockford, P.W., Hodgskiss, M.S.W., Wörndle, S., Creaser, R.A., Rainbird, R.H., Skulski, T.M., and Halverson, G. (2017). Precise age of Bangiomorpha pubescens dates the origin of eukaryotic photosynthesis. Geology *46*, 135–138.

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

Goldberg, A.V., Molik, S., Tsaousis, A.D., Neumann, K., Kuhnke, G., Delbac, F., Vivares, C.P., Hirt, R.P., Lill, R., and Embley, T.M. (2008). Localization and functionality of microsporidian iron-sulphur cluster assembly proteins. Nature *452*, 624–628.

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

Gould, S.B., Maier, U.G., and Martin, W.F. (2015). Protein import and the origin of red complex plastids. Curr. Biol. *25*, R515–R521.

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

Gruber, A., Vugrinec, S., Hempel, F., Gould, S.B., Maier, U.G., and Kroth, P.G. (2007). Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. Plant Mol. Biol. *64*, 519–530.

Hamilton, V.N., Singha, U.K., Smith, J.T., Weems, E., and Chaudhuri, M. (2014). Trypanosome alternative oxidase possesses both an N-terminal and internal mitochondrial targeting signal. Eukaryot. Cell *13*, 539–547.

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

Hirano, T., Tanidokoro, K., Shimizu, Y., Kawarabayasi, Y., Ohshima, T., Sato, M., Tadano, S., Ishikawa, H., Takio, S., Takechi, K., et al. (2016). Moss chloroplasts are surrounded by a peptidoglycan wall containing D-amino acids. Plant Cell *28*, 1521–1532.

Hsu, S.C., Patel, R., Bédard, J., Jarvis, P., and Inoue, K. (2008). Two distinct Omp85 paralogs in the chloroplast outer envelope membrane are essential for embryogenesis in Arabidopsis thaliana. Plant Signal. Behav. *3*, 1134–1135.

Hutin, C., Nussaume, L., Moise, N., Moya, I., Kloppstech, K., and Havaux, M. (2003). Early light-induced proteins protect Arabidopsis from photooxidative stress. Proc. Natl. Acad. Sci. U S A *100*, 4921–4926.

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

Jarvis, P., and Soll, J. (2001). Toc, Tic, and chloroplast protein import. Biochim. Biophys. Acta *1541*, 64–79.

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

Karniely, S., and Pines, O. (2005). Single translation-dual destination: mechanisms of dual protein targeting in eukaryotes. EMBO Rep. *6*, 420–425.

Kenrick, P., Wellman, C.H., Schneider, H., and Edgecombe, G.D. (2012). A timeline for terrestrialization: consequences for the carbon cycle in the Palaeozoic. Philos. Trans. R. Soc. B Biol. Sci. *367*, 519–536.

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

Khatoon, M., Inagawa, K., Pospišil, P., Yamashita, A., Yoshioka, M., Lundin, B., Horie, J., Morita, N., Jajoo, A., Yamamoto, Y., and Yamamoto, Y. (2009). 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.

Kikuchi, S., Asakura, Y., Imai, M., Nakahira, Y., Kotani, Y., Hashiguchi, Y., Nakai, Y., Takafuji, K., Bédard, J., Hirabayashi-Ishioka, Y., et al. (2018). A Ycf2-FtsHi heteromeric AAA-ATPase complex is required for chloroplast protein import. Plant Cell *30*, 2677–2703.

Kikuchi, S., Bédard, J., Hirano, M., Hirabayashi, Y., Oishi, M., Imai, M., Takase, M., Ide, T., and Nakai, M. (2013). Uncovering the protein translocon at the chloroplast inner envelope membrane. Science *339*, 571–574.

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

Köhler, D., Dobritzsch, D., Hoehenwarter, W., Helm, S., Steiner, J.M., and Baginsky, S. (2015). Identification of protein N-termini in Cyanophora paradoxa cyanelles: transit peptide composition and sequence determinants for precursor maturation. Front. Plant Sci. *6*, 1–11.

Law, Y.S., Zhang, R., Guan, X., Cheng, S., Sun, F., Duncan, O., Murcha, M.W., Whelan, J., and Lim, B.L. (2015). Phosphorylation and dephosphorylation of the presequence of precursor MULTIPLE ORGANELLAR RNA EDITING FACTOR3 during import into mitochondria from arabidopsis. Plant Physiol. *169*, 1344–1355.

Lee, D.W., Lee, S., Lee, J., Woo, S., Razzak, M.A., Vitale, A., and Hwang, I. (2019). Molecular mechanism of the specificity of protein import into chloroplasts and mitochondria in plant cells. Mol. Plant *12*, 951–966.

Lee, J., O'Neill, R.C., Park, M.W., Gravel, M., and Braun, P.E. (2006). 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.

Lichtenthaler, H.K., Buschmann, C., Doll, M., Fietz, H., Bach, T., Kozel, U., Meier, D., and Rahmsdorf, U. (1981). 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.

Ling, Q., Broad, W., Trösch, R., Töpel, M., Demiral Sert, T., Lymperopoulos, P., Baldwin, A., and Jarvis, R.P. (2019). Ubiquitin-dependent chloroplast-associated protein degradation in plants. Science *363*, eaav4467.

Martin, W., and Herrmann, R.G. (1998). Gene transfer from organelles to the nucleus: how much, what happens, and why? Plant Physiol. *118*, 9–17.

Martin, W. (2010). Evolutionary origins of metabolic compartmentalization in eukaryotes. Philos. Trans. R. Soc. Lond. B Biol. Sci. *365*, 847–855.

May, T., and Soll, J. (2000). 14-3-3 proteins form a guidance complex with chloroplast precursor proteins in plants. Plant Cell *12*, 53–63.

Mergaert, P., Kikuchi, Y., Shigenobu, S., and Nowack, E.C.M. (2017). Metabolic integration of bacterial endosymbionts through antimicrobial peptides. Trends Microbiol. *25*, 703–712.

Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K.K., Haas, F.B., Vanderstraeten, L., Becker, D., Lang, D., et al. (2018). The Chara genome: secondary complexity and implications for plant terrestrialization. Cell *174*, 448–464.

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

O'Neil, P.K., Richardson, L.G.L., Paila, Y.D., Piszczek, G., Chakravarthy, S., Noinaj, N., and Schnell, D. (2017). The POTRA domains of Toc75 exhibit chaperone-like function to facilitate import into chloroplasts. Proc. Natl. Acad. Sci. U S A *114*, E4868–E4876.

Paila, Y.D., Richardson, L.G.L., and Schnell, D.J. (2015). New insights into the mechanism of chloroplast protein import and its integration with protein quality control, organelle biogenesis and development. J. Mol. Biol. *427*, 1038–1060.

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

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

Pfanzagl, B., Zenker, A., Pittenauer, E., Allmaier, G., Martinez-Torrecuadrada, J., Schmid, E.R., De Pedro, M.A., and Löffelhardt, W. (1996). Primary structure of cyanelle peptidoglycan of Cyanophora paradoxa: a prokaryotic cell wall as part of an organelle envelope. J. Bacteriol. *178*, 332–339.

Robert, V., Volokhina, E.B., Senf, F., Bos, M.P., Van Gelder, P., and Tommassen, J. (2006). Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. PLoS Biol. *4*, 1984–1995.

Rodríguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Löffelhardt, W., Bohnert, H.J., Philippe, H., and Lang, B.F. (2005). Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr. Biol. *15*, 1325–1330.

Sato, R., Ito, H., and Tanaka, A. (2015). Chlorophyll b degradation by chlorophyll b reductase under high-light conditions. Photosynth. Res. *126*, 249–259.

Schleiff, E., and Becker, T. (2011). Common ground for protein translocation: access control for mitochondria and chloroplasts. Nat. Rev. Mol. Cell Biol. *12*, 48–59.

Schnell, D.J., Kessler, F., and Blobel, G. (1994). Isolation of components of the chloroplast protein import machinery. Science *266*, 1007–1012.

Shi, L.X., and Theg, S.M. (2013). The chloroplast protein import system: from algae to trees. Biochim. Biophys. Acta *1833*, 314–331.

Singer, A., Poschmann, G., Mühlich, C., Valadez-Cano, C., Hänsch, S., Hüren, V., Rensing, S.A., Stühler, K., and Nowack, E.C.M. (2017). Massive protein import into the early-evolutionary-stage photosynthetic organelle of the amoeba Paulinella chromatophora. Curr. Biol. *27*, 2763–2773.

Sommer, M.S., Daum, B., Gross, L.E., Weis, B.L.M., Mirus, O., Abram, L., Maier, U.G., Kühlbrandt, W., and Schleiff, E. (2011). Chloroplast Omp85 proteins change orientation during evolution. Proc. Natl. Acad. Sci. U S A *108*, 13841–13846.

Steiner, J.M., Yusa, F., Pompe, J.A., and Löffelhardt, W. (2005). Homologous protein import machineries in chloroplasts and cyanelles. Plant J. *44*, 646–652.

Sun, F., Carrie, C., Law, S., Murcha, M.W., Zhang, R., Law, Y.S., Suen, P.K., Whelan, J., and Lim, B.L. (2012). AtPAP2 is a tail-anchored protein in the outer membrane of chloroplasts and mitochondria. Plant Signal. Behav. *7*, 927–932.

Suzuki, E., and Suzuki, R. (2013). Variation of storage polysaccharides in phototrophic microorganisms. J. Appl. Glycosci. *60*, 21–27.

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

Timmis, J.N., Ayliff, M.A., Huang, C.Y., and Martin, W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. *5*, 123–135.

Tomitani, A., Okada, K., Miyashita, H., Matthijs, H.C.P., Ohno, T., and Tanaka, A. (1999). Chlorophyll b and phycobilins in the common ancestor of cyanobacteria and chloroplasts. Nature *400*, 159–162.

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

Töpel, M., Ling, Q., and Jarvis, P. (2012). Neofunctionalization within the Omp85 protein superfamily during chloroplast evolution. Plant Signal. Behav. *7*, 161–164.

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

de Vries, J., Stanton, A., Archibald, J.M., and Gould, S.B. (2016). Streptophyte terrestrialization in light of plastid evolution. Trends Plant Sci. *21*, 467–476.

de Vries, J., Archibald, J.M., and Gould, S.B. (2017). The carboxy terminus of YCF1 contains a motif conserved throughout >500 million years of streptophyte evolution. Genome Biol. Evol. *9*, 473–479.

Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl. Acad. Sci. U S A *111*, E4859–E4868.

Wiedemann, N., Kozjak, V., Chacinska, A., Schönfisch, B., Rospert, S., Ryan, M.T., Pfanner, N., and Meisinger, C. (2003). Machinery for protein sorting and assembly in the mitochondrial outer membrane. Nature *424*, 565–571.

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

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

Wu, G.Z., Meyer, E.H., Richter, A.S., Schuster, M., Ling, Q., Schöttler, M.A., Walther, D., Zoschke, R., Grimm, B., Jarvis, R.P., and Böck, R. (2019). Control of retrograde signalling by protein import and cytosolic folding stress. Nat. Plants *5*, 525–538.

Wunder, T., Martin, R., Löffelhardt, W., Schleiff, E., and Steiner, J.M. (2007). The invariant phenylalanine of precursor proteins discloses the importance of Omp85 for protein translocation into cyanelles. BMC Evol. Biol. *7*, 236.

Yoshioka-Nishimura, M., and Yamamoto, Y. (2014). Quality control of Photosystem II: the molecular basis for the action of FtsH protease and the dynamics of the thylakoid membranes. J. Photochem. Photobiol. B Biol. *137*, 100–106.

Zhu, J.K. (2016). Abiotic stress signaling and responses in plants. Cell *167*, 313–324.

Zimorski, V., Ku, C., Martin, W.F., and Gould, S.B. (2014). Endosymbiotic theory for organelle origins. Curr. Opin. Microbiol. *22*, 38–48.

iScience, Volume *23*

Supplemental Information

Major Changes in Plastid Protein Import and the Origin of the Chloroplastida

Michael Knopp, Sriram G. Garg, Maria Handrich, and Sven B. Gould

Transparent Methods

Culturing

C. reinhardtii (SCCAP K-1017) and *C. paradoxa* (SCCAP K-06262) were grown in NF2 medium and *P. purpureum* (SCCAP K-0515) cells were grown in MV10 medium (see http://www.sccap.dk/media/freshwater/4.asp and http://www.sccap.dk/media/marine/3.asp of the SCCAP for exact recipe and preparation), all three in aerated flasks at 20°C and illuminated with 50 µmol photons m^{-2} s^{-1} 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 600 µmol photons m^{-2} s^{-1}. RNA was sequenced and assembled exactly as described previously (Gould et al., 2019), 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 photons m^{-2} s^{-1}) and modulated blue light as saturation pulse (> 8000µmol photons 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-phase high pressure liquid chromatography (HPLC) with ultraviolet/visible spectroscopy detection (Hitachi/Merck). Pigment concentrations were determined using external pigment standards isolated from spinach thylakoids (Färber et al., 1997).

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 (Schindelin et al., 2012). 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 (Grabherr et al., 2011) (r2013-02-25), edgeR (Robinson et al., 2009) 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). For an overall comparison of all three conditions, for each transcript the fold changes were calculated relative to the average over all conditions. 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, although in cases where the hit did not provide enough information (hypothetical proteins, predicted proteins) the next best non-hypothetical hit was selected, accepting also E-values up to 10^{-3}. Additionally, InterProScan 5.39-77.0 (Quevillon et al., 2005) (Linux Standalone version, Pfam analysis) was used to retrieve Pfam, InterPro and GO Identifier.

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 (O'Leary et al., 2016) or GenBank (Clark et al., 2016) database via their respective gene identifiers (Suppl. table 1). Additionally, 28 genomes from Chlorophytes, Rhodophytes (Rossoni et al., 2019) and one Glaucophyte were downloaded either from the Refseq, GenBank or the JGI Genome Portal (Grigoriev et al., 2012) (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) (Altschul et al., 1997, Altschul et al., 1990). 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" (Katoh et al., 2002). 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 (Stamatakis, 2014, version 8.2.8) using the substitution model 'PROTCATWAGF' (WAG substitution Matrix and empirical base frequencies; model determined by the IQ-Tree 1.6.7 Modelfinder (Kalyaanamoorthy et al., 2017)) and 100 non-parametric bootstraps. An additional tree was constructed using the new RAxML-NG with the model LG+F+R5 and 1000 bootstraps (Kozlov et al., 2019) (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. All proteins were blasted (diamond blastp, identity cutoff: 25%, e-value 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* (Garg and Gould, 2016) 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 (Terashima et al., 2010, Facchinelli et al., 2013). 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 Gruber et al., 2007) and plotted using Seq2Logo (Thomsen and Nielsen, 2012).

References

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

Altschul, S. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

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

Eilers, P. H. C., and Peeters, J. C. H. (1988). A model for the relationship between light intensity and the rate of photosynthesis in phytoplankton. Ecol. Modell. 42, 199–215.

Facchinelli, F., Pribil, M., Oster, U., Ebert, N. J., Bhattacharya, D., Leister, D., and Weber, A. P. M. (2013). Proteomic analysis of the Cyanophora paradoxa muroplast provides clues on early events in plastid endosymbiosis. Planta 237, 637–651.

Färber, A., Young, A. J., Ruban, A. V., Horton, P., and Jahns, P. (1997). 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.

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

Gould, S. B., Garg, S. G., Handrich, M., Nelson-Sathi, S., Gruenheit, N., Tielens, A. G. M., and Martin, W. F. (2019). 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.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652.
Grigoriev, I. V, Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A., Minovitsky, S., Nikitin, R., Ohm, R. A., et al. (2012). The Genome Portal of the Department of Energy Joint Genome Institute. Nucleic Acids Res. 40, 26–32.

Gruber, A., Vugrinec, S., Hempel, F., Gould, S. B., Maier, U. G., and Kroth, P. G. (2007). Protein targeting into complex diatom plastids: Functional characterisation of a specific targeting motif. Plant Mol. Biol. 64, 519–530.

Inoue, K., and Potter, D. (2004). 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.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589.

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

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A., and Wren, J. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35, 4453–4455.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44, D733–D745.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. (2005). InterProScan: protein domains identifier. Nucleic Acids Res. 33, W116–W120.
Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

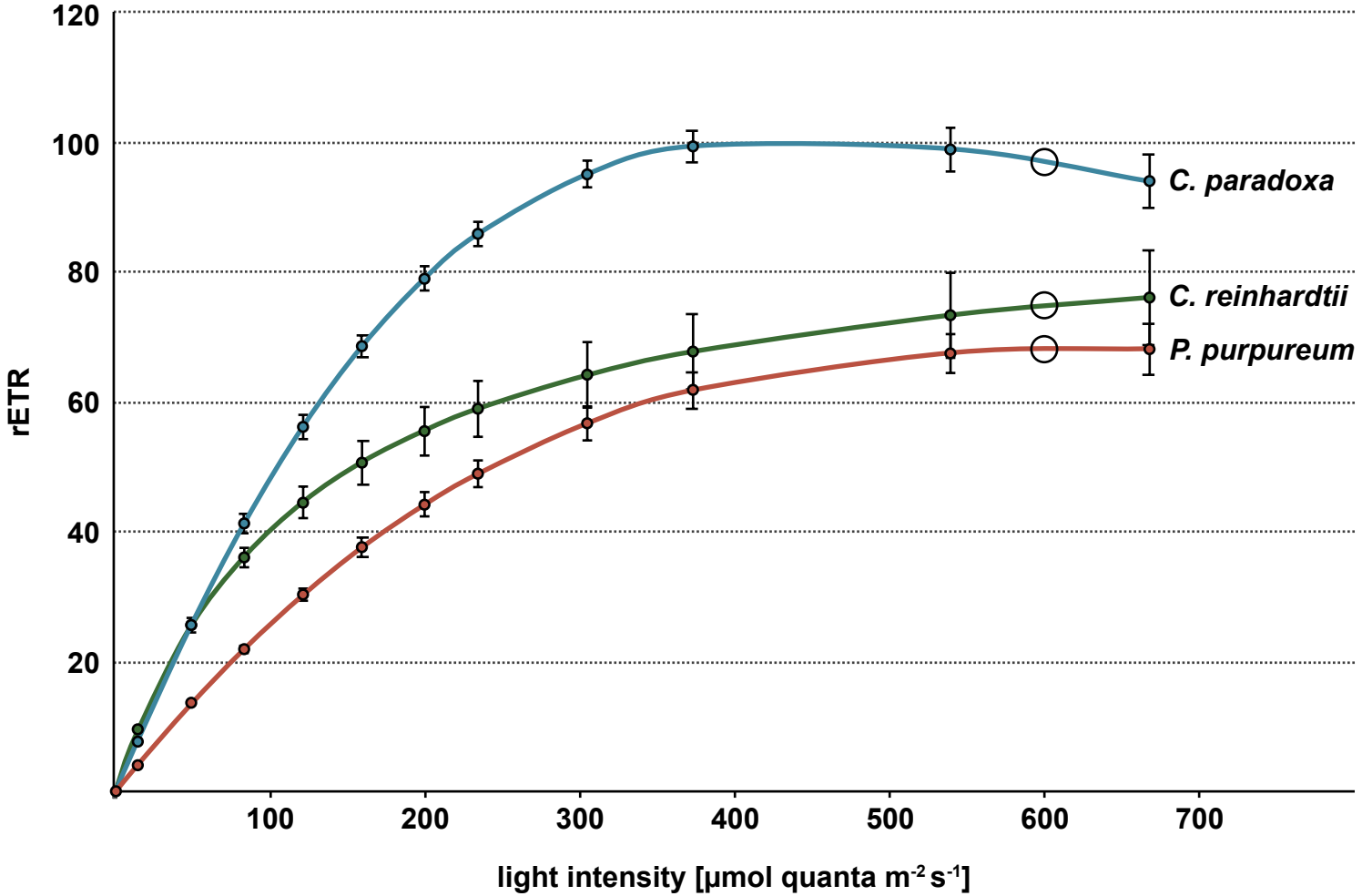
Rossoni, A. W., Price, D. C., Seger, M., Lyska, D., Lammers, P., Bhattacharya, D., and Weber, A. P. M. (2019). The genomes of polyextremophilic Cyanidiales contain 1% horizontally transferred genes with diverse adaptive functions. Elife 8, e45017.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: An open-source platform for biological-image analysis. Nat. Methods 9, 676–682.

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

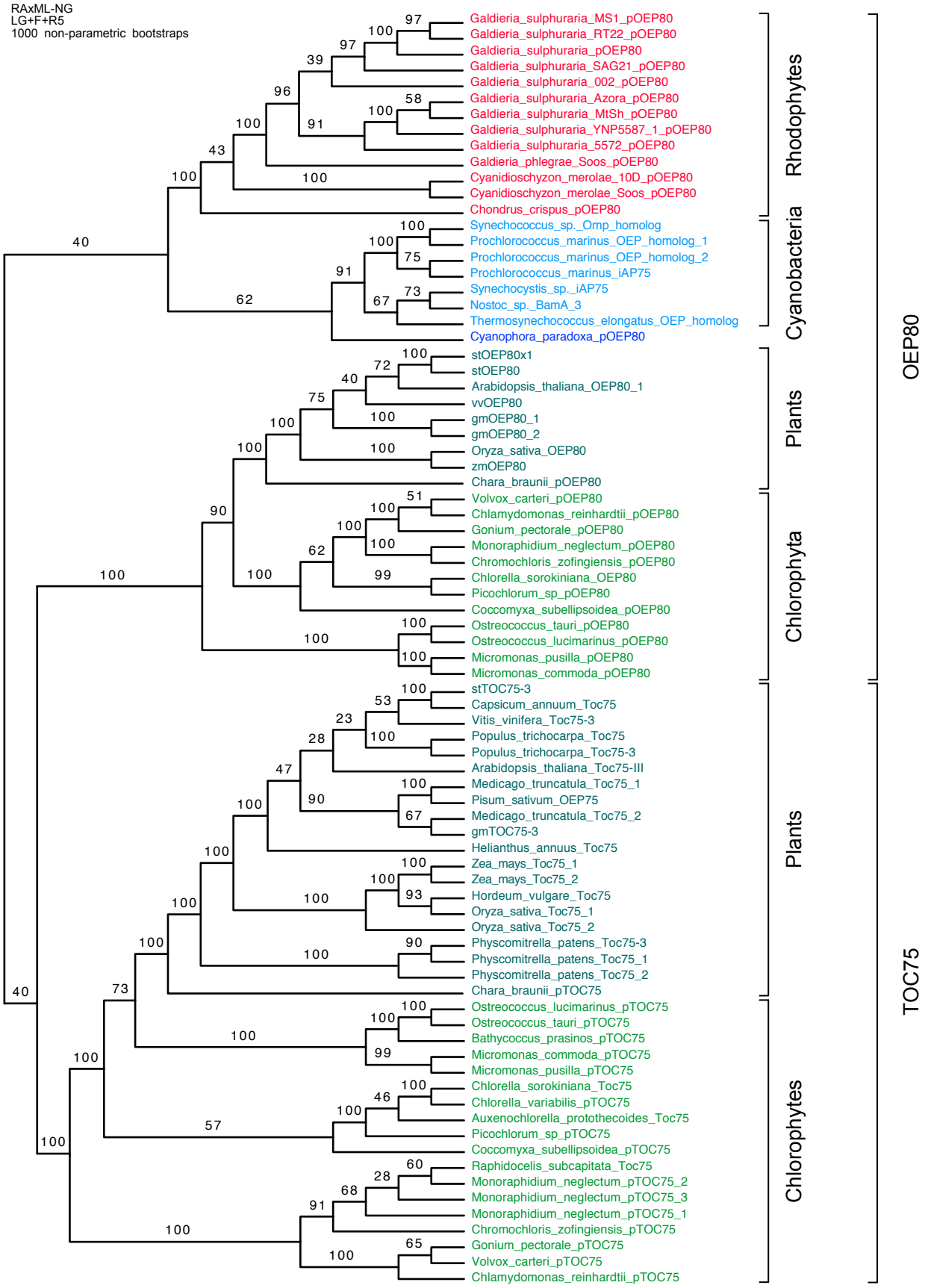
Terashima, M., Specht, M., Naumann, B., and Hippler, M. (2010). Characterizing the anaerobic response of Chlamydomonas reinhardtii by quantitative proteomics. Mol. Cell. Proteomics 9, 1514–1532.

Thomsen, M. C. F., and Nielsen, M. (2012). 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.



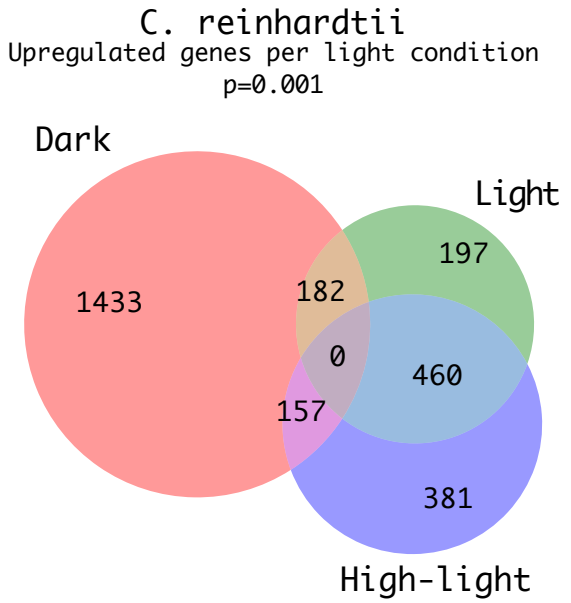
Suppl. Fig. 1: Related to Figure 2. Rapid light curves (RLCs) of the three primary algae species.

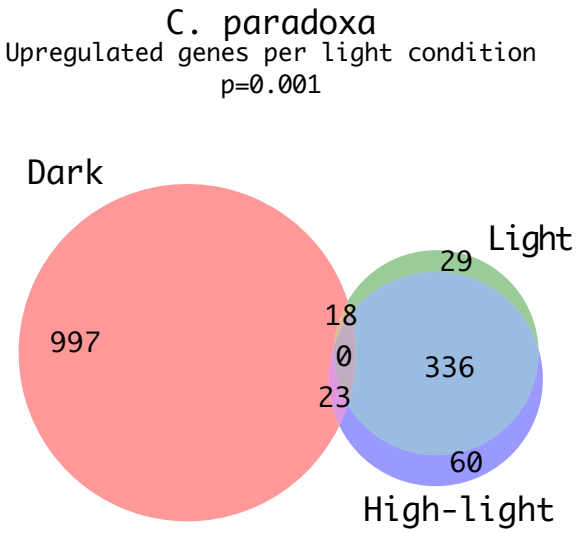
Relative electron transport rate (rETR) was determined using the FluorCam FC 800MF. Algae cultures were cultivated at 20°C with an illumination intensity of 50µE under a 12/12h day-night cycle. After 5 min of dark adaptation photosynthetic activity was measured with increasing light intensities (13, 48, 122, 160, 200, 235, 305, 375, 542, 670 µmol quanta m^{-2} s^{-1}). At 600 µE (indicated with circles) a saturation of the photosynthetic apparatus occurred in all three lineages analyzed.

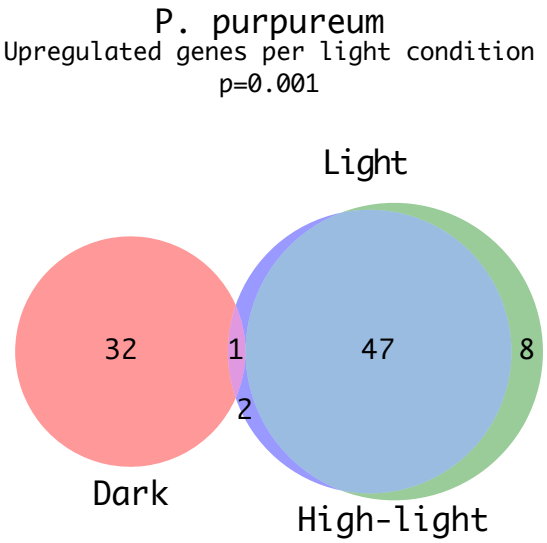


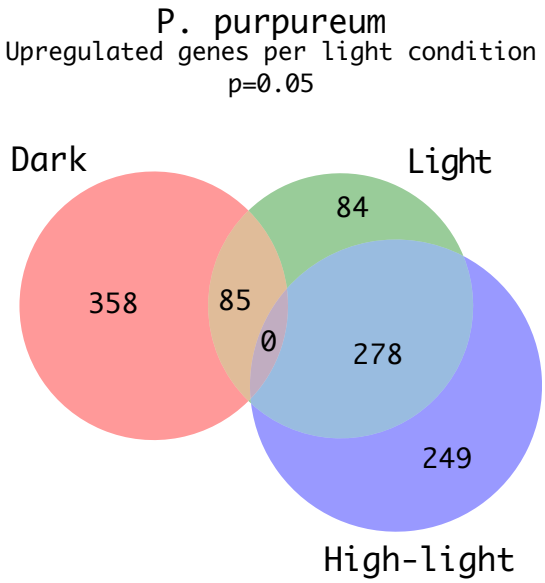
Suppl. Fig. 2: Related to Figure 4. Alternative phylogenetic analysis of Oep80 and Toc75 homologs

Phylogeny reconstruction via RaXML-NG (LG+R5+F) with 1000 bootstraps. The tree was rooted on the split between the monophyletic cyanobacteria and the eukaryotic sequences. Although the branching pattern differs from the tree reconstructed with RAxML 8 and the PROTCATWAGF Model, all TOC75 and OEP80 sequences from *Chloroplastida* still form a monophyletic clade. Like in the analysis underlying Fig. 4, TOC75 and OEP80 proteins of plants each form monophyletic clades.









Suppl. Fig. 3: Related to Figure 3. Venn diagrams of upregulated genes during each tested light condition.
Upregulation of transcripts between all conditions (Fig. 3). While *C. reinhardtii* upregulates distinct sets of genes during light and high-light conditions, *C. paradoxa* and *P. purpureum* upregulate predominantly the same genes during light and high-light conditions (p=0.001) illustrating C. reinhardtii's superior adaption to high light stress.

Suppl. Table 1: Related to Figure 4.

Tabular listing of all sequences used to construct phylogenetic trees (Fig. 4 and Suppl. Fig. 2) including the sequence header used in this study, accession number and taxonomic group of the source organism. All sequences that were found via BLAST searches are labelled as “Extended set”. Additionally, the genome sources of these sequences are listed including their source database, URL, download date and their respective assembly level.

Sequence header in this study	Dataset	Group	original Accession
Bathycoccus_prasinos_pTOC75	Extended set	Chlorophytes	XP_007509108.1
Chlamydomonas_reinhardtii_pOEP80	Extended set	Chlorophytes	XP_001695043.1
Chlamydomonas_reinhardtii_pTOC75	Extended set	Chlorophytes	XP_001703281.1
Chlorella_variabilis_pTOC75	Extended set	Chlorophytes	XP_005843553.1
Chromochloris_zofingiensis_pOEP80	Extended set	Chlorophytes	jgi|Chrzof1|9681|Cz04g11310.t1
Chromochloris_zofingiensis_pTOC75	Extended set	Chlorophytes	jgi|Chrzof1|819|Cz01g30030.t1
Coccomyxa_subellipsoidea_pOEP80	Extended set	Chlorophytes	XP_005648812.1
Coccomyxa_subellipsoidea_pTOC75	Extended set	Chlorophytes	XP_005649815.1
Gonium_pectorale_pOEP80	Extended set	Chlorophytes	jgi|Gonpec1|12926|rna3050
Gonium_pectorale_pTOC75	Extended set	Chlorophytes	jgi|Gonpec1|489|rna539
Micromonas_commoda_pOEP80	Extended set	Chlorophytes	XP_002509092.1
Micromonas_commoda_pTOC75	Extended set	Chlorophytes	XP_002504526.1
Micromonas_pusilla_pOEP80	Extended set	Chlorophytes	XP_003061676.1
Micromonas_pusilla_pTOC75	Extended set	Chlorophytes	XP_003059853.1
Monoraphidium_neglectum_pOEP80	Extended set	Chlorophytes	XP_013891338.1
Monoraphidium_neglectum_pTOC75_2	Extended set	Chlorophytes	XP_013895618.1
Ostreococcus_lucimarinus_pOEP80	Extended set	Chlorophytes	XP_001420427.1
Ostreococcus_lucimarinus_pTOC75	Extended set	Chlorophytes	XP_001416377.1
Ostreococcus_tauri_pOEP80	Extended set	Chlorophytes	XP_003081873.1
Ostreococcus_tauri_pTOC75	Extended set	Chlorophytes	XP_003074813.1
Picochlorum_sp_pOEP80	Extended set	Chlorophytes	jgi|Picsp_1|6034|NSC_03388-R1_protein
Picochlorum_sp_pTOC75	Extended set	Chlorophytes	jgi|Picsp_1|3926|NSC_01438-R1_chloroplast
Volvox_carteri_pOEP80	Extended set	Chlorophytes	XP_002948533.1
Volvox_carteri_pTOC75	Extended set	Chlorophytes	XP_002946938.1
Cyanophora_paradoxa_pOEP80	Extended set	Glaucophytes	tig00000900_g5389.t1
Chondrus_crispus_pOEP80	Extended set	Rhodophytes	XP_005711725.1
Cyanidioschyzon_merolae_10D_pOEP80	Extended set	Rhodophytes	XP_005534927.1
Cyanidioschyzon_merolae_Soos_pOEP80	Extended set	Rhodophytes	G3979.1
Galdieria_phlegrae_Soos_pOEP80	Extended set	Rhodophytes	G2108.1
Galdieria_sulphuraria_002_pOEP80	Extended set	Rhodophytes	G849.1
Galdieria_sulphuraria_074W_pOEP80	Extended set	Rhodophytes	XP_005703263.1
Galdieria_sulphuraria_5572_pOEP80	Extended set	Rhodophytes	G1359.1
Galdieria_sulphuraria_Azora_pOEP80	Extended set	Rhodophytes	G2556.1
Galdieria_sulphuraria_MS1_pOEP80	Extended set	Rhodophytes	G2257.1
Galdieria_sulphuraria_MtSh_pOEP80	Extended set	Rhodophytes	G1270.1
Galdieria_sulphuraria_RT22_pOEP80	Extended set	Rhodophytes	G2280.1
Galdieria_sulphuraria_SAG21_pOEP80	Extended set	Rhodophytes	G2095.1
Galdieria_sulphuraria_YNP5587_1_pOEP80	Extended set	Rhodophytes	G925.1
Auxenochlorella_protothecoides_Toc75	Initial sequence set	Chlorophytes	KFM25192.1
Chlorella_sorokiniana_OEP80	Initial sequence set	Chlorophytes	PRW60585.1
Chlorella_sorokiniana_Toc75	Initial sequence set	Chlorophytes	PRW61091.1
Raphidocelis_subcapitata_Toc75	Initial sequence set	Chlorophytes	GBF94877.1
Nostoc_sp._BamA_3	Initial sequence set	Cyanobacteria	BAB73968.1
Prochlorococcus_marinus_iAP75	Initial sequence set	Cyanobacteria	AAQ00463.1
Prochlorococcus_marinus_OEP80_1	Initial sequence set	Cyanobacteria	CAE19797.1
Prochlorococcus_marinus_OEP80_2	Initial sequence set	Cyanobacteria	CAE21588.1
Synechococcus_sp._Omp_homolog	Initial sequence set	Cyanobacteria	CAE07070.1
Synechocystis_sp._iAP75	Initial sequence set	Cyanobacteria	BAA17512.1
Thermosynechococcus_elongatus_OEP_homolog	Initial sequence set	Cyanobacteria	BAC09341.1
Arabidopsis_thaliana_OEP80_1	Initial sequence set	Streptophytes	NP_568378.1
Arabidopsis_thaliana_Toc75-III	Initial sequence set	Streptophytes	Q9STE8.1
Capsicum_annuum_Toc75	Initial sequence set	Streptophytes	PHT80225.1
Chara_braunii_pOEP80	Initial sequence set	Streptophytes	GBG89700.1
Chara_braunii_pTOC75	Initial sequence set	Streptophytes	GBG61930.1
Glycin_max_OEP80_1	Initial sequence set	Streptophytes	XP_003542049.2
Glycin_max_OEP80_2	Initial sequence set	Streptophytes	XP_003547118.1
Glycin_max_TOC75-3	Initial sequence set	Streptophytes	XP_003547008.1
Helianthus_annuus_Toc75	Initial sequence set	Streptophytes	OTG28259.1
Hordeum_vulgare_Toc75	Initial sequence set	Streptophytes	BAJ97575.1
Medicago_truncatula_Toc75_1	Initial sequence set	Streptophytes	XP_003606719.1
Medicago_truncatula_Toc75_2	Initial sequence set	Streptophytes	XP_003597400.3
Oryza_sativa_OEP80	Initial sequence set	Streptophytes	XP_015627644.1
Oryza_sativa_Toc75_1	Initial sequence set	Streptophytes	XP_015630560.1
Oryza_sativa_Toc75_2	Initial sequence set	Streptophytes	XP_025877151.1
Physcomitrella_patens_Toc75_1	Initial sequence set	Streptophytes	XP_024357726.1
Physcomitrella_patens_Toc75_2	Initial sequence set	Streptophytes	XP_024357726.1
Physcomitrella_patens_Toc75-3	Initial sequence set	Streptophytes	XP_024357726.1
Pisum_sativum_OEP75	Initial sequence set	Streptophytes	CAA58720.1
Populus_trichocarpa_Toc75-3	Initial sequence set	Streptophytes	XP_002303729.2
Solanum_tuberosum_OEP80_1	Initial sequence set	Streptophytes	XP_006354253.1
Solanum_tuberosum_OEP80_2	Initial sequence set	Streptophytes	XP_006351245.1
Solanum_tuberosum_TOC75-3	Initial sequence set	Streptophytes	XP_006350787.1
Vitis_vinifera_OEP80	Initial sequence set	Streptophytes	XP_002285507.2
Vitis_vinifera_Toc75-3	Initial sequence set	Streptophytes	XP_002280661.1
Zea_mays_OEP80	Initial sequence set	Streptophytes	XP_008645435.1
Zea_mays_Toc75_1	Initial sequence set	Streptophytes	PWZ07718.1
Zea_mays_Toc75_2	Initial sequence set	Streptophytes	PWZ53262.1

Organism	Source	Genome URL	Downloaded	Assembly level
Ostreococcus tauri	Refseq	ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/Ostreococcus_tauri/latest_assembly_versions/GCF_000214015.3_version_140606	20.03.19	chromosome
Chlamydomonas reinhardtii	Refseq	ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/Chlamydomonas_reinhardtii/latest_assembly_versions/GCF_000002595.1_v3.0	20.03.19	scaffold
Micromonas commoda	Refseq	ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/Micromonas_commoda/latest_assembly_versions/GCF_000090985.2_ASM9098v2	20.03.19	complete genome
Micromonas pusilla	Refseq	ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/Micromonas_pusilla/latest_assembly_versions/GCF_000151265.2_Micromonas_pusilla_CCMP1545_v2.0	20.03.19	scaffold
Ostreococcus lucimarinus	Refseq	ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/Ostreococcus_sp._lucimarinus_/latest_assembly_versions/GCF_000092065.1_ASM9206v1	20.03.19	complete genome
Chlorella variabilis	Refseq	ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/Chlorella_variabilis/latest_assembly_versions/GCF_000147415.1_v_1.0	20.03.19	scaffold
Volvox carteri	Refseq	ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/Volvox_carteri/latest_assembly_versions/GCF_000143455.1_v1.0	20.03.19	scaffold
Bathycoccus prasinos	Refseq	ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/Bathycoccus_prasinos/latest_assembly_versions/GCF_002220235.1_ASM222023v1	20.03.19	chromosome
Coccomyxa subellipsoidea	Refseq	ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/Coccomyxa_subellipsoidea/latest_assembly_versions/GCF_000258705.1_Coccomyxa_subellipsoidae_v2.0	20.03.19	contig
Monoraphidium neglectum	Refseq	ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/Monoraphidium_neglectum/latest_assembly_versions/GCF_000611645.1_mono_v1	20.03.19	scaffold
Gonium pectorale	JGI	https://genome.jgi.doe.gov/portal/pages/dynamicOrganismDownload.jsf?organism=Gonpec1	20.03.19	–
Chromochloris zofingiensis	JGI	https://genome.jgi.doe.gov/portal/pages/dynamicOrganismDownload.jsf?organism=Chrzof1	20.03.19	–
Picochlorum sp.	JGI	https://genome.jgi.doe.gov/portal/pages/dynamicOrganismDownload.jsf?organism=Picsp_1	20.03.19	–
Cyanophora paradoxa	Institute of Plant Biochemistry, HHU, Germany	http://cyanophora.rutgers.edu/cyanophora/Cyanophora_paradoxa_MAKER_gene_predictions-022111-aa.fasta	20.03.19	contig
Galdieria sulphuraria 074W	Refseq	ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/Galdieria_sulphuraria/latest_assembly_versions/GCF_000341285.1_ASM34128v1	20.03.19	scaffold
Galdieria phlegrae Soos	Institute of Plant Biochemistry, HHU, Germany	http://porphyra.rutgers.edu/Rossoni_et_al_2019.zip	–	–
Galdieria sulphuraria 002	Institute of Plant Biochemistry, HHU, Germany	http://porphyra.rutgers.edu/Rossoni_et_al_2019.zip	–	–
Galdieria sulphuraria 5572	Institute of Plant Biochemistry, HHU, Germany	http://porphyra.rutgers.edu/Rossoni_et_al_2019.zip	–	–
Galdieria sulphuraria Azora	Institute of Plant Biochemistry, HHU, Germany	http://porphyra.rutgers.edu/Rossoni_et_al_2019.zip	–	–
Galdieria sulphurariaMS1	Institute of Plant Biochemistry, HHU, Germany	http://porphyra.rutgers.edu/Rossoni_et_al_2019.zip	–	–
Galdieria sulphurariaMtSh	Institute of Plant Biochemistry, HHU, Germany	http://porphyra.rutgers.edu/Rossoni_et_al_2019.zip	–	–
Galdieria sulphuraria RT22	Institute of Plant Biochemistry, HHU, Germany	http://porphyra.rutgers.edu/Rossoni_et_al_2019.zip	–	–
Galdieria sulphuraria SAG21	Institute of Plant Biochemistry, HHU, Germany	http://porphyra.rutgers.edu/Rossoni_et_al_2019.zip	–	–
Galdieria sulphuraria YNP5587_1	Institute of Plant Biochemistry, HHU, Germany	http://porphyra.rutgers.edu/Rossoni_et_al_2019.zip	–	–
Cyanidioschyzon merolae Soos	Institute of Plant Biochemistry, HHU, Germany	http://porphyra.rutgers.edu/Rossoni_et_al_2019.zip	–	–
Cyanidioschyzon merolae 10D	Refseq	ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/Cyanidioschyzon_merolae/latest_assembly_versions/GCF_000091205.1_ASM9120v1	20.03.19	complete genome
Chondrus crispus	Refseq	ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/Chondrus_crispus/latest_assembly_versions/GCF_000350225.1_ASM35022v2	20.03.19	scaffold

Suppl. Table 2: Provided as Excel-Table

Suppl. Table 3: Provided as Excel-Table

Suppl. Table 4: Provided as Excel-Table

Suppl. Table 5: Related to Suppl. Tables 3 and 4.

List of all 112 Refseq plant and algal genomes that were used to annotate the transcripts of *C. reinhardtii*, *C. paradoxa* and *P. purpureum*.

Organism name	Assembly accession	TaxonomyID	Assembly level	Genome release date	Downloaded
Citrus sinensis	GCF_000317415.1	2711	Chromosome	12.12.12	28.06.19
Physcomitrella patens	GCF_000002425.4	3218	Chromosome	24.01.18	28.06.19
Papaver somniferum	GCF_003573695.1	3469	Chromosome	18.09.18	28.06.19
Gossypium hirsutum	GCF_000987745.1	3635	Chromosome	05.05.15	28.06.19
Theobroma cacao	GCF_000208745.1	3641	Chromosome	09.07.16	28.06.19
Cucumis sativus	GCF_000004075.2	3659	Chromosome	27.10.14	28.06.19
Cucurbita pepo subsp. pepo	GCF_002806865.1	3664	Chromosome	07.12.17	28.06.19
Populus trichocarpa	GCF_000002775.4	3694	Chromosome	24.01.18	28.06.19
Arabidopsis thaliana	GCF_000001735.4	3702	Chromosome	15.03.18	28.06.19
Brassica napus	GCF_000686985.2	3708	Chromosome	15.09.17	28.06.19
Brassica rapa	GCF_000309985.1	3711	Chromosome	07.11.12	28.06.19
Brassica oleracea var. oleracea	GCF_000695525.1	109376	Chromosome	27.05.14	28.06.19
Malus domestica	GCF_002114115.1	3750	Chromosome	03.05.17	28.06.19
Prunus persica	GCF_000346465.2	3760	Chromosome	02.02.17	28.06.19
Arachis hypogaea	GCF_003086295.2	3818	Chromosome	30.05.18	28.06.19
Cajanus cajan	GCF_000340665.1	3821	Chromosome	29.09.16	28.06.19
Cicer arietinum	GCF_000331145.1	3827	Chromosome	16.01.13	28.06.19
Glycine max	GCF_000004515.5	3847	Chromosome	24.07.18	28.06.19
Glycine soja	GCF_004193775.1	3848	Chromosome	21.02.19	28.06.19
Lupinus angustifolius	GCF_001865875.1	3871	Chromosome	16.11.16	28.06.19
Medicago truncatula	GCF_000219495.3	3880	Chromosome	18.06.14	28.06.19
Phaseolus vulgaris	GCF_000499845.1	3885	Chromosome	29.11.13	28.06.19
Vigna angularis	GCF_001190045.1	3914	Chromosome	31.07.15	28.06.19
Vigna unguiculata	GCF_004118075.1	3917	Chromosome	30.01.19	28.06.19
Manihot esculenta	GCF_001659605.1	3983	Chromosome	10.06.16	28.06.19
Daucus carota subsp. sativus	GCF_001625215.1	79200	Chromosome	06.05.16	28.06.19
Capsicum annuum	GCF_000710875.1	4072	Chromosome	11.03.15	28.06.19
Solanum lycopersicum	GCF_000188115.4	4081	Chromosome	18.04.18	28.06.19
Olea europaea var. sylvestris	GCF_002742605.1	158386	Chromosome	03.11.17	28.06.19
Sesamum indicum	GCF_000512975.1	4182	Chromosome	06.01.14	28.06.19
Helianthus annuus	GCF_002127325.1	4232	Chromosome	12.05.17	28.06.19
Cynara cardunculus var. scolymus	GCF_001531365.1	59895	Chromosome	17.04.18	28.06.19
Oryza sativa Japonica Group	GCF_001433935.1	39947	Chromosome	10.10.15	28.06.19
Oryza brachyantha	GCF_000231095.1	4533	Chromosome	19.01.12	28.06.19
Setaria italica	GCF_000263155.2	4555	Chromosome	30.10.15	28.06.19
Sorghum bicolor	GCF_000003195.3	4558	Chromosome	07.04.17	28.06.19
Zea mays	GCF_000005005.2	4577	Chromosome	07.02.17	28.06.19
Ananas comosus	GCF_001540865.1	4615	Chromosome	14.03.16	28.06.19
Musa acuminata subsp. malaccensis	GCF_000313855.2	214687	Chromosome	14.11.12	28.06.19
Asparagus officinalis	GCF_001876935.1	4686	Chromosome	06.02.17	28.06.19
Coffea arabica	GCF_003713225.1	13443	Chromosome	08.11.18	28.06.19
Brachypodium distachyon	GCF_000005505.3	15368	Chromosome	24.01.18	28.06.19
Solanum pennellii	GCF_001406875.1	28526	Chromosome	11.07.14	28.06.19
Gossypium arboreum	GCF_000612285.1	29729	Chromosome	28.09.15	28.06.19
Gossypium raimondii	GCF_000327365.1	29730	Chromosome	20.12.12	28.06.19
Vitis vinifera	GCF_000003745.3	29760	Chromosome	07.12.09	28.06.19
Coffea eugenioides	GCF_003713205.1	49369	Chromosome	08.11.18	28.06.19
Nicotiana attenuata	GCF_001879085.1	49451	Chromosome	15.11.16	28.06.19
Elaeis guineensis	GCF_000442705.1	51953	Chromosome	13.08.13	28.06.19
Fragaria vesca subsp. vesca	GCF_000184155.1	101020	Chromosome	24.02.11	28.06.19
Ostreococcus tauri	GCF_000214015.3	70448	Chromosome	02.10.14	28.06.19
Rosa chinensis	GCF_002994745.1	74649	Chromosome	15.01.19	28.06.19
Camelina sativa	GCF_000633955.1	90675	Chromosome	24.04.14	28.06.19
Prunus mume	GCF_000346735.1	102107	Chromosome	28.02.14	28.06.19
Arachis duranensis	GCF_000817695.2	130453	Chromosome	25.04.17	28.06.19
Arachis ipaensis	GCF_000816755.2	130454	Chromosome	25.04.17	28.06.19
Vigna radiata var. radiata	GCF_000741045.1	3916	Chromosome	28.10.15	28.06.19
Beta vulgaris subsp. vulgaris	GCF_000511025.2	3555	Chromosome	07.07.15	28.06.19
Panicum hallii	GCF_002211085.1	206008	Chromosome	01.05.18	28.06.19
Ziziphus jujuba	GCF_000826755.1	326968	Chromosome	06.02.15	28.06.19
Cyanidioschyzon merolae strain 10D	GCF_000091205.1	280699	Complete Genor	11.07.07	28.06.19
Micromonas commoda	GCF_000090985.2	296587	Complete Genor	10.04.09	28.06.19
Aegilops tauschii subsp. tauschii	GCF_001957025.1	169297	Contig	19.01.17	28.06.19
Prosopis alba	GCF_004799145.1	207710	Contig	17.04.19	28.06.19
Coccomyxa subellipsoidea C-169	GCF_000258705.1	574566	Contig	13.04.12	28.06.19
Chondrus crispus	GCF_000350225.1	2769	Scaffold	22.05.13	28.06.19
Chlamydomonas reinhardtii	GCF_000002595.1	3055	Scaffold	15.10.07	28.06.19
Volvoxcarteri f. nagariensis	GCF_000143455.1	3068	Scaffold	08.07.10	28.06.19
Spinacia oleracea	GCF_002007265.1	3562	Scaffold	27.02.17	28.06.19
Carica papaya	GCF_000150535.2	3649	Scaffold	06.05.08	28.06.19
Cucumis melo	GCF_000313045.1	3656	Scaffold	05.10.12	28.06.19
Cucurbita maxima	GCF_002738345.1	3661	Scaffold	31.10.17	28.06.19
Cucurbita moschata	GCF_002738365.1	3662	Scaffold	31.10.17	28.06.19
Momordica charantia	GCF_001995035.1	3673	Scaffold	26.12.16	28.06.19
Raphanus sativus	GCF_000801105.1	3726	Scaffold	29.09.15	28.06.19
Abrus precatorius	GCF_003935025.1	3816	Scaffold	11.12.18	28.06.19
Hevea brasiliensis	GCF_001654055.1	3981	Scaffold	01.06.16	28.06.19
Ricinus communis	GCF_000151685.1	3988	Scaffold	07.07.11	28.06.19
Nicotiana sylvestris	GCF_000393655.1	4096	Scaffold	16.05.13	28.06.19
Nicotiana tabacum	GCF_000715135.1	4097	Scaffold	29.05.14	28.06.19
Nicotiana tomentosiformis	GCF_000390325.2	4098	Scaffold	16.05.13	28.06.19
Solanum tuberosum	GCF_000226075.1	4113	Scaffold	19.09.11	28.06.19
Erythranthe guttata	GCF_000504015.1	4155	Scaffold	02.04.14	28.06.19
Lactuca sativa	GCF_002870075.1	4236	Scaffold	09.01.18	28.06.19
Nelumbo nucifera	GCF_000365185.1	4432	Scaffold	08.08.13	28.06.19
Camellia sinensis	GCF_004153795.1	4442	Scaffold	11.02.19	28.06.19
Amborella trichopoda	GCF_000471905.2	13333	Scaffold	30.09.13	28.06.19
Tarenaya hassleriana	GCF_000463585.1	28532	Scaffold	05.09.13	28.06.19
Ipomoea nil	GCF_001879475.1	35883	Scaffold	01.09.16	28.06.19
Micromonas pusilla CCMP1545	GCF_000151265.2	564608	Scaffold	09.04.09	28.06.19
Prunus avium	GCF_002207925.1	42229	Scaffold	12.06.17	28.06.19
Phoenixdactylifera	GCF_000413155.1	42345	Scaffold	24.06.13	28.06.19
Juglans regia	GCF_001411555.1	51240	Scaffold	22.10.15	28.06.19
Quercus suber	GCF_002906115.1	58331	Scaffold	29.01.18	28.06.19
Arabidopsis lyrata subsp. lyrata	GCF_000004255.2	81972	Scaffold	26.11.16	28.06.19
Chenopodium quinoa	GCF_001683475.1	63459	Scaffold	19.05.17	28.06.19
Durio zibethinus	GCF_002303985.1	66656	Scaffold	15.09.17	28.06.19
Eucalyptus grandis	GCF_000612305.1	71139	Scaffold	02.05.14	28.06.19
Eutrema salsugineum	GCF_000478725.1	72664	Scaffold	05.11.13	28.06.19
Populus euphratica	GCF_000495115.1	75702	Scaffold	12.08.14	28.06.19
Phalaenopsis equestris	GCF_001263595.1	78828	Scaffold	07.08.15	28.06.19
Capsella rubella	GCF_000375325.1	81985	Scaffold	23.04.13	28.06.19
Citrus clementina	GCF_000493195.1	85681	Scaffold	08.11.13	28.06.19
Selaginella moellendorffii	GCF_000143415.4	88036	Scaffold	06.07.10	28.06.19
Herrania umbratica	GCF_002168275.1	108875	Scaffold	05.06.17	28.06.19
Galdieria sulphuraria	GCF_000341285.1	130081	Scaffold	25.02.13	28.06.19
Monoraphidium neglectum	GCF_000611645.1	145388	Scaffold	26.02.15	28.06.19
Jatropha curcas	GCF_000696525.1	180498	Scaffold	02.06.14	28.06.19
Pyrus xbretschneideri	GCF_000315295.1	225117	Scaffold	03.12.12	28.06.19
Chlorella variabilis	GCF_000147415.1	554065	Scaffold	16.09.10	28.06.19
Dendrobium catenatum	GCF_001605985.2	906689	Scaffold	12.12.17	28.06.19
Morus notabilis	GCF_000414095.1	981085	Scaffold	07.08.13	28.06.19