Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution 45 (2007) 384–391
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Short communication

Phylogeny of Calvin cycle enzymes supports Plantae monophyly

Adrian Reyes-Prieto, Debashish Bhattacharya *

University of Iowa, Department of Biological Sciences and the Roy J. Carver Center for Comparative Genomics, 446 Biology Building, Iowa City, IA 52242, USA

Received 10 November 2006; revised 5 February 2007; accepted 13 February 2007
Available online 4 March 2007

1. Introduction

Photosynthesis is a critical biochemical process on our planet providing food for most life. The common ancestor of plants and their algal sisters gained photosynthesis through the engulfment and retention of a cyanobacterial primary endosymbiont that evolved into a photosynthetic organelle, the plastid (Bhattacharya et al., 2004). In photosynthetic eukaryotes, the essential series of reactions that capture the products of photosynthetic light reactions (ATP and $NADPH_2$) to fix CO_2 (Fig. 1), known as the Calvin cycle (CC; Calvin and Benson, 1948), takes place in the plastid stroma. The eukaryotic CC involves 11 different enzymes (Table 1) that are nuclear encoded and plastid targeted to express their function, with the exception of ribulose-1,5-bisphosphate carboxylase (RuBisCO) subunits (large and small) that remain plastid encoded in red and glaucophyte algae. In green algae (and land plants) the RuBisCO large subunit is encoded in the plastid genome but the small subunit is nuclear encoded. Photosynthetic eukaryotes also contain cytosolic enzymes involved in glycolysis and gluceoneogenesis that catalyze reactions similar to those in the CC and were present in eukaryotes before plastid origin (Martin and Schnarrenberger, 1997). Molecular phylogenetic analyses suggest that land plants (Martin and Schnarrenberger, 1997) and red algae acquired at least a subset of the CC enzymes via intracellular (endosymbiotic) gene transfer (EGT) from the captured cyanobacterium prior to the divergence of green and red algae (Matsuzaki et al., 2004). However, it is well known that some CC enzymes in land plants and red algae have a non-cyanobacterial origin (Martin and Schnarrenberger, 1997; Matsuzaki et al., 2004). A likely explanation is that these pre-existing host enzymes took over the role of the original cyanobacterial proteins that have been lost over evolutionary time. It is unknown whether these putative gene replacements were an ancient feature of eukaryotic CC evolution or whether gene recruitments-replacements occurred more recently in different photosynthetic lineages. To gain a clearer picture of early CC enzyme evolution it is critical to analyze genome data from the three major groups that presumably diversified from the first photosynthetic eukaryote: the red algae, the green algae (including land plants), and the glaucophyte algae. These taxa, referred to as the Plantae (Cavalier-Smith, 1981) or Archaeplastida (Adl et al., 2005), are postulated to share a single origin. Phylogenetic analyses using multi-gene nuclear and plastid data strongly support Plantae monophyly (e.g., Rodriguez-Ezpeleta et al., 2005), however some single- and multi-gene analyses do not recover this clade (e.g., Nozaki et al., 2003; Stiller and Harrell, 2005).

Phylogenetic analysis of enzymes involved in conserved, ancient metabolic processes is a potentially valuable source of information for elucidating the evolutionary history of major eukaryotic groups. Here, we use this comparative approach to elucidate the phylogeny of CC enzymes with molecular data from all three primary photosynthetic lineages that share the cyanobacterial endosymbiont; i.e., including expressed sequence tag (EST) data generated in our lab from the glaucophyte alga *Cyanophora paradoxa* (Reyes-Prieto et al., 2006).

2. Materials and methods

2.1. cDNA library construction

Total RNA from a culture of *Cyanophora paradoxa* Pringsheim strain (CCMP329) was extracted with Trizol (GibcoBRL) and the mRNA purified using the Oligotex

* Corresponding author. Fax: +1 319 335 1069.
E-mail address: debashi-bhattacharya@uiowa.edu (D. Bhattacharya).

1055-7903/$ - see front matter
doi:10.1016/j.ympev.2007.02.026

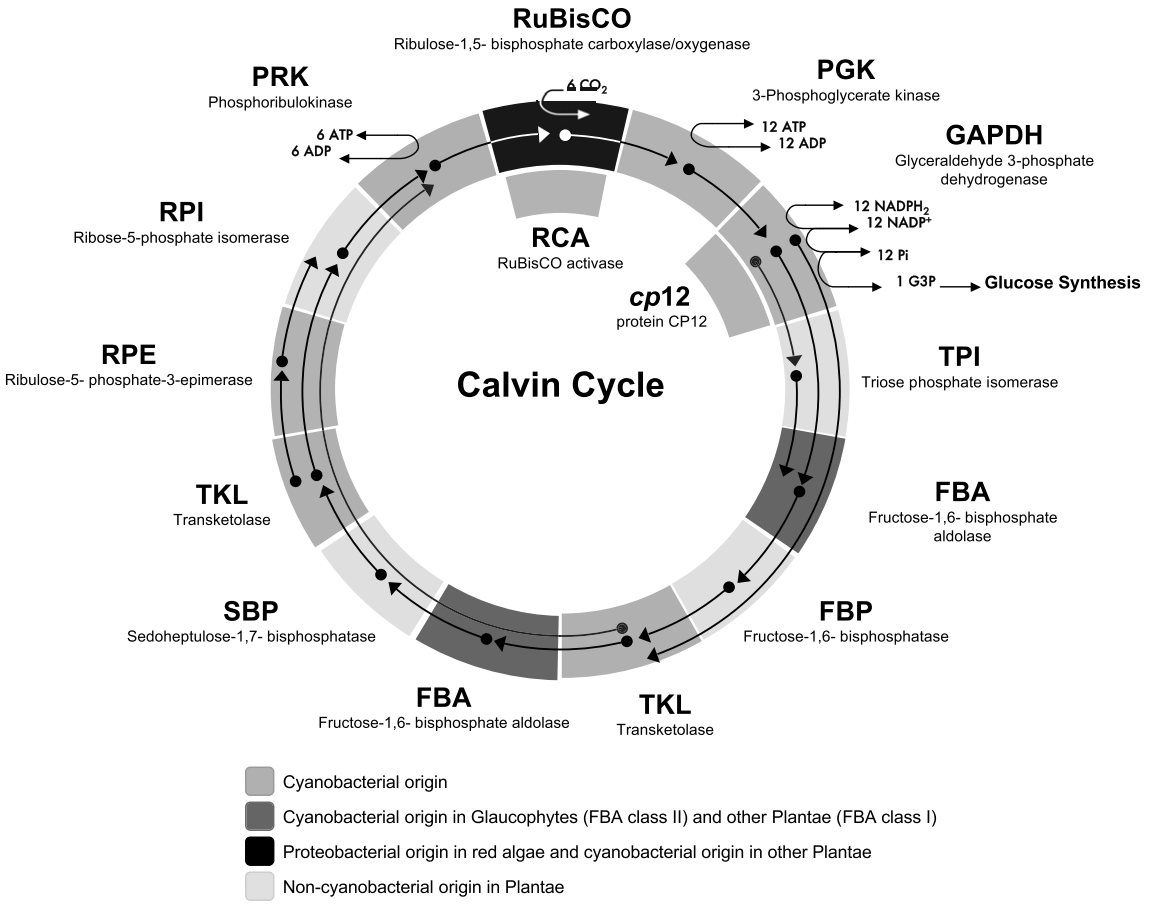

Fig. 1. The Calvin cycle of algae and plants. The CC is comprised of proteobacterial RuBisCO that is present in red algae (black circle region) and cyanobacterial PGK, GAPDH, TKL, RPE, and PRK (medium gray circle regions) that is present in all Plantae. The cyanobacterial TPI, FBP/SBP, and RPI (light gray circle regions) genes were replaced with host genes in the Plantae ancestor, whereas two unrelated FBA classes (I and II; dark gray circle regions) are present in cyanobacteria (i.e., both coexist in *Synechococcus* sp. CC9902), and were differentially lost during Plantae evolution (see text). Circular lines connect the enzymatic product (dot) used as substrate for the next corresponding enzyme (arrowhead) in the CC sequential reactions.

Table 1
Plastid Calvin cycle enzymes in Plantae

	Enzyme		Subunits and/or class
1	Ribulose-1,5-bisphosphate carboxylase large subunit	RuBisCO	RBCL class I
	Ribulose-1,5-bisphosphate carboxylase small subunit		RBCS class I
2	Phosphoglycerate kinase	PGK	
3	Glyceraldehyde-3-phosphate dehydrogenase subunit A	GAPDH	GAPDH A
	Glyceraldehyde-3-phosphate dehydrogenase subunit B		GAPDH B[a]
4	Triosephosphate isomerase	TPI	
5	Fructose-1,6-bisphosphate aldolase class I	FBA	FBA class I[b]
	Fructose-1,6-bisphosphate aldolase class II		FBA class II[c]
6	Fructose-1,6-bisphosphatase	FBP	FBP F-II
7	Sedoheptulose-1,7-bisphosphatase	SBP	
8	Transketolase	TKL	
9	Ribulose-phosphate 3-epimerase	RPE	
10	Ribose 5-phosphate isomerase	RPI	
11	Phosphoribulokinase	PRK	PRK class II
	Rubisco activase[d]	RCA	
	CP12[d]	CP12	

[a] Present in streptophyta.
[b] Present in red and green algae.
[c] Present only in glaucophyta.
[d] Regulatory proteins.

mRNA Midi Kit (Qiagen). Starter and normalized cDNA libraries were constructed as previously described (Reyes-Prieto et al., 2006). The cDNA clones (11,000) were submitted to 3′ single-pass sequencing. The *C. paradoxa* ESTs are available in dbEST (Reyes-Prieto et al., 2006).

2.2. Sequence handling

Sequences from the CC enzymes of the large (RBCL) and small (RBCS) subunits of RuBisCO, phosphoglycerate kinase (PGK), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), triosephosphate isomerase (TPI), fructose 1,6-bisphosphate aldolase class I (FBA I), fructose 1,6-bisphosphate aldolase class II (FBA II), fructose-1,6-bisphosphatase (FBP), sedoheptulose-1,7 bisphosphatase (SBP), transketolase (TKL), ribulose-phosphate 3-epimerase (RPE), ribose 5-phosphate isomerase (RPI), phosphoribulokinase (PRK), and other proteins involved in the CC such as RuBisCO activase (RCA) and the protein CP12 (CP12) were obtained and/or deduced (in the case of ESTs) from GenBank, dbEST, JGI (genome.jgi-psf.org, JGI), TBestDB (tbestdb.bcm.umontreal.ca/searches/welcome. php, TBestDB), *Cyanidoschyzon merolae* genome project (merolae.biol.s. u-tokyo.ac.jp, CGP), *Galdieria sulphuraria* genome project (genomics.msu.edu/galdieria, GGP), and the *Porphyra yezoensis* EST project (www.kazusa.or.jp/en/plant/porphyra/EST/,PGP). Homologous sequences were identified using a BLAST search with the *C. paradoxa* sequences as query. Multiple protein alignments wereprepared with Clustal X (Chenna et al., 2003) followed by manual refinement. These alignments are available upon request from D.B.

2.3. Sequence analysis

The GenBank accession numbers or identification codes (the corresponding source database is indicated in the brackets) of the sequences analyzed in Fig. 2 are listed below (see also Supplementary Table S1). Triose phosphate isomerase (TPI), *Taenia* (AAG21132), *Acetabularia* (AAL00002542), *Schistosoma* (AAP06170), *Encope* (AAT06239), *Ptychodera* (AAT06251), *Helicosporidium* (AAU93945), *Prototheca* (AAV65344), *Porphyra* (AAV65489), *Chlamydomonas* (AAV65490), *Haliotis* (AAZ30690), *Mus* (BAB27194), *Oryza* 2 (BAD34212), *Chlorella* 1 (BAE48228), *Chlorella* 2 (BAE48229), *Tetraodon* (CAF90849), *Pongo* (CAH91732), *Cyanidioschyzon* (CMQ172C) CGP, *Cyanophora* (EG946532), *Chaetomium* (EAQ91292), *Aedes* (EAT46226), *Glaucocystis* (EC125016), *Galdieria* 1 (Contig02102.g11, GGP), *Galdieria* 2 (Contig02102.g107, GGP), *Synechocystis* (NP_442075), *Thermosynechococcus* (NP_681756), *Secale* 2 (P46225), *Debraromyces* (Q6BMB8), *Candida* (Q9P940), *Magnaporthe* (XP_364060), *Pseudomonas* (YP_276315), *Thiomicrospira* (YP_391085), *Prochlorococcus* (YP_397467), *Nitrosospira* (YP_411784), *Synechococcus* (YP_473880), *Rhodoferax* (YP_522755), *Nostoc* (ZP_00111349), *Crocosphaera* (ZP_00518524), *Trichodesmium* (ZP_00675084), *Marinomonas* (ZP_01075692), *Oceanospirillum* (ZP_01168141), *Arabidopsis* 1 (AAA03449), *Zea* (AAB81110), *Secale* 1 (P46226), *Garcilaria* (P48492), *Oryza* 1 (XP_462797), *Spinacia* (P48496), *Arabidopsis* 2 (Q9SKP6), *Thalassiosira* 1, 2, 3 (jgi| Thaps3| 36462| e_gw1. 10.171.1, jgi| Thaps3| 40958| estExt_gwp_gw1.C_chr_60155, jgi| Thaps3| 30380| estExt_Genewise1.C_chr_180081), *Phaeodactylum* 1, 2, 3 (jgi| Phatr2| 18228 | estExt_gwp_ gw1.C_chr_20269, jgi| Phatr2| 50738| estExt_fgenesh1_pm. C_chr_20046, jgi| Phatr2| 54738| estExt_Phatr1_ua_kg.C_ chr_130035, JGI), *Pavlova* (EC175615), *Isochrysis* (EC145473), *Karenia* (CO062693), *Karlodinium* (EC158977); Fructose -1,6- bisphosphatase (FBP), *Lactuca* (AAF19790), *Vibrio* (AAF95685), *Cryptococcus* (AAW40656), *Ralstonia* (AAZ61806), *Bombyx* (ABF51284), *Oryza* 1 (ABF95158), *Oryza* 2 (Q9SDL8), *Arabidopsis* 1 (CAA41154), *Drosophila* (CAC35155), *Cyanidioschyzon* 1 (CMO245C, CGP), *Cyanidioschyzon* 2 (CMP129C, CGP), *Cyanophora* 1 (EG946632, EG945330), *Chlamydomonas* (jgi| Chlre3| 169775| fgenesh2_pg. C_scaffold_11000214, JGI), *Galdieria* 1 (Contig06901.g19, GGP), *Galdieria* 2 (Contig06101.g25, GGP), *Galdieria* 3 (Contig07501.g50, GGP), *Danio* (NP_001004008), *Synechocystis* (NP_441738), *Rattus* (NP_446168), *Thermosynechococcus* (NP_681331), *Spinacia* 1 (P22418), *Solanum* (P46276), *Arabidopsis* 2 (Q9MA79), *Oryza* (Q9SDL8), *Musa* (Q9XF47), *Photobacterium* (YP_128618), *Anabaena* (YP_322197), *Rhodospirillum* (YP_427490), *Escherichia* (YP_543771), *Crocosphaera* (ZP_00516972), *Trichodesmium* (ZP_00674296), *Flavobacteriales* (ZP_01105388), *Polaribacter* (ZP_01118035), *Halorhodospira* (ZP_01151819), *Psychromonas* (ZP_01214915), *Cyanidioschyzon* 3 (CMI196C, CGP), *Spinacia* 2 (O20252), *Arabidopsis* 3 (P46283), *Magnaporthe* (XP_367798), *Neurospora* (XP_957052), *Cyanophora* 2 (EG943965, EG946954,), *Yarrowia* (XP_500111), *Candida* (XP_710934), *Aspergillus* (XP_751698), *Ustilago* (XP_758850), *Cytophaga* (YP_679484), *Thalassiosira* 1, 2, 3 4 (jgi| Thaps3| 264556| thaps1_ua_kg.chr_18000078, jgi| Thaps3| 16295| gw1.5.288.1 jgi| Thaps3| 33021| e_gw1.3. 399.1, jgi| Thaps3| 261823| thaps1_ua_kg.chr_3000201), *Phaeodactylum* 1, 2, 3 4 (jgi| Phatr2| 31994| fgenesh1_pg. C_chr_1000595, jgi| Phatr2| 2793| gw1.4.138.1 jgi| Phatr2| 9359| e_gw1.1.334.1, jgi| Phatr2| 8744| e_gw1.1.338.1) *Guillardia* (AW342537), *Alexandrium* (CF947867), *Emiliania* (CX776165), *Isochrysis* (EC147059, EC147040) *Karenia* (ABF73028), *Amphidinium* (CF067225, CF064690); Sedoheptulopse -1,6- bisphosphatase (SBP),*Cyanophora* (EG943965, EG946954,), *Glaucocystis* (EC121758), *Griffithsia* (Q7XZ85), *Porphyra* (AV433409, PGP), *Galdieria* 1 (Contig09202.g18, GGP), *Cyanidioschyzon* 1 (CMI196C, CGP), *Physcomitrella* (BJ957298), *Arabidopsis* (P46283), *Spinacia* (O20252), *Oryza* (AAO22558), *Mesostigma* (DN263714), *Chlamydomonas* (P46284), *Galdieria* 2 (Contig07801.g56, GGP), *Magnaporthe* (XP_367798), *Neurospora* (XP_957052), *Cyanidioschyzon* 2 (CMT362C, CGP), *Wolinella* FBP (NP_906476), *Thiomicrospira* FBP (YP_393576), *Campylobacter*FBP (ZP_00370304).

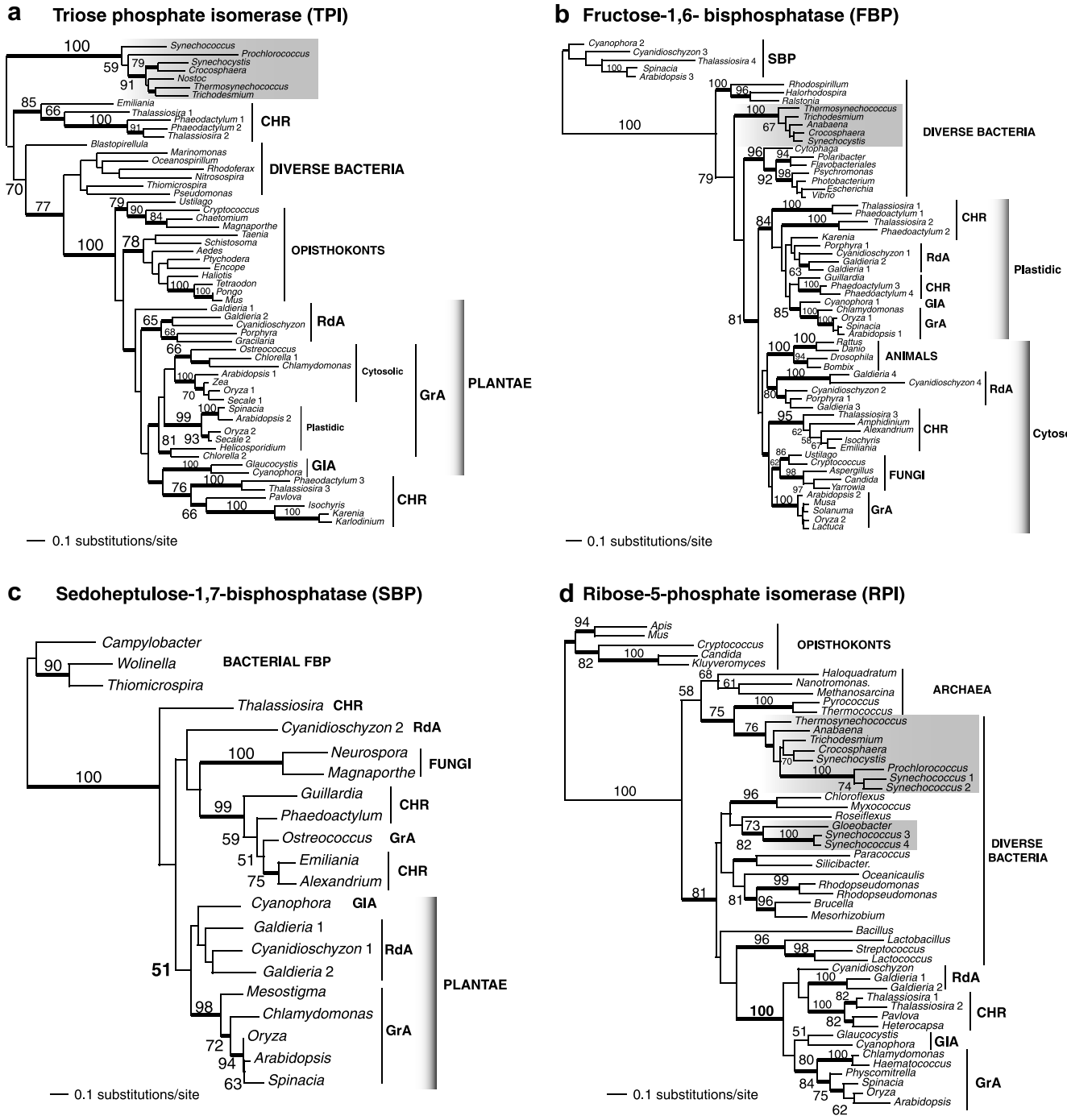

Fig. 2. Maximum likelihood trees of putative non-cyanobacterial CC enzymes with bootstrap values (when >50%) indicated at the nodes. The thick lines indicate branches with a Bayesian posterior probability >0.95. The branch lengths are proportional to the number of substitutions per site (see scales in the figure). The cyanobacteria are in shaded boxes and the Plantae are indicated with a gray bar. Plantae members are GrA (green algae and plants), GlA (glaucophytes), and RdA (red algae). Chromalveolates are denoted as CHR.

Thalassiosira (jgi| Thaps3| 261823| thaps1_ua_kg.chr_ 3000201, JGI), *Guillardia* (ABD51940), *Emiliania* (CX778868), *Alexandrium* (CK786334), *Ostreococcus* (CAL53197), *Phaeodactylum* (jgi| Phatr2| 9124| e_gw1. 1.382.1, JGI); Ribose 5-phosphate isomerase (RPI), *Arabidopsis* (AAF04905), *Spinacia* (AAL77589), *Gloeobacter* (BAC87971), *Physcomitrella* (BJ603408), *Cyanidioschyzon* (CMO291C, CGP), *Galdieria* 1 (Contig09202.g4, GGP), *Galdieria* 2 (Contig05201.g20, GGP), *Cyanophora* (EG944895, EG945227), *Haematococcus* (DV203830), *Glaucocystis* (EC124738), *Mus* (NP_033101), *Lactococcus* (NP_268390), *Synechocystis* (NP_442343), *Brucella* (NP_539891), *Thermosynechococcu*s (NP_682063), *Thermococcus* (Q5JH26), *Pyrococcus* (Q9V0L6), *Oryza* (XP_476826), *Apis* (XP_623160), *Methanosarcina* (YP_306636), *Anabaena* (YP_324984), *Natronomonas*

(YP_326053), *Synechococcus* 2 (YP_382366), *Synechococcus* 4 (YP_475715), *Synechococcus* 3 (YP_478248), *Rhodopseudomonas* (YP_486993), *Rhodopseudomonas* (YP_531949), *Silicibacter* (YP_613954), *Myxococcus* (YP_633632), *Haloquadratum* (YP_659340), *Mesorhizobium* (YP_674020), *Bacillus* (ZP_00239745), *Lactobacillus* (ZP_00384858), *Crocosphaera* (ZP_00517843), *Mesorhizobium* (ZP_0061 2727), *Paracoccus* (ZP_00631489), *Trichodesmium* (ZP_00672737), *Chloroflexus* (ZP_00768235), *Streptococcus* (ZP_00789039), *Oceanicaulis* (ZP_00957984), *Prochlorococcus* (ZP_01006556), *Synechococcus* 1 (ZP_01084139), *Roseiflexus* (ZP_01359698), *Chlamydomonas* (jgi|Chlre3|128644|estExt_gwp_1H.C_140217, JGI), *Thalassiosira* 1, 2 (jgi|Thaps3|32332|e_gw1.2.419.1, jgi|Thaps3|32252|e_gw1.2.426.1, JGI), *Pavlova* (ABA55587), *Heterocapsa* (AAW79354).

All *C. paradoxa* EST (Supplementary Table S2) and other sequences analyzed in this study are presented in the Supplementary data (Figs. S1–S5).

2.4. Phylogenetic analysis

The protein alignments were analyzed under maximum likelihood (ML). The best-fitting amino acid substitution model was selected according to the Akaike informational criterion with ProtTest using the default values (Abascal et al., 2005). The WAG (Whelan-Goldman) model with heterogeneous gamma rate distribution across sites (+Γ) and proportion of invariant sites (+I) was selected by ProtTest for all of the protein data sets. The WAG model parameter values were used under PHYML V2.4.3 (Guindon and Gascuel, 2003) for the ML tree searches. The stability of monophyletic groups was assessed using PHYML with 100 bootstrap replicates. In addition, the posterior probabilities of nodes in the PHYML trees were calculated using MrBayes (Ronquist and Huelsenbeck, 2003) with a Metropolis-coupled Markov chain Monte Carlo run from a random tree and sampled each 100 from a total of a million generations. One heated and three cold chains were simultaneously started and the best fitting substitution model for each protein set was used in the analyses. The initial 25% of sampled trees were discarded as ''burn in'' prior to construction of the consensus phylogeny.

2.5. Cellular target prediction

The probability of plastid targeting was assessed with neural network-based approaches with the Predotar V1.03 (urgi.infobiogen.fr/predotar) and TargetP V1.1 (www.cbs.dtu.dk/services/TargetP) servers using the default values.

3. Results

3.1. Cyanobacterial enzymes

Our ML and Bayesian phylogenetic analyses show that the Plantae CC enzymes have a cyanobacterial core. The phylogenetic trees of PGK, GAPDH, TKL, RPE and PRK with and without chromalveolate homologs (see Appendix A, Supplementary Figs. S1 and S5) recover a monophyletic group comprising Plantae and cyanobacteria (i.e., 79–100% bootstrap support). This is the expected topology for proteins derived from the ancestral cyanobacterial endosymbiont (Martin and Schnarrenberger, 1997). Some chromalveolate genes have a red algal origin (e.g., PGK) as expected under the chromalveolate hypothesis, others potentially have a green algal origin through horizontal gene transfer (HGT) or EGT from a green algal endosymbiont (e.g., PRK), and others have undergone gene replacement from a duplicated host homolog (GAPDH) (for details see Fast et al., 2001; Bhattacharya et al., 2004; Nosenko et al., 2006).

The GAPDH tree (Figs. S1 and S5) shows that the heterodimeric enzyme (GAPDH-A/GAPDH-B) evolved in land plants (see Petersen et al., 2006), with a single copy gene for plastid-targeted GAPDH in red, non-streptophyte green algae and glaucophytes. The multimeric enzyme Class I RuBisCO (RBCL and RBCS) has a cyanobacterial origin in green and glaucophyte algae (G-type). In contrast, in red algae, plastid encoded RuBisCO has a proteobacterial provenance (R-type) likely via HGT (Delwiche and Palmer, 1996) (Fig. S2). Other proteins involved in the regulation of the CC have an endosymbiotic origin. The chaperone-like RuBisCO activase (RCA) is of cyanobacterial provenance in green and glaucophyte algae (Fig. S2), but no homologs were identified in red algae. This result is consistent with the origin and distribution of RuBisCO in Plantae. The cyanobacterial-derived enzyme of green and, presumably, glaucophyte algae is activated via non-covalent interactions with RCA. In contrast, the proteobacterial-like RuBisCO of red algae may not require RCA-mediated activation, and the RCA gene was probably lost. RCA homologs are absent from red algal and proteobacterial genomes. An alternative scenario is that both G and R RuBisCO types were present in the plastid ancestor and these systems were differentially lost in Plantae (Martin and Schnarrenberger, 1997). Both types of RuBisCO are apparently present in some proteobacteria but do not co-occur in cyanobacteria (Uchino and Yokota, 2003). Additionally, protein CP12, a presumed linker mediator of PRK-GAPDH aggregation in the inactive complex (Wedel et al., 1997) is present only in Plantae and cyanobacteria (Fig. S5).

3.2. Fructose-1,6-bisphospate aldolase distribution in the Plantae

FBA has a complex evolutionary history. Two unrelated plastid targeted enzymes (known as class I and II, Marsh and Lebherz, 1992) are present in Plantae. Plastid FBA class I (FBA-I) was not found in the *C. paradoxa* phylogenomic analyses, however a cytosolic-like FBA-I is present in the glaucophyte *Glaucocystis nostochinearum* and it groups with the red algae, green algae and animal homologs (however with no bootstrap support, Fig. S3). The

plastid FBA-I version is present in red and green algae, and in our phylogeny forms a moderately well supported (86%) monophyletic group (Fig. S3). The plastid FBA-I of red algae groups with cyanobacterial homologs (100%), suggesting the endosymbiotic origin of this enzyme in red algae. Our analysis does not resolve the origin of the green algal plastid sequence, but suggests a cyanobacterial provenance (Fig. S3). Other studies suggested the origin of plastid FBA-I from a duplication of the gene encoding the eukaryotic cytosolic enzyme (Schnarrenberger et al., 1994; Rogers and Keeling, 2004). There are two homolog variants of FBA-II, termed Type A and Type B. The plastid enzyme uniquely present in glaucophytes is the FBA-II Type B (FBA-IIB) and likely has a cyanobacterial origin (71%, Fig. S4B; Nickol et al., 2000). A sequence that is distantly related to FBA-IIB and is comprised of an apparent gene fusion of tagatose-1, 6-bisphosphate aldolase and 2 tandemly repeated copies of 3-hydroxyisobutyrate dehydrogenase is also present in land plants (TBP aldolases, Fig. S4). A non-cyanobacterial FBA-IIA was identified in glaucophytes (possibly cytosolic) and is related to homologs from red algae, fungi, and non-plastid-targeted proteins in chromalveolates (Fig. S4). The presence of both FBA II type sequences (cytosolic-like type A and plastid type B) in *C. paradoxa* is consistent with previous detailed biochemical analyses (Gross et al., 1994). These trees suggest that both plastid FBA class I and class II genes are likely of cyanobacterial origin in Plantae, however glaucophytes (Fig. S4B) have the typical class II gene of plastid function, whereas red and green algae (Fig. S3) retain the class I gene. Both plastid FBA classes are bifunctional for the reversible condensation of fructose 1,6-bisphosphate and sedoheptulose 1,7-bisphosphate (Flechner et al., 1999). Remarkably, the cyanobacterium *Synechococcus* sp. BL107 possesses both classes of FBA genes (accession numbers EAU72138 and EAU72139). Moreover, experimental data show that enzymes of both FBA classes are active in the cyanobacterium *Synechocystis* sp. PCC 6803, with the FBA-II enzyme being used in the CC (Nakahara et al., 2003).

3.3. Non-cyanobacterial replacements

The remainder of the CC enzymes originated from non-cyanobacterial sources. These CC proteins (i.e., TPI, FBP, SBP, and RPI) may therefore not have arisen from the plastid endosymbiont but are shared among the Plantae. This is consistent with the ancient recruitment of these genes in the Plantae common ancestor. In our analysis of TPI (Fig. 2a), we recover a monophyletic Plantae (albeit with no bootstrap support) within a well-supported clade of eukaryotes (100%) that is distantly related to cyanobacteria. The association of chromalveolates with the Plantae clade is explained by secondary EGT in these taxa (e.g., Nosenko et al., 2006) and does not contradict the notion of Plantae monophyly. This topology suggests that the gene encoding plastid targeted TPI in the Plantae likely evolved from the eukaryotic homolog via gene duplication and acquisition of a plastid import signal, as previously suggested for land plants (Henze et al., 1994; Schmidt et al., 1995). Bioinformatic targeting signal predictions for plastid TPI in green algae and plants using the computer programs Predotar and TargetP provide robust support for a plastid location of these enzymes (probability = 0.70–0.98, respectively), consistent with Schmidt et al. (1995). In contrast to TPI, the FBP (F-II type, Tamoi et al., 1998) tree (Fig. 2b) shows moderate bootstrap support (72%) for the monophyly of the plastid-targeted enzyme in Plantae and chromalveolates as sister to the branch containing the cytosolic homologs. This topology suggests that plastidic FBP may have arisen in the Plantae algal ancestor via an ancient gene duplication involving the host cytosolic isoform. This is a clade of distinct eukaryotic origin (89%) that is not closely related to cyanobacterial homologs. SBP homologs (which share limited sequence similarity with FBP) were not detected in available cyanobacterial genomes and the closest relatives of the Plantae genes were some fungal and chromalveolate hypothetical proteins and FBP sequences in diverse bacteria (3 representatives are shown in Fig. 2b). The origin of plastid SBP has been proposed from a common ancestral eukaryotic enzyme in land plants (possibly mitochondrial derived; Martin et al., 1996) and in red algae (Matsuzaki et al., 2004). The results of the FBP and SBP analyses indicate that these related proteins replaced the original cyanobacterial FBP/SBP (F-I type, with dual hydrolytic activity for fructose 1,6-bisphosphate and sedoheptulose 1,7-bisphosphate, Tamoi et al., 1998) in the last common ancestor of Plantae. And finally, the RPI tree (Fig. 2d) provides another example of a non-cyanobacterial CC enzyme that was recruited in the Plantae ancestor. In this case, the branch uniting the Plantae is robustly supported (100%) with the gene arising from an as yet unidentified lineage, although due to a lack of support for the backbone of the tree, a cyanobacterial origin of RPI in Plantae cannot yet be unambiguously excluded. These phylogenetic data indicate that at least three likely host-derived enzyme-encoding genes (TPI, FBP, SBP) replaced the ancestral cyanobacterial CC enzymes in the single common ancestor of red, green, and glaucophyte algae. RPI gene origin remains inconclusive, although a single origin in the Plantae ancestor is robustly supported by our analysis.

4. Discussion

4.1. The evolution of RuBIsCO and FBA in Plantae

Current evidence shows the presence of only G-type RuBIsCO in cyanobacteria (out of 19 complete and 10 partial genomes that were analyzed), but it is possible given extensive HGT among prokaryotes, that the plastid cyanobacterial ancestor also contained the R-type RuBIsCO prior to endosymbiosis (Martin and Schnarrenberger,

1997). Under this scenario, differential gene losses in green + glaucophyte versus red algae may explain the current distribution in Plantae. The presence of RuBIsCO genes in the plastid genome of most Plantae (except for nuclear-encoded RBCS in land plants) combined with the rarity of HGT to this organelle (for exception, see Rice and Palmer, 2006), favor the endosymbiotic origin of both RuBisCO types in the Plantae ancestor. A similar scenario may explain the cyanobacterial origin and distribution of the plastid-targeted FBA enzymes in Plantae. Both plastid FBA classes could have been encoded in the cyanobacterial endosymbiont genome (e.g., *Synechococcus* sp. BL107), and after plastid establishment and the split of the main Plantae groups, red and green algae retained FBA-I for the CC, whereas glaucophytes retained FBA-II for this function.

4.2. The monophyly of Plantae

The cyanobacterial provenance of the CC core enzymes (PGK, GAPDH, TKL, RPE, and PRK) is consistent with the single endosymbiotic origin of the pathway in the common ancestor of Plantae.

However, an important piece of evidence to support this scenario comes from the phylogeny of enzymes that were recruited post-endosymbiosis to give rise to the eukaryotic version of the CC. It is highly unlikely that multiple independent acquisitions of these non-cyanobacterial genes (i.e., encoding TPI, FBP, SBP, and possibly RPI) in the different Plantae lineages would produce the observed phylogenetic pattern. Under this scenario the three lineages would have independently recruited the *same* 4 CC enzymes from the *same or closely related donor* thereby supporting group monophyly by virtue of shared multiple horizontal gene transfer events and not vertical inheritance. Although this scenario is possible it is far less parsimonious than the single origin of the 4 non-cyanobacterial genes in the Plantae ancestor followed by vertical inheritance in descendants. A different type of scenario is that the endosymbiont contained these genes that are both of eukaryotic and prokaryotic non-cyanobacterial origin (probably acquired by HGT; Martin, 1999) and Plantae acquired these through the canonical EGT process. This hypothesis implies that the plastid descends from a cyanobacterium with a radically different genomic complement than in studied taxa (Martin and Schnarrenberger, 1997). In either case, our data provide convincing evidence that extant Plantae lineages are descendants of a single ancestor.

A recent analysis of 143 nuclear genes (ca. 30,000 amino acid positions) provides strong support for Plantae host monophyly (Rodriguez-Ezpeleta et al., 2005). However, a contentious issue in Plantae evolution is the order of divergence among its constituent lineages. The TPI, FBP, SBP, RPI (excluding chromalveolate sequences) RPE, and GAPDH trees suggest that red algae (without bootstrap support) are the earliest diverging Plantae (see also Rodriguez-Ezpeleta et al., 2005). This result is inconsistent with many plastid gene trees that identify glaucophytes as the earliest divergence (Rodriguez-Ezpeleta et al., 2005; Yoon et al., 2006). To address this issue we require far more sequence data from glaucophytes and other basal algal Plantae members (e.g., prasinophyte green algae and mesophilic bangiophyte red algae). This is a challenging task with respect to glaucophytes because only a handful of taxa exist in culture collections.

In summary, enzyme substitutions have played an important role in plastid establishment and provide a fertile source of data for understanding algal and plant evolution. Studies of other algae that have undergone secondary or tertiary endosymbiosis (i.e., engulfment of an existing alga) such as in chromalveolates (Nosenko et al., 2006) show an even more complex pattern of gene recruitment for the CC (e.g., see Fig. S1). Interestingly, recent phylogenetic analyses of the plastid shikimate pathway (biosynthesis of chorismate) suggest that 5 of the 7 ancestral cyanobacterial enzymes were replaced (potentially via HGT) "before the divergence of green plants and red algae" (Richards et al., 2006). The land plants and red algae apparently share only 2 enzymes of cyanobacterial origin. These results are also consistent with the monophyly of Plantae, although glaucophytes were not included in this study. Another piece of data that supports Plantae monophyly is the phylogeny of plastid targeted metabolite antiporters. These proteins regulate metabolite (monosaccharides, phosphoenolpyruvate) movement across the plastid membranes (Weber et al., 2006) and phylogenetic analysis supports the origin of red and green algal plastid antiporters through the co-option and diversification of a single host gene encoding an existing endomembrane translocator (Weber et al., 2006). The intricacy of these gene acquisitions and recruitments that underlie plastid establishment begin to provide an explanation for the singular nature of primary plastid endosymbiosis that laid the foundation for eukaryotic biodiversity (Bhattacharya et al., 2004).

Acknowledgments

This work was supported by grants from the National Science Foundation and the National Aeronautics and Space Administration awarded to D.B. (EF 04-31117, NNG04GM17G, respectively).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2007.02.026.

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105.

Adl, S.M., Simpson, A.G., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R., Bowser, S.S., Brugerolle, G., Fensome, R.A., Fredericq, S., James, T.Y., Karpov, S., Kugrens, P., Krug, J., Lane, C.E., Lewis, L.A., Lodge, J., Lynn, D.H., Mann, D.G., McCourt, R.M., Mendoza, L., Moestrup, O., Mozley-Standridge, S.E., Nerad, T.A., Shearer, C.A., Smirnov, A.V., Spiegel, F.W., Taylor, M.F., 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J. Eukaryot. Microbiol. 52, 399–451.

Bhattacharya, D., Yoon, H.S., Hackett, J.D., 2004. Photosynthetic eukaryotes unite: endosymbiosis connects the dots. BioEssays 26, 50–60.

Calvin, M., Benson, A.A., 1948. The path of carbon in photosynthesis. Science 107, 476–480.

Cavalier-Smith, T., 1981. Eukaryote kingdoms: seven or nine? Biosystems 14, 461–481.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., Thompson, J.D., 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 31, 3497–3500.

Delwiche, C.F., Palmer, J.D., 1996. Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. Mol. Biol. Evol. 13, 873–882.

Fast, N.M., Kissinger, J.C., Roos, D.S., Keeling, P.J., 2001. Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. Mol. Biol. Evol. 18, 418–426.

Flechner, A., Gross, W., Martin, W.F., Schnarrenberger, C., 1999. Chloroplast class I and class II aldolases are bifunctional for fructose-1,6-biphosphate and sedoheptulose-1,7-biphosphate cleavage in the Calvin cycle. FEBS Lett. 447, 200–202.

Gross, W., Bayer, M.G., Schnarrenberger, C., Gebhart, U.B., Maier, T.L., Schenk, H., 1994. Two distinct aldolases of class II type in the cyanoplasts and in the cytosol of the alga *Cyanophora paradoxa*. Plant Physiol. 105, 1393–1398.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Henze, K., Schnarrenberger, C., Kellermann, J., Martin, W., 1994. Chloroplast and cytosolic triosephosphate isomerases from spinach: purification, microsequencing and cDNA cloning of the chloroplast enzyme. Plant Mol. Biol. 26, 1961–1973.

Marsh, J.J., Lebherz, H.G., 1992. Fructose-bisphosphate aldolases: an evolutionary history. Trends Biochem. Sci. 17, 110–113.

Martin, W., 1999. Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. Bioessays 21, 99–104.

Martin, W., Schnarrenberger, C., 1997. The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. Curr. Genet. 32, 1–18.

Martin, W., Mustafa, A.Z., Henze, K., Schnarrenberger, C., 1996. Higher-plant chloroplast and cytosolic fructose-1,6-bisphosphatase isoenzymes: origins via duplication rather than prokaryote-eukaryote divergence. Plant Mol. Biol. 32, 485–491.

Matsuzaki, M., Misumi, O., Shin-I, T., Maruyama, S., Takahara, M., Miyagishima, S.Y., Mori, T., Nishida, K., Yagisawa, F., Nishida, K., Yoshida, Y., Nishimura, Y., Nakao, S., Kobayashi, T., Momoyama, Y., Higashiyama, T., Minoda, A., Sano, M., Nomoto, H., Oishi, K., Hayashi, H., Ohta, F., Nishizaka, S., Haga, S., Miura, S., Morishita, T., Kabeya, Y., Terasawa, K., Suzuki, Y., Ishii, Y., Asakawa, S., Takano, H., Ohta, N., Kuroiwa, H., Tanaka, K., Shimizu, N., Sugano, S., Sato, N., Nozaki, H., Ogasawara, N., Kohara, Y., Kuroiwa, T., 2004. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. Nature 428, 653–657.

Nakahara, K., Yamamoto, H., Miyake, C., Yokota, A., 2003. Purification and characterization of class-I and class-II fructose-1,6-bisphosphate aldolases from the cyanobacterium *Synechocystis* sp. PCC 6803. Plant Cell Physiol. 44, 326–333.

Nickol, A.A., Muller, N.E., Bausenwein, U., Bayer, M.G., Maier, T.L., Schenk, H.E., 2000. *Cyanophora paradoxa*: nucleotide sequence and phylogeny of the nucleus encoded muroplast fructose-1,6-bisphosphate aldolase. Z. Naturforsch. 55, 991–1003.

Nosenko, T., Lidie, K.L., Van Dolah, F.M., Lindquist, E., Cheng, J.F., Bhattacharya, D., 2006. Chimeric plastid proteome in the Florida "red tide" dinoflagellate *Karenia brevis*. Mol. Biol. Evol. 23, 2026–2038.

Nozaki, H., Matsuzaki, M., Takahara, M., Misumi, O., Kuroiwa, H., Hasegawa, M., Shin-i, T., Kohara, Y., Ogasawara, N., Kuroiwa, T., 2003. The phylogenetic position of red algae revealed by multiple nuclear genes from mitochondria-containing eukaryotes and an alternative hypothesis on the origin of plastids. J. Mol. Evol. 56, 485–497.

Petersen, J., Teich, R., Becker, B., Cerff, R., Brinkmann, H., 2006. The GapA/B gene duplication marks the origin of Streptophyta (charophytes and land plants). Mol. Biol. Evol. 23, 1109–1118.

Reyes-Prieto, A., Hackett, J.D., Bonaldo, M.F., Soares, M.B., Bhattacharya, D., 2006. Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. Curr. Biol. 16, 2320–2325.

Rice, D.W., Palmer, J.D. 2006. An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. BMC Biol. 4, 31.

Richards, T.A., Dacks, J.B., Campbell, S.A., Blanchard, J.L., Foster, P.G., McLeod, R., Roberts, C.W., 2006. Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. Eukaryot. Cell 5, 1517–1531.

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

Rogers, M., Keeling, P.J., 2004. Lateral transfer and recompartmentalization of Calvin cycle enzymes of plants and algae. J. Mol. Evol. 58, 367–375.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Schmidt, M., Svendsen, I., Feierabend, J., 1995. Analysis of the primary structure of the chloroplast isozyme of triosephosphate isomerase from rye leaves by protein and cDNA sequencing indicates a eukaryotic origin of its gene. Biochim. Biophys. Acta 1261, 257–264.

Schnarrenberger, C., Pelzer-Reith, B., Yatsuki, H., Freund, S., Jacobshagen, S., Hori, K., 1994. Expression and sequence of the only detectable aldolase in *Chlamydomonas reinhardtii*. Arch. Biochem. Biophys. 313, 173–178.

Stiller, J.W., Harrell, L., 2005. The largest subunit of RNA polymerase II from the Glaucocystophyta: functional constraint and short-branch exclusion in deep eukaryotic phylogeny. BMC Evol. Biol. 5, 71.

Tamoi, M., Murakami, A., Takeda, T., Shigeoka, S., 1998. Acquisition of a new type of fructose-1,6-bisphosphatase with resistance to hydrogen peroxide in cyanobacteria: molecular characterization of the enzyme from *Synechocystis* PCC 6803. Biochim. Biophys. Acta 1383, 232–244.

Uchino, Y., Yokota, A., 2003. "Green-like" and "red-like" RubisCO cbbL genes in *Rhodobacter azotoformans*. Mol. Biol. Evol. 20, 821–830.

Weber, A., Linka, K., Bhattacharya, D., 2006. Single, ancient origin of a plastid metabolite translocator family in Plantae from an endomembrane-derived ancestor. Eukaryot. Cell 5, 609–612.

Wedel, N., Soll, J., Paap, B.K., 1997. CP12 provides a new mode of light regulation of Calvin cycle activity in higher plants. Proc. Natl. Acad. Sci. USA 94, 10479–10484.

Yoon, H.S., Müller, K.M., Sheath, R.G., Ott, F.D., Bhattacharya, D., 2006. Defining the major lineages of red algae (Rhodophyta). J. Phycol. 42, 482–492.