Analysis of the Genome of *Cyanophora paradoxa*: An Algal Model for Understanding Primary Endosymbiosis

Debashish Bhattacharya, Dana C. Price, Cheong Xin Chan, Jeferson Gross, Jürgen M. Steiner, and Wolfgang Löffelhardt

Abstract Algae and plants rely on the plastid (e.g., chloroplast) to carry out photosynthesis. This organelle traces its origin to a cyanobacterium that was captured over a billion years ago by a single-celled protist. Three major photosynthetic lineages (the green algae and plants [Viridiplantae], red algae [Rhodophyta], and Glaucophyta) arose from this primary endosymbiotic event and are putatively united as the Plantae (also known as Archaeplastida). Glaucophytes comprise a handful of poorly studied species that retain ancestral features of the cyanobacterial endosymbiont such as a peptidoglycan cell wall. Testing the Plantae hypothesis and elucidating glaucophyte evolution has in the past been thwarted by the absence of complete genome data from these taxa. Furthermore, multigene phylogenetics has fueled controversy about the frequency of primary plastid acquisitions during

D. Bhattacharya (✉)
Rutgers University, Department of Ecology, Evolution and Natural Resources, Institute of Marine and Coastal Science, 59 Dudley Road, Rutgers University, New Brunswick, NJ 08901, USA
e-mail: bhattacharya@aesop.rutgers.edu

D.C. Price • J. Gross
Department of Ecology, Evolution and Natural Resources, Rutgers University, New Brunswick, NJ 08901, USA

C.X. Chan
Department of Ecology, Evolution and Natural Resources, Rutgers University, New Brunswick, NJ 08901, USA

The University of Queensland, Institute for Molecular Bioscience and ARC Centre of Excellence in Bioinformatics, Brisbane, QLD 4072, Australia

J.M. Steiner
Department of Plant Physiology, Martin-Luther-University Halle-Wittenberg, 06120 Halle (Saale), Germany

W. Löffelhardt
Max F. Perutz Laboratories, Department of Biochemistry and Cell Biology, University of Vienna, 1030 Vienna, Austria

eukaryote evolution because these approaches have generally failed to recover Plantae monophyly and often provide conflicting results. Here, we review some of the key insights about Plantae evolution that were gleaned from a recent analysis of a draft genome assembly from *Cyanophora paradoxa* (Glaucophyta). We present results that conclusively demonstrate Plantae monophyly. We also describe new insights that were gained into peptidoglycan biosynthesis in glaucophytes and the carbon concentrating mechanism (CCM) in *C. paradoxa* plastids.

Contents

Introduction

Independent primary endosymbioses involving Gram-negative bacteria led to the origin of the bioenergetic organelles mitochondria and plastids (e.g., Margulis 1970; Cavalier-Smith and Lee 1985; Gray 1992; Gross and Bhattacharya 2009). These events had lasting impacts on our planet with plastid endosymbiosis giving rise to algae and plants that became a driving force behind Earth's climate, geochemistry, and ecology (Falkowski et al. 2004). Primary plastids are shared by three extant lineages that are referred to as the Plantae (Cavalier-Smith 1981) or Archaeplastida (Adl et al. 2005): the Glaucophyta (glaucophyte algae), the Rhodophyta (red algae), and the Viridiplantae (green algae and plants). Determining the number of primary endosymbioses (i.e., single or multiple) that gave rise to the plastid in these three phyla has long been an open question in algal evolution and directly impacts inference of the tree of life. If it happened as many as three times in the different Plantae lineages that would suggest that these taxa are polyphyletic and establishing the combination of a eukaryotic and a prokaryotic cell is relatively "easy," in evolutionary terms. Most importantly, the resulting chimera converged on similar plastid features for each separate occurrence. If it occurred only once, then the plastid in all major autotrophic eukaryotic lineages traces its origin to this remarkable event in evolution.

A large body of data stemming from phylogenetic and comparative analyses of plastids suggest that primary endosymbiosis occurred a single time in the Plantae ancestor and all plastids [except in the *Paulinella* lineage (e.g., Yoon et al. 2006; Nowack et al. 2011)] trace their ancestry to this singular event (e.g., Bhattacharya et al. 2004; Delwiche 1999; Palmer 2003; Rodriguez-Ezpeleta et al. 2005;

Chan et al. 2011). Nevertheless, many recent nuclear multigene trees provide little (Burki et al. 2007; Patron et al. 2007) or no (Nozaki et al. 2009; Baurain et al. 2010; Parfrey et al. 2010) support for Plantae monophyly and often provide conflicting results. The inability to conclusively support or reject Plantae monophyly (and thereby resolve the number of plastid primary endosymbioses in the eukaryote tree of life) is largely explained by the lack of complete genome data from glaucophytes. This hurdle was recently crossed with the completion and analysis of a draft genome assembly from the glaucophyte *Cyanophora paradoxa* (Price et al. 2012). Intriguingly, the *C. paradoxa* plastid (often referred to as the "muroplast") maintains the ancestral cyanobacterial trait of a peptidoglycan wall (Pfanzagl et al. 1996). This and other traits such as an unconventional carbon-concentrating mechanism (CCM) have made this glaucophyte a model for photosynthesis and endosymbiosis research. Here, we review some of the major features of the *C. paradoxa* genome and the insights it provides into Plantae evolution.

Genome Data

To generate an initial genome draft, a total of 8.3 billion base pairs (Gbp) of Roche 454 and Illumina GAIIx sequence data from *C. paradoxa* CCMP329 (Pringsheim strain) were coassembled with 279 Mbp of random-shear Sanger sequence from this taxon. The resulting assembly comprised 60,119 contigs totaling 70.2 Mbp in size with an N50 of 2.7 kbp (minimum 100 bp and maximum 66 kbp). This highly fragmented assembly is currently being improved by the addition of significant Illumina mate-pair library sequence data. Pulsed field gel electrophoresis analysis of *C. paradoxa* shows the presence of at least seven chromosomes with the smallest being less than 3 Mbp in size (Price et al. 2012). A previous fluorescence-activated cell sorting (FACS) study suggested the haploid genome size in *C. paradoxa* was 140 Mbp (Löffelhardt et al. 1997). Given that the draft genome assembly converged on ca. 70 Mbp, it is likely that the original FACs sorting was done with diploid cells and the haploid genome size of *C. paradoxa* is closer to 70 Mbp.

Analysis of the *C. paradoxa* nuclear genome data (Price et al. 2012) demonstrates a highly enriched G + C-content (i.e., 83.8 % at third codon positions), resulting in difficulties in sequence generation and assembly. To assess whether the assembly was deficient in coding regions due to this issue, we used BLASTN with 3,900 Sanger-derived EST unigenes from *C. paradoxa* to query the draft assembly. This analysis showed that 99.1 % of the ESTs (i.e., putative protein coding regions) had hits (at $e \leq 10^{-10}$). This suggests the majority of expressed genes are represented in the genome data. Thereafter, a total of 15 Gbp of Illumina mRNA-seq data was used to train ab initio gene predictors, resulting in 27,921 weighted consensus gene structures. The organelle genomes of *C. paradoxa* were also analyzed, including novel data from its sister glaucophyte, *Glaucocystis nostochinearum*, but will not be discussed here (for details, see Price et al. 2012).

Testing Plantae Monophyly

Several approaches were taken to ascertain support for the monophyly of Plantae. These included phylogenetic analysis of single proteins, elucidating the extent and type of endosymbiotic or horizontal gene transfer (E/HGT), comparative analysis of groups of proteins such as components of the plastid protein translocons and fermentation pathways, and analysis of plastid solute transporters. All of these data strongly support a single origin of Plantae and therefore a single primary plastid endosymbiosis in their common ancestor (Price et al. 2012). Here, we present the results of the single protein and plastid translocon analyses.

For the single proteins, we first used BLASTP to analyze the evolutionary affiliations of the 27,921 predicted protein models in *C. paradoxa*. A total of 4,628 proteins had significant BLASTP hits ($e \leq 10^{-10}$) to sequences in a comprehensive local database that we use for comparative analysis (e.g., Moustafa et al. 2009; Chan et al. 2011). A simplified reciprocal BLAST best-hits approach (Chan et al. 2011) identified a total of 606 proteins that had hits only to one other phylum (i.e., exclusive gene sharing). With the requirement of an increasing number of hits per query (x) from *C. paradoxa* and the second phylum, as $x \geq 2$ (606 proteins), $x \geq 10$ (125 proteins), and $x \geq 20$ (23 proteins), we found that *C. paradoxa* and Viridiplantae shared the largest number of exclusive genes (Fig. 1), indicating a close evolutionary relationship between these lineages. Next, using an automated approach (Chan et al. 2011), we generated 4,445 maximum likelihood (ML) trees for *C. paradoxa* proteins that had significant database hits. To minimize the impact of taxon sampling on this analysis, we considered trees that contained ≥ 3 phyla and a minimum number of terminal taxa (N) that ranged from 4 to 40 (Fig. 2a). Using this approach, we found that >60 % of all trees support (at bootstrap ≥ 90 %) a sister group relationship between glaucophytes and red and/or green algae. The glaucophytes were most often positioned as sister to Viridiplantae (105, 83, 48, 19, and 10 trees at $N = 4$, 10, 20, 30, and 40), consistent with the analysis of exclusive gene sharing with only a small number of trees favoring the monophyly of glaucophytes and red algae. This result was found even though a significant number of trees favored glaucophyte–red–green (Plantae) monophyly (44, 40, 32, 18, and 16 trees at $N = 4$, 10, 20, 30, and 40) and we had substantial red algal genome data in our database (361,625 sequences). Interestingly, many of the trees showed *C. paradoxa* to be monophyletic with other Plantae in a clade ("shared") that also included non-Plantae phyla (GlR/GlGr/GlRGr in Fig. 2a). When we sorted the phylogenomic output using the red or green algae as the query to test Plantae monophyly, these results also identified Plantae as the most frequently recovered clade (Fig. 2b, c). However, both red and green algae show far more gene sharing than glaucophytes because they, unlike glaucophytes, are implicated in secondary endosymbioses that have resulted in their genes being spread throughout the tree of life to groups such as "chromalveolates" and euglenids (Harper and Keeling 2003; Moustafa et al. 2009; Baurain et al. 2010; Chan et al. 2011). Given that single protein

Fig. 1 Exclusive gene sharing between Glaucophyta and one other taxon, when the total number of hits (*x*) was ≥2 (**a**), ≥10 (**b**), and ≥20 (**c**). The *green slices* get larger as *x* increases. The vira matches include prasinophyte (*Bathycoccus*, *Micromonas*, and *Ostreococcus*) and *Chlorella* viruses
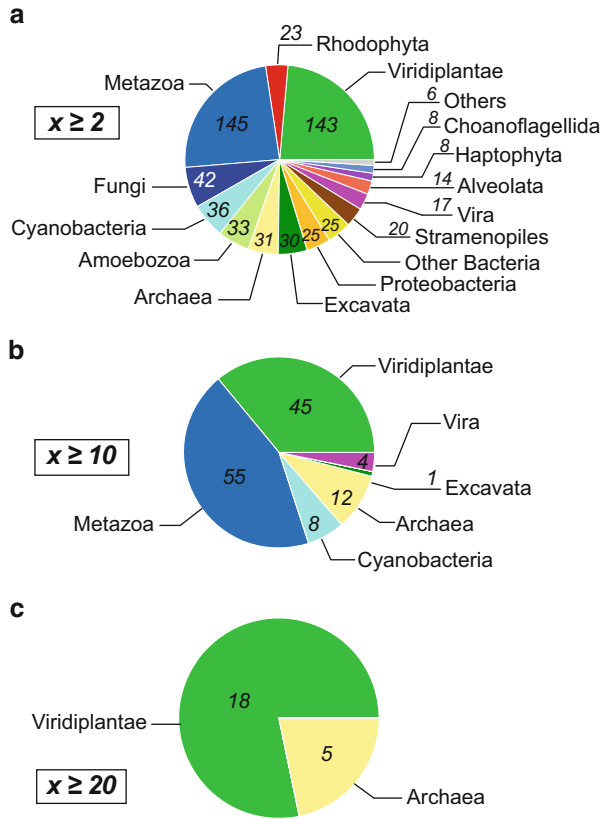
trees firmly establish glaucophytes as members of the Plantae we analyzed a landmark trait of Plantae, the plastid protein translocons.

A key innovation required for the cyanobacterium-to-plastid evolutionary transition in primary endosymbiosis was the establishment of protein translocons for protein targeting into the emergent organelle (e.g., Reumann et al. 2005; Gross and Bhattacharya 2008, 2009). Components of the *T*ranslocons at the *o*uter and *i*nner envelope membranes of *c*hloroplasts (Toc and Tic, respectively) have been described in higher plants, and algae of the green, red, and "chromalveolate" lineages (McFadden and van Dooren 2004). The existence of an analogous protein import system in *C. paradoxa* is suggested by immunological detection of epitopes in this alga using plant Toc75 and Tic110 antibodies, and heterologous protein import assays (Steiner et al. 2005; Yusa et al. 2008). These data suggest that all Plantae share a key invention that laid the foundation for plastid integration within the host cell. Our analysis of the *C. paradoxa* genome identified homologs of Toc75 and Tic110 that are OEM (outer envelope membrane) and IEM (inner envelope membrane) protein conducting channels, respectively, two Toc34-like receptors, as well as homologs of the plastid Hsp70 and Hsp93 chaperones, and stromal
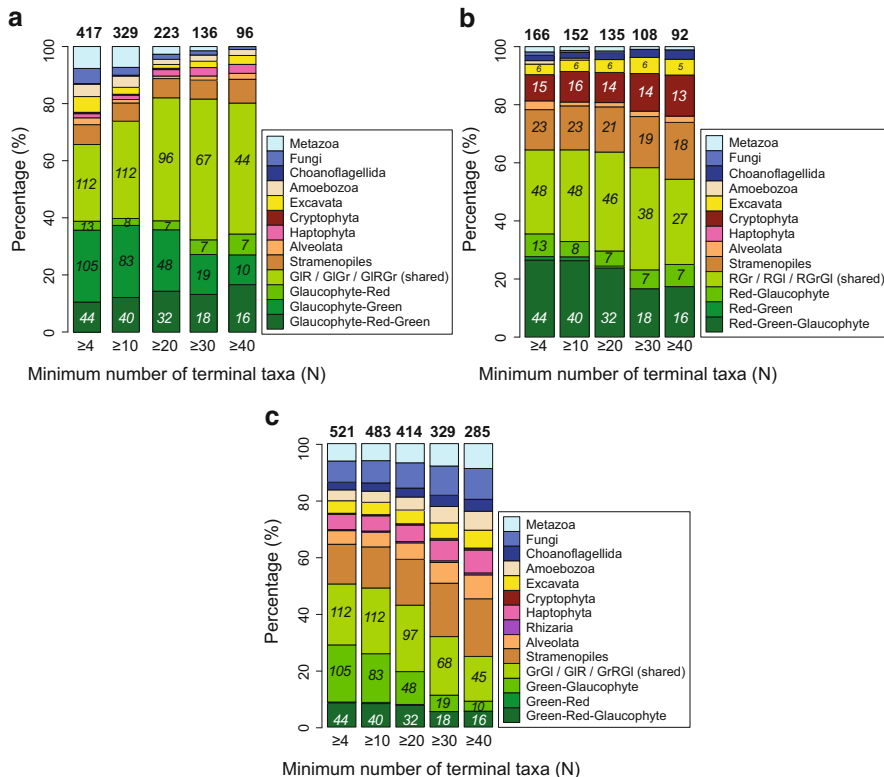
Fig. 2 Testing Plantae monophyly. (**a**) Percentage of maximum likelihood single protein trees that support the monophyly of Glaucophyta (bootstrap ≥90 %) with other members of the Plantae, or in combination with non-Plantae taxa that interrupt this clade. These latter groups of trees are primarily explained by red or green algal endosymbiotic gene transfer (EGT) into the nuclear genome of "chromalveolates" and euglenids. For each of these algal lineages, the set of trees with different numbers of terminal taxa (*N*) ≥4, ≥10, ≥20, ≥30, and ≥40 and distinct phyla ≥3 in a tree are shown. Similar analyses that were done using red algae and Viridiplantae as the query to test for Plantae monophyly are shown in panels (**b**) and (**c**), respectively

processing peptidase (Price et al. 2012). Such a minimal set of components is likely to have formed the primitive protein translocation system in the Plantae ancestor (Gross and Bhattacharya 2008, 2009). Candidates for additional translocon subunits were also detected in *C. paradoxa*. Furthermore, the Tic20 and Toc22 ML phylogenies provide unambiguous evidence for a cyanobacterial provenance of these genes in Plantae and a monophyletic relationship of *C. paradoxa* with plants and other algae (see Price et al. 2012). In summary, the evolution of protein translocons to the nascent plastid has long been held as a formative event in the emergence of the Plantae ancestor. Analysis of the *C. paradoxa* genome revealed the presence of the conserved core of translocon subunits derived from the cyanobacterial endosymbiont (i.e., Toc75, Tic20, Tic22) as well as novel genes

that apparently evolved de novo in the host (i.e., Toc34 and Tic110). These data provide further unambiguous evidence that the primary plastid was established in a single common ancestor of the Plantae.

Enzymes of Peptidoglycan Biosynthesis

The muroplast wall of glaucophyte algae is the sole documented example of peptidoglycan (PG) in Plantae and its origin from the plastid endosymbiont is noncontroversial. The PG consists of one giant molecule ("sacculus") and belongs to the A1gamma type, like the cell walls of *Escherichia coli* and cyanobacteria, but is thicker and more cross-linked than in the former and more reduced than in the latter. A unique feature of the PG is its modification with *N*-acetylputrescine (Löffelhardt and Bohnert 2001). The space between the inner and outer envelope membranes of muroplasts, the "periplasmic space", harbors the peptidoglycan layer and enzymes for its synthesis, modification, and degradation.

PG biosynthesis has been studied in great detail in *E. coli* and can be divided into cytoplasmic, membrane-bound, and periplasmic steps. This three-step compartmentalization process is also present in *C. paradoxa*: (1) biosynthesis of the disaccharide–pentapeptide precursor occurs in the muroplast stroma (activities of MurA and MurF have been shown), (2) its transfer to the lipid carrier at the inner envelope membrane, and (3) its insertion into growing PG chains in the periplasmic space (Löffelhardt and Bohnert 2001). The latter step is catalyzed by penicillin binding proteins (PBPs) that have transglycosylase and/or transpeptidase activities. Seven PBPs that range in size from 35 to 110 kDa were identified in the muroplast envelope by labeling with a radioactive derivative of ampicillin. In addition, enzymatic activities of DD- and LD-carboxypeptidases and DD-endopeptidase that hydrolyze defined bonds in PG have been demonstrated in muroplasts (Löffelhardt and Bohnert 2001).

Here, three different approaches were used for PBP gene identification (1) domain search; (2) BLAST search against the eight PBP genes of *Synechocystis* sp. PCC6803 (Marbouty et al. 2009) and the *Anabaena* sp. PCC7120 homologs; and (3) BLAST search against *Physcomitrella patens* PBP-like genes. In most cases, the results converged leading to at least 11 genes or gene fragments being identified in *C. paradoxa* (Table 1). In general, sequence similarity was higher to homologs in cyanobacteria than those in *P. patens*. No *C. paradoxa* homologs to the small PBPs 6 and 7 were identified. The PBP numbering scheme applied here is from *E. coli*. However, sequence similarity (especially among the large PBPs) is significant which is reflected in their redundant function.

In some cases of periplasmic proteins, bipartite presequences consisting of a transit peptide and a signal peptide can be envisaged. This suggests import to the muroplast stroma, followed by export to the periplasmic space. This special variant of "conservative sorting" would necessitate a dual location of Sec (already

Table 1 Nuclear genes involved in the biosynthesis of plastid peptidoglycan in *C. paradoxa*

Gene/protein	Function[a]	Contig
PBP3, PBP1, PBP2	PG transglycosylase/transpeptidase	11,577
PBP3, PBP1, PBP2	PG transglycosylase/transpeptidase	55,323
PBP3, PBP1	PG transglycosylase/transpeptidase	10,395
ftsI/PBP4	PG transglycosylase/transpeptidase	11,029
PBP?	PG transpeptidase	15,041
PBP 8	D-Ala-D-Ala-carboxypeptidase C	40,415
dacB/PBP 5	D-Ala-D-Ala-carboxypeptidase	7,465
	D-Ala-D-Ala-endopeptidase	
dacB/PBP5	D-Ala-D-Ala-carboxypeptidase	6,898
	D-Ala-D-Ala-endopeptidase	
vanX	D-Ala-D-Ala-dipeptidase	54,463
vanY/endolysin	D-Ala-D-Ala-carboxypeptidase	15,693
Lysozyme-like	Muramidase	54,844
mlt	Lytic transglycosylase	9,267
glmS	Glucosamin-6-P synthase	39,008
murA	UDP-*N*-acetylglucosamine-1-carboxyvinyl transferase	53,904
murA	UDP-*N*-acetylglucosamine-1-carboxyvinyl transferase	7,908
murB	UDP-*N*-acetylenolpyruvoyl-glucosamine reductase	8,011
murB	UDP-*N*-acetylenolpyruvoyl-glucosamine reductase	40,266
murC	UDP-*N*-acetylmuramate:L-Ala ligase	17,182
murI	Glutamate racemase	25,539
murD	D-Glu adding enzyme	39,147
murE	DAP-adding enzyme	9,035
Alr	Alanine racemase	53,160; 386
Ddl	D-Ala:D-Ala ligase	7,167
murF	UDP-*N*-acetylmuramoyl tripeptide/D-Ala-D-Ala ligase	52,912
mraY	Lipid I synthesis	37,977
murG	Lipid II synthesis	54,819

[a]The high MW (1–4) and the medium MW (5–8) PBPs are redundant in *Synechocystis* sp. PCC6803

documented) and Tat (seems possible as another parallel to cyanobacteria) translocases on thylakoid and inner envelope membranes of muroplasts. In a Gram-negative background, the low molecular weight (MW) peptidases VanX and VanY are not linked to vancomycin resistance but rather to D-alanine recycling and to an additional endolysin, respectively. Peptidoglycan biosynthesis requires cleavage of existing glycan chains to allow for insertion of new material. This is

performed by soluble and membrane-bound lytic transglycosylases: one gene of this kind could also be identified in *C. paradoxa*. A lysozyme family protein with significant similarity to protist lysozymes displays a signal peptide indicating a vacuolar (lysosomal) location that is likely involved in the autophagosomal digestion of damaged muroplasts. Genes for stromal proteins that are involved in the synthesis of the soluble precursor are denoted as *glm*. The N-terminal transit peptide identifies one protein in *C. paradoxa* (*glmS*) as a member of the muroplast-resident PG biosynthesis pathway, whereas a cytosolic counterpart would be expected to participate in protein glycosylation. The complete list of enzymes in the alga that are involved in UDP-*N*-acetylmuramate biosynthesis as well as the peptide side-chain adding enzymes, and the alanine (Alr) and glutamate (MurI) racemases are listed in Table 1. The membrane-bound or associated MraY and MurG proteins complete this compilation.

Genes for enzymes of PG biosynthesis were transferred twice into Plantae in the course of evolution—from the mitochondrial ancestor and from the cyanobacterial ancestor of plastids. These remain recognizable in sequence from *Arabidopsis thaliana* (few genes) to the moss *P. patens* (almost complete set), but their functions are likely to have changed. As long as chemical and structural proof is lacking (pleiotropic), effects of antibiotics or gene knock-outs of plastid division do not provide sufficient evidence to claim the presence and biosynthesis of PG in the plastids of bryophytes (Takano and Takechi 2010). Glaucophyte PG is unique in Plantae. In *Paulinella*, the situation is different: there is also PG in this eukaryote, but all genes necessary for its biosynthesis (Marin et al. 2007) are encoded on the endosymbiont (i.e., "chromatophore", photosynthetic organelle) genome, which exceeds the size of plastid genomes by a factor of 5–10. Unlike their counterparts in *C. paradoxa*, these genes retain their prokaryotic character, i.e., they were not transferred to the nuclear genome and thus no import of precursor proteins is required for biosynthesis of the sacculus in photosynthetic *Paulinella* species.

The correlation of more than one gene to a given function is not uncommon among cyanobacteria. A second gene with high sequence similarity to *murG* is more closely related to MGDG synthases, the likely function of "MurG" in plants. In an analogous fashion, *murD*-like genes might instead play a role in folate biosynthesis. Until the presence of PG in *P. patens* is unequivocally proven, one should expect modified functions for "mur-like" genes. The fact that the cyanobacterial counterparts are often, but not always the top hits suggests a mosaic structure of the gene complement for PG biosynthesis in *C. paradoxa*. HGT from bacteria (e.g., Firmicutes and Verrucomicrobia) is likely to be prominent when the transferred genes provide a required function, i.e., PG biosynthesis in the case of glaucophytes. In addition, gene replacement might have occurred in some cases.

The Rubisco-Containing Microcompartment of Muroplasts: Carboxysome Versus Pyrenoid

The conspicuous, electron-dense central body of *C. paradoxa* muroplasts described in most publications was named a carboxysome (Raven 2003; Fathinejad et al. 2008). This coinage did not take into account the fact that eukaryotes contain pyrenoids to fulfill the function of a carbon-concentrating mechanism (CCM) and emphasized the often-postulated transitional position of glaucophytes between plastids and cyanobacteria. However, all of our attempts to identify carboxysomal shell proteins in the *C. paradoxa* genome failed, either with domain searches (BMC = Pfam 00936 or Pfam 03319) or with a concatenated dataset of cyanobacterial CcmKLMNO sequences. Indeed, it might be problematic to harbor shell protein genes in the nucleus, because they have high affinities to each other and likely self-assemble as carboxysomal prestructures (Kinney et al. 2011), thereby interfering with protein import into muroplasts. Thus far, *Paulinella* constitutes the only example of "eukaryotic carboxysomes". Again, the necessary genes remain on the plastid genome, interestingly derived via HGT (Marin et al. 2007). In any case, the hypothesis of peptidoglycan retention in *C. paradoxa* (Raven 2003) to stabilize the plastid against the osmotic pressure of bicarbonate that is enriched more than 1,000-fold in the stroma through the action of the carboxysomal CCM could not be verified. In contrast, evidence was obtained (Table 2) for a number of proteins (LciB, C, and D) with functions in the pyrenoidal CCM of *Chlamydomonas reinhardtii* (Yamano et al. 2010). LciB and LciC were shown to form a hexameric complex (ca. 360 kDa) under active operation of the CCM: light and low concentration of CO_2. This complex localizes close to the pyrenoid but is relocalized from the pyrenoid to the stroma upon high CO_2 concentration or darkness. There seems to be no connection to pyrenoid development and/or starch sheath formation. A role is assumed in trapping of CO_2 that has escaped from the pyrenoid via interaction with the carbonic anhydrase Cah6 and, eventually, also in accumulating CO_2 reaching the stroma from the cytosol, i.e., in the active uptake of CO_2 in *C. reinhardtii* (Wang et al. 2011). Alternatively, physical blockage of CO_2 from escaping the pyrenoid by the complex has been postulated (Yamano et al. 2010). The complex is not required under high levels of CO_2. In this case, a function similar to the cyanobacterial shell proteins CcmK and CcmL (which, however, are present under all conditions) can be envisaged. Some putative cyanobacterial plastid ancestors contain LciB and LciC, given their filamentous nature (*Lyngbya*) or capability of producing a starch-like reserve carbohydrate (*Cyanothece*). These bacteria might use mechanisms of the type discussed above that are superimposed on their carboxysomal CCM. If carboxysomes were transferred to early plastids via endosymbiosis, the separation between carboxysomal and pyrenoidal CCM could have occurred within the phylum Glaucophyta, i.e., *C. paradoxa* and *G. nostochinearum* already progressed towards a pyrenoidal CCM, whereas *Gloeochaete wittrockiana* and *Cyanoptyche gloeocystis*, with their polyhedral microcompartments confined by an electron-dense shell-like

Table 2 Genes for proteins involved in the CCM of *Cyanophora paradoxa*

Gene	Function	Contig	Comments
LciA[a]	Bicarbonate transport	8,717	TP, complete
LciA	Bicarbonate transport	53,293	TP, complete
LciB[a]	CCM	37,097	TP, complete
LciB	CCM	53,135	TP, complete
LciB, *LciD*?	CCM	54,037	TP, 3′-truncated
LciB, *LciC*?	CCM	25,875	Fragment
rca[a]	Rubisco activase	26,296	TP, complete
?[a]	Carbonic anhydrase	52,891	Beta-CA superfamily, cytosolic
Cah4[a]	Carbonic anhydrase	54,421	Beta-CA superfamily, mitochondrial
Cah5[a]	Carbonic anhydrase	9,670	Beta-CA superfamily, mitochondrial
?	Carbonic anhydrase	38,132	Gamma-CA family, cytosolic?
?	Carbonic anhydrase	53,783	Gamma-CA family, cytosolic

[a]CO_2-responsive gene

TP muroplast transit peptide

layer, might have retained the carboxysomal CCM (Fathinejad et al. 2008). Under such a scenario, the *ccmKLMNO* genes would be expected to reside on the muroplast genomes of *G. wittrockiana* and *C. gloeocystis*. The PG wall, though no longer necessary, was retained for unknown reasons in the plastids of *C. paradoxa* and *G. nostochinearum*. Table 2 includes two genes encoding the putative bicarbonate transporter LciA and several genes with strong sequence similarity to genes for LciB, LciC, and LciD from *C. reinhardtii*. Because these are closely related, an exact assignment is difficult. However, whenever the N-termini are intact, unequivocal muroplast presequences were found for these enzymes.

A key enzyme of the CCM is carbonic anhydrase, either copackaged with Rubisco in cyanobacterial carboxysomes or located in the lumen of thylakoids traversing the pyrenoid of *C. reinhardtii*. The number of CAs can vary among algae, e.g., from 9 in *C. reinhardtii* to 13 in some diatoms (Tachibana et al. 2011). Five CAs from *C. paradoxa* are shown in Table 2. Two of these belong to the gamma-CA family with high sequence similarity to homologs in plants. The other three contain the conserved Zn-binding site (VCGHSHCGAMKG) of (cyano) bacterial beta-CAs. In the case of the putative mitochondrial CAs, high sequence similarity to *C. reinhardtii* Ca1 and Ca2 is observed. A bona fide muroplast CA (e.g., the stromal Cah6 or the lumenal Cah3 of *C. reinhardtii*) is missing from this compilation. If we assume a pyrenoidal CCM in *C. paradoxa*, the organism must utilize a mechanism different from that in *C. reinhardtii*. There is no evidence of a thylakoid-lumenal CA or a muroplast microcompartment traversed by thylakoid membranes. In the diatom *Phaeodactylum tricornutum*, the carbonic anhydrase CA-1 (CO_2 responsive) is copackaged with pyrenoidal Rubisco and does not reside in the lumen of the traversing thylakoid (Tachibana et al. 2011). Mass spectrometric analysis of central body proteins from *C. paradoxa* did not reveal a CA-like protein. The only outcome of these studies (in addition to Rubisco LSU and SSU) was

Rubisco activase that was also corroborated by Western blotting and assembly studies after in vitro import into isolated muroplasts (Fathinejad et al. 2008). *C. paradoxa* activase, while showing high sequence similarity to both cyanobacterial and plant homologs, lacks the C-terminal extension typical for filamentous cyanobacteria. This protein contains a domain that shares high sequence similarity with repetitive regions found in the largest carboxysome shell protein CcmM. An N-terminal extension present in plant homologs is present in the *C. paradoxa* protein. Taken together, the domain structure of Rubisco activase from *C. paradoxa* does not support the carboxysome concept. Several genes listed in Table 2 were shown to be CO_2 responsive in the closely related *C. paradoxa* SAG 45.84 (Kies strain) underlining their postulated role in the CCM (Burey et al. 2007).

Conclusions

Rather than being a relict lineage, the analyses presented here and in Price et al. (2012) paint a picture of the "living fossil" *C. paradoxa* as a gene- (and function)-rich species that provides many clues to early events in plastid endosymbiosis and Plantae evolution. These data unambiguously support Plantae monophyly, thereby answering a fundamental question about the eukaryote tree of life. The components of the peptidoglycan biosynthetic pathway in *C. paradoxa* were identified and indicated a cyanobacterial provenance of many key enzymes with likely instances of recruitment of additional genes via HGT from other prokaryote sources. Finally, evidence was found that strongly argues against the existence of a proposed eukaryotic carboxysome in *C. paradoxa*. The available data are more consistent with a pyrenoidal CCM in this species and in its sister *G. nostochinearum*. However, the mechanism of CCM function is likely to be different from that found in *C. reinhardtii*.

References

Adl SM, Simpson AG, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, James TY, Karpov S, Kugrens P, Krug J, Lane CE, Lewis LA, Lodge J, Lynn DH, Mann DG, McCourt RM, Mendoza L, Moestrup O, Mozley-Standridge SE, Nerad TA, Shearer CA, Smirnov AV, Spiegel FW, Taylor MF (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol 52:399–451

Baurain D, Brinkmann H, Petersen J, Rodríguez-Ezpeleta N, Stechmann A, Demoulin V, Roger AJ, Burger G, Lang BF, Philippe H (2010) Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. Mol Biol Evol 27:1698–1709

Bhattacharya D, Yoon HS, Hackett JD (2004) Photosynthetic eukaryotes unite: endosymbiosis connects the dots. Bioessays 26:50–60

Burey SC, Poroyko V, Ozturk ZN, Fathi-Nejad S, Schüller C, Ohnishi N, Fukuzawa H, Bohnert HJ, Löffelhardt W (2007) Acclimation to low [CO2] by an inorganic carbon concentrating mechanism in *Cyanophora paradoxa*. Plant Cell Environ 30:1422–1435

Burki F, Shalchian-Tabrizi K, Pawlowski J (2007) Phylogenomics reshuffles the eukaryotic supergroups. PLoS ONE 2:e790

Cavalier-Smith T (1981) Eukaryote kingdoms: seven or nine? BioSystems 14:461–481

Cavalier-Smith T, Lee JJ (1985) Protozoa as hosts for endosymbioses and the conversion of symbionts into organelles. J Protozool 32:376–379

Chan CX, Yang EC, Banerjee T, Yoon HS, Martone PT, Estevez JM, Bhattacharya D (2011) Red-and-green algal monophyly and extensive gene sharing found in a rich repertoire of red algal genes. Curr Biol 21:328–333

Delwiche CF (1999) Tracing the thread of plastid diversity through the tapestry of life. Am Nat 154:S164–S177

Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O, Taylor FJ (2004) The evolution of modern eukaryotic phytoplankton. Science 305:354–360

Fathinejad S, Steiner JM, Reipert S, Marchetti M, Allmaier G, Burey SC, Ohnishi N, Fukuzawa H, Löffelhardt W, Bohnert HJ (2008) A carboxysomal CCM in the cyanelles of the "coelacanth" of the algal world, *Cyanophora paradoxa*? Physiol Plant 133:27–32

Gray MW (1992) The endosymbiont hypothesis revisited. Int Rev Cytol 141:233–357

Gross J, Bhattacharya D (2008) Revaluating the evolution of the Toc/Tic protein translocons. Trends Plant Sci 14:13–20

Gross J, Bhattacharya D (2009) Mitochondrial and plastid evolution in eukaryotes: an outsider's perspective. Nat Rev Genet 10:495–505

Harper JT, Keeling PJ (2003) Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. Mol Biol Evol 20:1730–1735

Kinney JN, Axen SD, Kerfeld CA (2011) Comparative analysis of carboxysome shell proteins. Photosynth Res 109:21–32

Löffelhardt W, Bohnert HJ (2001) The cyanelle (muroplast) of *Cyanophora paradoxa*: a paradigm for endosymbiotic organelle evolution. In: Seckbach J (ed) Symbiosis. Kluwer, Dordrecht, pp 111–130

Löffelhardt W, Bohnert HJ, Bryant DA (1997) The complete sequence of the *Cyanophora paradoxa* cyanelle genome (Glaucocystophyceae). Plant Syst Evol 11(Suppl):149–162

Marbouty M, Mazouni K, Saguez C, Cassier-Chauvat C, Chauvat F (2009) Characterization of the *Synechocystis* strain PCC 6803 penicillin-binding proteins and cytokinetic proteins FtsQ and FtsW and their network of interactions with ZipN. J Bacteriol 191:5123–5133

Margulis L (1970) Origin of eukaryotic cells. Yale University Press, New Haven, CT, p 349

Marin B, Nowack ECM, Glöckner G, Melkonian M (2007) The ancestor of the *Paulinella* chromatophore obtained a carboxysomal operon by horizontal gene transfer from a *Nitrococcus*-like gamma-proteobacterium. BMC Evol Biol 7:85

McFadden GI, van Dooren GG (2004) Evolution: red algal genome affirms a common origin of all plastids. Curr Biol 14:R514–R516

Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D (2009) Genomic footprints of a cryptic plastid endosymbiosis in diatoms. Science 324:1724–1726

Nowack EC et al (2011) Endosymbiotic gene transfer and transcriptional regulation of transferred genes in *Paulinella chromatophora*. Mol Biol Evol 28:407–422

Nozaki H, Maruyama S, Matsuzaki M, Nakada T, Kato S, Misawa K (2009) Phylogenetic positions of Glaucophyta, green plants (Archaeplastida) and Haptophyta (Chromalveolata) as deduced from slowly evolving nuclear genes. Mol Phylogenet Evol 53:872–880

Palmer JD (2003) The symbiotic birth of plastids: how many times and whodunit? J Phycol 39:4–11

Parfrey LW, Grant J, Tekle YI, Lasek-Nesselquist E, Morrison HG, Sogin ML, Patterson DJ, Katz LA (2010) Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. Syst Biol 59:518–533

Patron NJ, Inagaki Y, Keeling PJ (2007) Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. Curr Biol 17:887–891

Pfanzagl B, Zenker A, Pittenauer E, Allmaier G, Martinez-Torrecuadrada J, Schmid ER, De Pedro MA, 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

Price DC, Chan CX, Yoon HS, Yang EC, Qiu H, Weber APM, Schwacke R, Gross J, Blouin NA, Lane C, Reyes-Prieto A, Durnford DG, Neilson JAD, Lang BF, Burger G, Steiner JM, Löffelhardt W, Meuser JE, Posewitz MC, Ball S, Arias MC, Henrissat B, Coutinho PM, Rensing SA, Symeonidi A, Doddapaneni H, Green BR, Rajah VD, Boore J, Bhattacharya D (2012) *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. Science 335:843–847

Raven JA (2003) Carboxysomes and peptidoglycan walls of cyanelles: possible physiological functions. Eur J Phycol 38:47–53

Reumann S, Inoue K, Keegstra K (2005) Evolution of the general protein import pathway of plastids (review). Mol Membr Biol 22:73–86

Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W, Bohnert HJ, Philippe H, Lang BF (2005) Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr Biol 15:1325–1330

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

Tachibana M, Allen AE, Kikutani S, Endo Y, Bowler C, Matsuda Y (2011) Localization of putative carbonic anhydrases in two marine diatoms, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. Photosynth Res 109:205–221

Takano H, Takechi K (2010) Plastid peptidoglycan. Biochim Biophys Acta 1800:144–151

Wang Y, Duanmu D, Spalding MH (2011) Carbon dioxide concentrating mechanism in *Chlamydomonas reinhardtii*: inorganic carbon transport and CO2 recapture. Photosynth Res 109:115–122

Yamano T, Tsujikawa T, Hatano K, Ozawa S, Takahashi Y, Fukuzawa H (2010) Light and low-CO2 -dependent LCIB–LCIC complex localization in the chloroplast supports the carbon-concentrating mechanism in *Chlamydomonas reinhardtii*. Plant Cell Physiol 51:1453–1468

Yoon HS, Reyes-Prieto A, Melkonian M, Bhattacharya D (2006) Minimal plastid genome evolution in the *Paulinella* endosymbiont. Curr Biol 16:R670–R672

Yusa F, Steiner JM, Löffelhardt W (2008) Evolutionary conservation of dual Sec translocases in the cyanelles of *Cyanophora paradoxa*. BMC Evol Biol 8:304