Published in : Protist (2007), vol. 158, pp. 263-276.
Status : Postprint (Author's version)

Origin and Distribution of Calvin Cycle Fructose and Sedoheptulose Bisphosphatases in Plantae and Complex Algae: A Single Secondary Origin of Complex Red Plastids and Subsequent Propagation via Tertiary Endosymbioses

René Teich[a], Stefan Zauner[b], Denis Baurain[c,d], Henner Brinkmann[c], and Jörn Petersen[a]

[a]*Institut für Genetik, Technische Universität Braunschweig, Spielmannstraße 7, D-38106 Braunschweig, Germany*

[b]*Cell Biology, Philipps-University Marburg, D-35032 Marburg, Germany*

[c]*Département de Biochimie, Université de Montréal, C.P. 6128, Canada*

[d]*Département des Sciences de la Vie, Université de Liège, B22, B-4000 Liège, Belgium*

Abstract

Sedoheptulose-1,7-bisphosphatase (SBPase) and fructose-1,6-bisphosphatase (FBPase) are essential nuclear-encoded enzymes involved in land plant Calvin cycle and gluconeogenesis. In this study, we cloned seven *SBP* and seven *FBP* cDNAs/genes and established sequences from all lineages of photosynthetic eukaryotes, in order to investigate their origin and evolution. Our data are best explained by a single recruitment of plastid-targeted *SBP* in Plantae after primary endosymbiosis and a further distribution to algae with complex plastids. While *SBP* is universally found in photosynthetic lineages, its presence in apicomplexa, ciliates, trypanosomes, and ascomycetes is surprising given that no metabolic function beyond the one in the plastid Calvin cycle is described so far. Sequences of haptophytes, cryptophytes, diatoms, and peridinin-containing dinoflagellates (complex red lineage) strongly group together in the *SBP* tree and the same assemblage is recovered for plastid-targeted *FBP* sequences, although this is less supported. Both *SBP* and plastid-targeted *FBP* are most likely of red algal origin. Including phosphoribulokinase, fructose bisphosphate aldolase, and glyceraldehyde-3-phosphate dehydrogenase, a total of five independent plastid-related nuclear-encoded markers support a common origin of all complex rhodoplasts via a single secondary endosymbiosis event. However, plastid phylogenies are incongruent with those of the host cell, as illustrated by the cytosolic FBP isoenzyme. These results are discussed in the context of Cavalier-Smith's far-reaching chromalveolate hypothesis. In our opinion, a more plausible evolutionary scenario would be the establishment of a unique secondary rhodoplast and its subsequent spread via tertiary endosymbioses.

Keywords : origin of algae ; primary metabolism ; eukaryote-to-eukaryote endosymbioses ; plastid evolution ; host cell evolution ; chromalveolate hypothesis.

Introduction

The insights of the Russian botanist Constantin Mereschkowsky who already described the principles of a cyanobacterial origin of plastids in 1905 were unfortunately forgotten for decades (Martin and Kowallik 1999; Mereschkowsky 1905). Mereschkowsky explains the benefit of photoautotrophy in his allegory of a palm tree and a hungry and blood-thirsty lion by the proposal that the presence of "green slaves" (plastids) within each cell of the animal would result in a calm predator lying in the sun without any interest in his natural prey. The endosymbiotic theory was rediscovered in the 1970s and the endosymbiotic gene transfer to the nucleus is nowadays considered as a major driving force of eukaryotic evolution (Bhattacharya et al. 2004; Gray et al. 1999; Margulis 1975).

With the notable exception of the testate amoeba *Paulinella chromatophora* (Marin et al. 2005; Yoon et al. 2006), eukaryotic photosynthesis originated from a single primary endosymbiosis event with a cyanobacterium (Bhattacharya and Medlin 1995; Douglas 1998). Their direct descendants, the Plantae or Archaeplastida contain primary plastids surrounded by two membranes and consist of green plants (green algae and land plants), rhodophytes, and glaucophytes (Adl et al. 2005; Rodriguez-Ezpeleta et al. 2005). The great algal diversity (especially of the marine phytoplankton) can mainly be traced back to secondary and tertiary endosymbioses, where eukaryotic host cells engulfed photosynthetic eukaryotes and reduced them to complex plastids (Stoebe and Maier 2002). Cryptophyte and chlorarachniophyte plastids still harbor highly reduced nucleomorphs, which

are respectively the remnants of the nucleus of the red and green algal endosymbionts (Douglas et al. 2001; Gilson and McFadden 1997, 2002). With respect to their complex plastids, euglenophytes and chlorarachniophytes constitute the green lineage, whereas haptophytes, heterokonts, cryptophytes, and peridinin-containing dinoflagellates are representatives of the red lineage (Archibald et al. 2003; Delwiche 1999; Li et al. 2006). The affiliation of the reduced plastids of apicomplexa, including the malaria parasite *Plasmodium falciparum,* is unclear since both green and red algal ancestries for the apicoplast are controversially discussed (Cai et al. 2003; Funes et al. 2002, 2003; Waller et al. 2003). Among complex green plastids, euglenophytes are classified together with diplonemids and kinetoplastids as Euglenozoa based on the host cell (Dooijes et al. 2000; Maslov et al. 1999), while chlorarachniophytes are part of the cercomonads (Bhattacharya et al. 1995; McFadden et al. 1994). The distinct origin of these host lineages provides strong evidence for two independent secondary endosymbioses with green algae; this assumption was recently confirmed at the plastid level (Rogers et al. 2007). In contrast, a common ancestry of stramenopiles (oomycetes, heterokont algae), haptophytes, cryptophytes, and alveolates (ciliates, apicomplexa, dinoflagellates) via a single secondary endosymbiosis with a rhodophyte was recently proposed and is known as the "chromalveolate hypothesis" (Cavalier-Smith 1999). This hypothesis is supported by plastid phylogenies in which the "chromalveolate" lineages are monophyletic and nested within red algae (Bachvaroff et al. 2005; Yoon et al. 2002). At the nuclear level, three genes of the Calvin cycle, i.e. glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), class II fructose-1,6-bisphosphate aldolase (*FBA-II*), and phosphoribulokinase (*PRK*) also support the monophyly of "chromalveolate plastids", although none of these nuclear-encoded genes displays the expected red algal affiliation (Harper and Keeling 2003; Patron et al. 2004; Petersen et al. 2006). Moreover, the monophyly of all "chromalveolate host cells" is at best elusive (Gray et al. 1998; Li et al. 2006; Sánchez Puerta et al. 2004; Van de Peer and De Wachter1997).

Sedoheptulose-1,7-bisphosphatase (SBP) and fructose-1,6-bisphosphatase (FBP) are two other enzymes essential for the plastid Calvin cycle of land plants (Raines 2003), where they are subjected to ferredoxin/thioredoxin-dependent light regulation (Buchanan 1980). An additional cytosolic FBPase isoenzyme, which is generally present in eukaryotes, catalyzes an irreversible step of gluconeogenesis (Plaxton 1996). SBP is especially relevant to crop breeding research since it has been shown to represent the crucial metabolic bottleneck of the Calvin cycle with major influence on the control of the whole pathway (reviewed by Raines 2003). Furthermore, overexpression of SBPase in tobacco significantly improves carbon fixation and growth (Lefebvre et al. 2005; Miyagawa et al. 2001).

Eukaryotic *SBP* and *FBP* are distantly related and nuclear-encoded genes that are both of bacterial ancestry (Martin et al. 1996). Chloroplast SBPase and FBPase of land plants have replaced the bispecific enzyme of cyanobacteria that either uses sedoheptulose-1,7-bisphosphate or fructose-1,6-bisphosphate as a substrate (Gerbling et al. 1986). Thus, *SBP* may represent a promising test system for open questions of eukaryotic evolution including the examination of the relationships of complex algae that originated via eukaryote-to-eukaryote endosymbioses (see above). Given its presumably specific function within the Calvin cycle, the presence of a *SBP* gene in the agent of sleeping sickness, *Trypanosoma brucei*, has been regarded as a strong argument for a photosynthetic ancestry of the Euglenozoa (Hannaert et al. 2003; Martin and Borst 2003; Rogers and Keeling 2004).

In this study, we present comprehensive analyses of *SBP* and *FBP* sequences from all lineages of photosynthetic eukaryotes and investigate their distribution. Our results challenge the predictions of the chromalveolate hypothesis and are best interpreted as evidence for independent tertiary endosymbioses at the origin of the extant diversity of algae with complex red plastids.

Results

Isolation and Characterization of SBP and FBP Sequences

In this study, we established altogether 14 new *SBP* and *FBP* sequences from the cryptophyte *Guillardia theta,* the dinophyte *Lingulodinium polyedrum*, the haptophyte *Prymnesium parvum*, the red algae *Chondrus crispus* and *Porphyra yezoensis,* the euglenophyte *Euglena gracilis,* and the liverwort *Marchantia polymorpha* (see supplementary Table S1 for details). Except for the genomic *SBP* clone of *Prymnesium*, all newly established sequences are cDNA clones. The *SBP* from *Euglena* is encoded as a biprotein precursor (815 amino acids (aa) in total) that is separated by a hinge region of approximately 30 aa. Both subunits exhibit an identity of 75% and probably originated from a recent gene duplication and fusion event (see below). Since the native SBP enzyme in land plants is a dimer of two identical subunits (Martin and Schnarrenberger 1997), the biprotein of *Euglena* could be the active holoenzyme; however, post-translational processing cannot be ruled out.

Additional *SBP* and *FBP* sequences were retrieved from the public databases through extensive BLAST searches. Raw sequence data (e.g., NCBI Traces) were merged into contigs, and subsequently the deduced aa sequences were used for phylogenetic analyses (see supplementary Table S1). Surprisingly, the green algal genomes of *Chlamydomonas reinhardtii* and *Volvox carteri* contain only the plastid-targeted *FBP* gene (see below), indicating a secondary loss of the cytosolic isoform. This conclusion is compatible with previous biochemical and molecular studies that have shown that cytosolic glycolysis/gluconeogenesis is lacking in chlorophytes (Schnarrenberger et al. 1990, 1994). In contrast, genomes of diatoms and land plants contain three *FBP* genes while rhodophytes harbor up to four different copies.

Phylogenetic Analyses of FBP Sequences

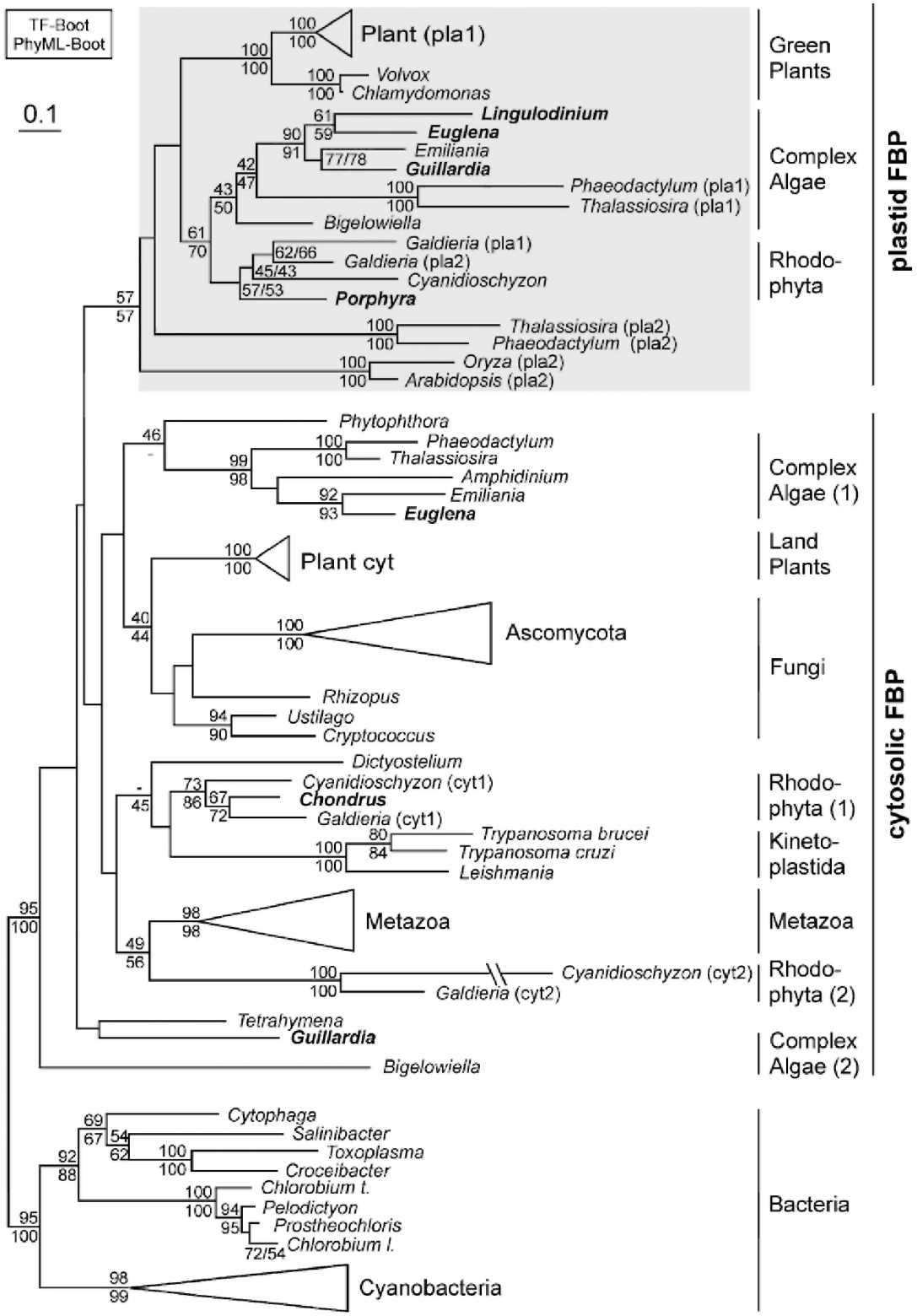
Figure 1 shows the maximum likelihood tree (TreeFinder) of 94 eukaryotic and 12 bacterial *FBP* sequences based on 231 aa positions. The tree is rooted with the bacterial sequences and contains a broad variety of *FBP* sequences from photosynthetic eukaryotes: cytosolic and plastid genes from green plants (chlorophytes and streptophytes), rhodophytes and representatives of all lineages of complex algae including a cryptophyte (*Guillardia*), a haptophyte (*Emiliania*), two heterokonts (*Thalassiosira*, *Phaeodactylum*), two dinoflagellates (*Lingulodinium*, *Amphidinium*), a chlorarachniophyte (*Bigelowiella*), and a euglenophyte (*Euglena*). Altogether 53 cytosolic sequences of animals, fungi, kinetoplastids, the slime mould *Dictyostelium,* the oomycete *Phytophthora,* the ciliate *Tetrahymena,* and the apicomplexan parasite *Toxoplasma* were included.

With the exception of the *FBP* from *Toxoplasma* that clearly groups among the bacterial sequences (Rogers and Keeling 2004), all eukaryotic *FBP* sequences are clustered in a clade sister to the Bacteria. Despite comprehensive database searches, no archaeal homolog was found, which suggests a common bacterial origin for both the cytosolic and plastid-targeted isoenzymes involved in gluconeogenesis and the Calvin cycle respectively. Plastid-targeted sequences including representatives of all complex lineages form a distinct subtree that originated from an ancient gene duplication early in eukaryotic evolution (Fig. 1; shaded in gray). The initially low support for this clade reaches nearly maximum values (57/57% versus 97/100%) after removing the fast-evolving additional plastid homologs (pla2) of plants and diatoms (supplementary Fig. S1a and 1b). The remaining *FBP* sequences (pla1) comprise a highly supported subtree of green plants (chlorophytes and streptophytes; 100/100% BP) branching as a sister group to a clade composed of rhodophytes and an assemblage of complex algae (55/61%; supplementary Fig. S1b). The latter subtree (53/61%; supplementary Fig. S1b) contains species of all complex lineages with red plastids (haptophytes, cryptophytes, diatoms, and dinoflagellates), but it also includes *Euglena* and *Bigelowiella*, which acquired their plastids independently via secondary endosymbioses with chlorophytes (Delwiche 1999; Rogers et al. 2007). Thus, both complex "green algae" appear to have recruited their plastid-targeted *FBP* sequences via horizontal gene transfer (HGT) in a non-endosymbiotic context. The diatom sequences (pla1) have very long branches corresponding to an evolutionary rate roughly two times higher than their cytosolic counterpart (see Fig. 1). Long branches often lead to a reduced statistical support (compare supplementary Fig. S1b and c), but in the presence of a distant outgroup or a fast-evolving ingroup, they may furthermore yield an artifactual topology (long-branch attraction artifact [LBA; Brinkmann et al. 2005; Felsenstein 1978]). Accordingly, the statistical support for the plastid subtree containing all lineages of complex algae increases after the removal of divergent ascomycete and distantly related bacterial outgroup sequences (compare supplementary Fig. S1b and d [F]; 53/61% versus 64/74%) while the rhodophycean affiliation is also strengthened (55/61 % versus 86/92%).

The cytosolic *FBP* sequences are divided into several distinct lineages of uncertain branching order. Whereas the affiliation of three basal sequences *(Bigelowiella, Guillardia,* and *Tetrahymena*) is unresolved, several groups including land plants, metazoa, ascomycetes, and kinetoplastids (inc. trypanosomes) are well supported (Fig. 1) and their ingroup topology reflects the organismal evolution (data not shown). The lack of support for the monophyly of fungi is probably caused by the conspicuously long branch separating the ascomycetes (e.g. *Saccharomyces*) from the basidiomycetes and zygomycetes (e.g. *Ustilago*, *Rhizopus*). In contrast, a well-supported subtree (98/99%) with a very long basal branch contains cytosolic *FBP* sequences from complex algae (1) including those of diatoms, *Amphidinium*, *Emiliania*, and *Euglena.* However, this ensemble does not contain the cryptophycean sequence of *Guillardia* and its phylogeny is inconsistent with the accepted organismal relationships. In particular, the grouping of *Euglena* and the haptophyte *Emiliania* is surprising and likely results from relatively recent HGT. A common ancestry of the whole subtree is strongly supported by two specific insertions of two and five amino acids (see supplementary Fig. S3). The presence of the first insertion in the *FBP* gene of the heterotrophic oomycete *Phytophthora* (only one additional amino acid [K]) suggests that the otherwise weakly supported position sister to the complex algae may be correct (Fig. 1). The alveolate superensemble (ciliates, apicomplexa, and dinoflagellates) cannot be recovered due to the presence of paralogous genes in *Amphidinium* (dinoflagellate) and *Toxoplasma* (apicomplexa) that harbors a xenologous gene related to

that of the bacterium *Croceibacter* (Fig. 1).

Figure 1. *Maximum likelihood tree inferred using TreeFinder (TF) with a WAG+F+Γ4 model based on 106 FBP sequences and 231 amino acid positions.* The phylogenetic tree is rooted with a bacterial outgroup. Sequences established in this study are in **boldface** and the subtree of the plastid sequences is shaded in gray. The horizontal length of the triangles is equivalent to the average branch length of the sequences, their numbers are as follows: Plant pla1 (6), Plant cyt (6), Ascomycota (15), Metazoa (28), and Cyanobacteria (5). The statistical support for internal nodes was determined by bootstrap analyses (100 replicates) and is indicated at the corresponding branches (TF above, PhyML below). Only support values ≥40% are shown, cyt: cytosolic; pla: plastid-targeted.



Published in : Protist (2007), vol. 158, pp. 263-276.
Status : Postprint (Author's version)

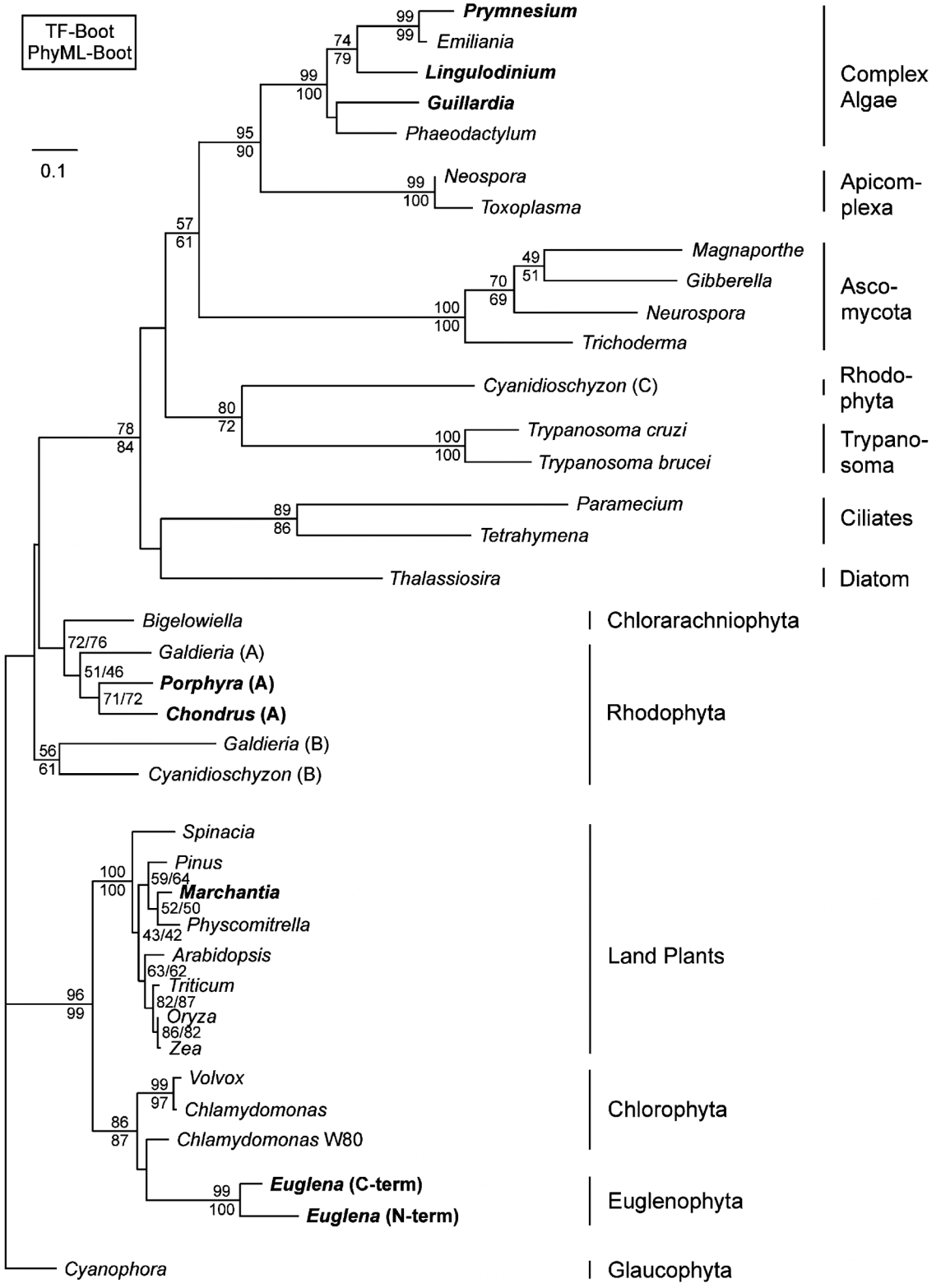
Phylogenetic Analyses of SBP Sequences

Several phylogenetic analyses of *SBP* and *FBP* sequences have shown that they are homologous and demonstrated the highly distinct nature of the monophyletic *SBP* subtree (Hannaert et al. 2003; Martin et al. 1996; Rogers and Keeling 2004). Figure 2 shows the maximum likelihood tree (PhyML) of 37 eukaryotic *SBP* sequences based on 197 aa positions. To maximize the number of alignment positions and to reduce the LBA effects, we excluded the highly divergent bacterial and eukaryotic *FBP* sequences from our analyses. The two fused *SBP* sequences from *Euglena* were analyzed separately. The data set contains species of the primary photosynthetic lineage including green plants, rhodophytes, and the glaucophyte *Cyanophora* as well as representatives of all complex lineages harboring red and green plastids (see *FBP* above). Moreover, all other available *SBP* sequences including those of non-photosynthetic apicomplexa, ciliates, trypanosomes, and fungi were retrieved and analyzed. The *SBP* phylogeny was arbitrarily rooted with the glaucophyte *Cyanophora* (Fig. 2). Among several distinct subtrees is found a highly supported clade (96/99%) that contains sequences from green plants as well as two very fast-evolving sequences of the complex "green" alga *Euglena*. Our analyses show that the N- and C-terminal copies of the *Euglena* gene originated from a recent gene duplication/fusion event and that they exhibit a close affiliation to chlorophycean sequences (86/87% support; Fig. 2), indicating endosymbiotic recruitment from the engulfed green alga. In contrast, the *SBP* from *Bigelowiella,* which also harbors a complex green plastid, groups with several red algal sequences (72/76% support). The presence of another "red" gene in *Bigelowiella* has previously been reported for plastid phosphoribulokinase (*PRK*; Petersen et al. 2006). Furthermore, an ancient gene duplication in rhodophytes led to an additional *SBP* copy only found in the Cyanidiales (*Cyanidioschyzon* [B] plus *Galdieria* [B]). From within the red algae emerges a group of divergent *SBP* sequences that comprises complex algae with red plastids, *Cyanidioschyzon* [C], and four non-photosynthetic lineages. Although the red algal affiliation of this group is not statistically substantiated, its common origin is solidly supported (78/84%). Since all sequences of this group are very fast evolving, basal relationships are expected to be dominated by ingroup LBA artifacts (Brinkmann et al. 2005; Philippe et al. 2005). The true affiliation of the *Cyanidioschyzon* sequence "C" remains unclear, but it might represent a very fast-evolving duplicate of the *Cyanidioschyzon* and *Galdieria SBP-"B"* gene. Support for this interpretation is given by an intron position shared between *Galdieria* [B] and *Guillardia* (see next section).

The most striking assembly of the *SBP* phylogeny is the distinct subtree (99/100%) containing sequences from all four complex lineages with red plastids including a diatom (*Phaeodactylum*), a cryptophyte (*Guillardia*), a dinoflagellate (*Lingulodinium*), and two haptophytes (*Prymnesium* and *Emiliania)*. Moreover, the sequences of *Toxoplasma* and *Neospora* (apicomplexa), both parasites with a reduced plastid (apicoplast), branch as sister to those of algae with complex rhodoplasts with a high support (90/95%). It seems possible that this apicomplexan *SBP* gene stems from the common red algal endosymbiont. If this is true, the absence of sequences from other completely sequenced apicomplexan genomes (e.g. *Plasmodium*, *Theileria*, and *Cryptosporidium*) would reflect a secondary loss in the course of gradual adaptation to a parasitic life style. However, it is conspicuous that these apicomplexan *SBP* sequences do not group together with the dinoflagellate *Lingulodinium*, while they are known to be sister groups within the alveolate superensemble. The topology of the complex algal sequences suggests a moderately supported sister group relationship between *Lingulodinium* and haptophytes (74/79%), whereas the relative branching order of *Guillardia* and *Phaeodactylum* is not resolved. The separate *SBP* from the second diatom *Thalassiosira* may represent a paralogous gene that originated either from an ancient gene duplication or HGT in a non-endosymbiotic context, but more sequences from stramenopiles are necessary to substantiate this interpretation. All available *SBP* sequences of non-photosynthetic and aplastidial eukaryotes (ciliates, trypanosomes, and fungi) form distinct and highly supported lineages (Fig. 2). The fungal subtree contains the longest branches, which clearly points towards accelerated evolutionary rates of the considered *SBP* sequences (more than two times higher than the already fast-evolving cytosolic *FBP*). A comparison with the phylogenetic analyses of cytosolic *FBP* genes (supplementary Fig. S1e) further reveals that the fungal *SBP* sequences are restricted to a small subgroup of closely related ascomycetes, whereas all other available fungal genomes contain no *SBP* homologs. The observed distribution in ascomycetes is therefore suggestive of HGT rather than common ancestry. Finally, it is important to highlight the patchy distribution of *SBP* not only in diatoms but also in kinetoplastids. The independent origin of *SBP* genes in *Euglena* and trypanosomes as well as their absence in *Leishmania* (completely sequenced genome) strengthens a HGT interpretation for trypanosomes.

Published in : Protist (2007), vol. 158, pp. 263-276.
Status : Postprint (Author's version)

Figure 2. *Maximum likelihood tree inferred using PhyML and a WAG+F+Γ4 model based on 37 SBP sequences and 197 amino acid positions. Sequences established in this study are in* ***boldface.*** The phylogenetic tree was arbitrarily rooted with Cyanophora. The statistical support for internal nodes was determined by bootstrap analyses and is indicated at the corresponding branches (TF above, PhyML below). Only support values ≥40% are shown.



Characterization of Genomic SBP Sequences

We determined the exon/intron distribution of *SBP* genes from the haptophyte *Prymnesium*, the cryptophyte *Guillardia*, and the rhodophyte *Chondrus* by comparison of newly established genomic and cDNA sequences, and retrieved further genomic sequences from public databases. The closely related *SBP* genes of *Prymnesium, Guillardia* and the diatom *Phaeodactylum* contain six, seven, and one intron at unique positions, indicating lineage-specific losses or gains within complex algae (supplementary Fig. S2). However, the intron position 296-0 is shared between the cryptophyte *Guillardia* and the red alga *Galdieria* (*SBP-"B"*) and may represent a shared ancestral character (symplesiomorphy).

Discussion

Origin, Function, and Distribution of SBP Genes

The Calvin cycle fructose-1,6-bisphosphatase/sedoheptulose-1,7-bisphosphatase (FBPase/SBPase) of cyanobacteria exhibits a dual substrate specificity, but was replaced by two monospecific isoenzymes in land plant chloroplasts. Both are encoded in the nucleus by genes of bacterial origin that are clearly homologous, but only distantly related. The plastid-targeted *FBP* is a duplicate of the closely related cytosolic *FBP* (Fig. 1), whereas the very divergent *SBP* displays no close affinity to any of the other *FBP* sequences (Martin et al. 1996).

SBPase was previously described as an essential enzyme of green plants that was acquired after primary endosymbiosis (Martin et al. 1996). In this study, we document the presence of SBPase in all lineages of photosynthetic eukaryotes containing primary plastids including green plants, rhodophytes, and glaucophytes (Plantae) as well as in all lineages with complex green (euglenophytes, chlorarachniophytes) or complex red plastids (haptophytes, cryptophytes, diatoms, dinoflagellates) (Fig. 2). Thus, it is likely that the eukaryotic architecture of the Calvin cycle including an additional enzyme specific for the SBPase function was established early in the primary photosynthetic lineage and then spread via subsequent eukaryote-to-eukaryote endosymbioses.

The presence of Calvin cycle SBP is universal among plants and algae. However, this is only part of the story since non-photosynthetic apicomplexa (harboring a non-photosynthetic apicoplast) and at least three aplastidial lineages (trypanosomes, fungi, ciliates) contain *SBP* homologs (Fig. 2) despite the lack of a Calvin cycle. The metabolic function of the respective enzymes is unclear, but they do not act as FBP substitutes since these heterotrophic eukaryotes also contain cytosolic *FBP* genes (Fig. 1; supplementary Fig. S1e). The presence of *SBP* within the apicomplexan parasites *Toxoplasma* and *Neospora* is especially noteworthy, because it might represent the ancient Calvin cycle enzyme of the endosymbiont, provided that it was not acquired via HGT Our phylogenetic analyses yield a strong association of apicomplexa and complex algae with red plastids (Fig. 2), and thus favor a red algal origin of the apicoplast in the long-standing dispute about its endosymbiotic ancestry (Funes et al. 2002, 2003; Rogers and Keeling 2004; Waller et al. 2003). The trypanosomal *SBP* was previously proposed to be a relict of a common photosynthetic ancestor of all Euglenozoa (Hannaert et al. 2003). However, our analyses strongly indicate that the *SBP* of *Euglena* is of green algal endosymbiotic origin (Fig. 2), which suggests an independent acquisition by kinetoplastids and considerably weakens the proposition of a photosynthetic life style for their common ancestors. Similarly, the very fast-evolving fungal *SBP* sequences are restricted to closely related ascomycetes (Fig. 2; supplementary Fig. S1e) and are obviously no indicators for fungal photosynthesis. Since the majority of more than 40 broadly sampled fungal genomes (ascomycetes, basidiomycetes, and zygomycetes) contain no *SBP* genes, a late acquisition by a subgroup of ascomycetes seems more likely than a genuine fungal origin. Finally, following the chromalveolate hypothesis, various recent publications speculate about a photosynthetic ancestry of ciliates (Cavalier-Smith 1999; Fast et al. 2001; Patron et al. 2004). While the presence of *SBP* sequences in *Paramecium* and *Tetrahymena* could in principle support such claims, the strongly supported common origin of alveolates (Rodriguez-Ezpeleta et al. 2005; Van de Peer and De Wachter 1997) is clearly not recovered in the *SBP* tree (Fig. 2). Therefore, these non-orthologous genes provide no evidence for a cryptic plastid in ciliates, which is in agreement with the recent analysis of the complete genome of *Tetrahymena thermophila* (Eisen et al. 2006). On the contrary, they likely fulfill a non-photosynthetic function, as proposed for their kinetoplastid and ascomycete counterparts. Taken together, two scenarios can explain the distribution of *SBP* genes in the aplastidial lineages. First, early in eukaryotic evolution a common ancestor of Plantae recruited this gene of which a divergent rhodophycean descendant, possibly related to the *Cyanidioschyzon* duplicate (*SBP-"C"*), was subsequently transmitted to the complex red lineage. Ascomycetes, trypanosomes, and ciliates could have acquired their fast-evolving *SBPs* by independent and late HGTs. Second, *SBP* is an old eukaryotic gene with a non-photosynthetic function that was already present in the host cell prior to primary plastid endosymbiosis. If the latter scenario is true, the *Cyanidioschyzon* (Cyanidiales)

sequence "C" is a relict of an ancient gene duplication and this divergent copy was lost in all other primary photosynthetic eukaryotes. Moreover, the absence of *SBP* from most fungi (zygomycetes, basidiomycetes, and a majority of ascomycetes), metazoa, and several protists [*Leishmania* (Euglenozoa), *Phytophthora* (oomycetes), *Dictyostelium*] implies several additional lineage-specific gene losses. In our opinion, the assumption of three HGTs (first scenario) is more likely than multiple independent losses. In conclusion, irrespective of which scenario is true, the sole presence of *SBP* genes in non-photosynthetic lineages does not constitute evidence for photosynthetic ancestry.

Diagnostic Gene Transfers of SBP and FBP Document the Common Origin of Complex Rhodoplasts

Endosymbiotic gene transfer (EGT) to the host cell nucleus after primary and eukaryote-to-eukaryote endosymbioses represents a major source of new nuclear genes in primary and complex plastids (Martin and Herrmann 1998). A clear example for EGT is the *SBP* from *Euglena* that was most likely recruited from the engulfed green alga (Fig. 2).

However, their Calvin cycle *FBP* appears to be of lateral origin from a complex alga with a red plastid (e.g. a dinoflagellate; Fig. 1) and a similar red algal origin for this plastidial gene in *Bigelowiella* seems likely. Comparable cases of HGT have already been reported for these two phagotrophic complex "green algae" (Archibald et al. 2003; Petersen et al. 2006), and such lineage-specific acquisitions represent the fortuity of gene recruitment (Martin and Herrmann 1998). In this study, we identified two EGTs concerning *SBP* and plastid-targeted *FBP* that have important evolutionary implications. Our phylogenetic analyses provide strong support (99/100%) for the common origin of *SBP* genes from haptophytes (*Prymnesium*, *Emiliania*), cryptophytes (*Guillardia*), diatoms (*Phaeodactylum*), and dinoflagellates (*Lingulodinium*). These lineages of complex algae harboring red plastids in turn display a strongly supported sister group relationship with apicomplexan *SBPs* (*Toxoplasma*, *Neospora* [95/90%]) (Fig. 2). The exact phylogenetic origin of this assemblage remains inconclusive due to the fast evolutionary rate, but its association with red algae and an intron position shared between *Guillardia* and *Galdieria* (*SBP-"B"*; supplementary Fig. S2) both argue in favor of an endosymbiotic origin from a rhodophyte. Plastid-targeted *FBP* probably exhibits a second but less pronounced example for common branching of all photosynthetic lineages harboring a complex rhodoplast. Indeed, the closely related homologous sequences of *Emiliania*, *Guillardia*, *Lingulodinium*, and diatoms likely share the same red algal origin (Fig. 1; supplementary Fig. S1d), assuming that *Euglena* and *Bigelowiella* recruited their genes in a non-endosymbiotic context (see above). Thus, we present here two plastid-related genes *(SBP* and *FBP)* of red algal origin that support a common ancestry of plastids from haptophytes, diatoms, cryptophytes, and dinoflagellates. The monophyly of these taxonomic groups was also obtained for other plastid-related genes, i.e. *GAPDH* (*GapC-I*), class II aldolase (*FBA-II*), and phosphoribulokinase (*PRK*) (Fast et al. 2001; Patron et al. 2004; Petersen et al. 2006). However, plastid *SBP* and *FBP* are the first nuclear-encoded genes that reflect the expected rhodophycean origin since the other genes display no specific relationship to the red lineage. The *PRK* of algae with complex red plastids even shows a strong affiliation to the green lineage and was most likely obtained from a chlorophyte via HGT (Petersen et al. 2006). In conclusion, irrespective of the mode of recruitment, the independent fixation of at least five markers for the plastid Calvin cycle in the nucleus of the host cell provides strong evidence for a common origin of complex red plastids via a single secondary endosymbiosis with a red alga.

Implications for the Evolution of Complex Algae

A single secondary endosymbiosis with a red alga in a common ancestor of all "chromists" (haptophytes, cryptophytes, stramenopiles) and alveolates (ciliates, apicomplexa, dinoflagellates) is the central proposal of the chromalveolate hypothesis (Cavalier-Smith 1999). A common origin of "chromalveolates" implies the monophyly of the complex red plastids and of their host cells as well as the congruence of plastid and nuclear phylogenies. Therefore, it is suspicious that the five markers that are consistent with the chromalveolate hypothesis all encode plastid-located proteins of the Calvin cycle, which were likely fixed in the nucleus soon after the secondary endosymbiosis with a rhodophyte. On the other hand, there is not a single gene representing the evolution of the host cell that delivers comparable support (see e.g. cytosolic *FBP*; Fig. 1). While there is no reason to believe that haptophytes, cryptophytes, diatoms, and dinoflagellates recruited these independent markers by chance, their unusually high resolving power (in most cases 100% bootstrap support) is striking for proteins of barely 300 aa. In contrast, more than 5000 aa positions were required to achieve significant support (95% BP) for the common origin of alveolates and stramenopiles (Rodriguez-Ezpeleta et al. 2005). However, a closer look at the ingroup topologies of these five "lucky genes" (Bapteste et al. 2002) reveals discrepancies with the accepted relationships of the host cells. The *SBP* phylogeny (Fig. 2) obviously does not reflect a common origin of alveolate sequences (discussed above) and a similar situation, where *Toxoplasma* displays no

association with the dinophycean sequence, was previously reported for plastid GAPDH (Harper and Keeling 2003). Based on plastid genome analyses, cryptophytes and diatoms branch within red algae to the exclusion of the early branching Cyanidiales (Rodriguez-Ezpeleta et al. 2005), while the presence of a chlorophycean *PRK* in all orders of complex algae with red plastids indicates that the secondary endosymbiosis occurred after the split between streptophytes and chlorophytes (Petersen et al. 2006). This implies that "chromalveolates" originated relatively late in terms of global eukaryotic evolution, i.e. after the three primary photosynthetic groups (Plantae) had separated and their major lineages were well established. If this were true, the monophyly of the "chromalveolate" lineages should not be that elusive.

Taken together, the three points raised above (suspicious resolving power, plastid/host incongruencies, and the timing paradox) cannot be adequately explained in the framework of the chromalveolate hypothesis. Instead, we propose that the origin of haptophytes, cryptophytes, diatoms, and peridinin-containing dinoflagellates is best interpreted as the result of a late and unique secondary endosymbiosis with a red alga followed by subsequent tertiary or even quaternary endosymbioses (Bachvaroff et al. 2005; Delwiche 1999; Petersen et al. 2006). Eukaryote-to-eukaryote endosymbioses with complex algae have already been documented for dinoflagellates that independently reduced either a haptophyte, or a diatom or a cryptophyte to a complex plastid (Chesnick et al. 1997; Hackett et al. 2003; Tengs et al. 2000). Once the plastid tool box was assembled in the host nucleus (Stoebe and Maier 2002), subsequent endosymbioses were likely facilitated. Hence, the genes for Calvin cycle enzymes including *SBP* and *FBP* were recruited only once in the course of secondary endosymbiosis, then adapted to the complex red plastid, and subsequently spread to unrelated lineages as part of third hand plastids. Tertiary endosymbioses would also explain the otherwise contradictory data of mitochondrial genes and 18S rDNA (Sanchez Puerta et al. 2004, Van de Peer and De Wachter 1997), as well as account for the different types of mitochondria across "chromists" (Gray et al. 1998). The location of the complex algal clade in the cytosolic *FBP* tree (Fig. 1; supplementary Fig. S3) indicates that a phagotrophic stramenopile might be a potential candidate for the host cell that underwent single secondary endosymbiosis with a rhodophyte. Besides cytosolic *FBP,* we predict the presence of other non-plastid-related remnants of the secondary host in the nucleus of "tertiary" algae. Identifying more of these genes is probably the best way to validate our alternative scenario and to clarify which lineage(s) still harbor secondary and which one(s) already contain(s) tertiary plastids.

Methods

*Algal material***:** The haptophyte *Prymnesium parvum* (strain 127.79) was obtained from the "Sammlung von Algenkulturen" at the University of Göttingen (SAG) and the liverwort *Marchantia polymorpha* was obtained from the Deutsche Sammlung für Mikroorganismen and Zellkulturen (DSMZ). Culturing was performed as previously described (Petersen et al. 2003, 2006). The algal material of the rhodophyte *Chondrus crispus* was field-collected on the North Sea island Helgoland.

Isolation of nucleic acids and construction of libraries: The isolation of nucleic acids, the preparation of λZAPII libraries from *Chondrus crispus* and *Marchantia polymorpha,* and the preparation of a λEMBL3 library from *Prymnesium parvum* have been previously described (Petersen et al. 2003, 2006). Both mRNA and the λZAPII cDNA library from *Euglena gracilis* were provided by William Martin (University of Düsseldorf), the λZAPII library from *Lingulodinium polyedrum* was donated by Woodland Hastings (Harvard University).

RT-PCR amplification: Specific probes were amplified via RT-PCR using the Thermoscript RT-PCR System (Invitrogen). Degenerated primers for PCR amplification were designed based on the universally conserved N-terminal SBP/FBP motif WFDPLDG (5'-GTSGTSTTCGACCCSCTNGAY-GG-3'), the C-terminal SBP motif YTGGMVPD (5'-ACSGGSACCATSCCSCCSGTRTA-3'), and the C-terminal FBP motif EQAGG(K/Q)(G/A) (5'-SCYTKSCCSCCSGCYTGYTC-3'). Reverse transcription, PCR amplification, and cloning was performed as previously described (Petersen et al. 2006). The different *SBP* and *FBP* clones were identified by sequencing using radioactive and fluorescence techniques.

Isolation and sequencing of cDNA and genomic clones: Prymnesium parvum, Euglena gracilis, Chondrus crispus, and *Marchantia polymorpha* libraries were screened using ^{32}P-labeled homologous RT-PCR probes for *SBP* and *FBP* genes (Petersen et al. 2003, 2006). *SBP* screening of the *Lingulodinium polyedrum* library was performed with previously identified probes under less stringent conditions. *FBP* screening was performed with a homologous probe, amplified via RT-PCR with primers specific for EST clone CD810868 [GenBank]. cDNA clones from *Guillardia theta* were identified from an ongoing EST project. The cDNA clones from λZAPII cDNA libraries were subcloned in pBluescript II SK(+) by single clone excision. The genomic EMBL3 clone from *Prymnesium parvum* was subcloned into the *Pst*I site of pBluescript II SK (+). The Kazusa Institute (Japan) established *SBP* and *FBP* ESTs [GenBank: AV432345, AV432933] of the rhodophyte *Porphyra yezoensis* and

kindly donated the respective cDNA clones. All clones were sequenced on both strands using pBluescript or gene-specific primers.

Data deposition: Sequences established in this study have been deposited in the DDBJ/EMBL/ GenBank International Nucleotide Sequence Database under the following accession numbers: DQ508151-DQ508166.

Identification of clones by database analyses: The *SBP* sequence of the glaucophyte *Cyanophora paradoxa* was obtained from the "Taxonomically Broad EST Database" (TBestDB; University of Montreal), and *SBP* and *FBP* sequences from the red algae *Galdieria sulphur-aria* and *Cyanidioschyzon merolae* were identified from the *Galdieria* database (GDB; Michigan State University) and the "*Cyanidioschyzon merolae* Genome Project" (CMGP; http://www.merolae.biol.s.u-tokyo.ac.jp/) respectively. Those of the diatom *Thalassiosira pseudonana* were established at the "Joint Genome Institute" (JGI; USA). Nucleotide and deduced amino acid sequences of *SBP* and *FBP* were used as query sequences for BLAST searches in the NCBI database (TBLASTN; BLASTP; MEGABLAST). The respective clones were retrieved from GenBank and if necessary assembled into contigs. The authenticity of the *SBP* and *FBP* sequences from *Tetrahymena* and *Paramecium* was assessed by the ciliate-specific codon usage. Table S1 contains a list of all sequences established and analyzed in this study *(SBP* and *FBP* alignments are available upon request).

Sequence handling and phylogenetic analyses: The initial alignments, obtained with CLUSTAL X (Thompson et al. 1997), were manually refined using the ED option of the MUST program package (Philippe 1997). Unreliable parts of the alignments (e.g. indels, low complexity areas, highly variable regions) were automatically removed using the program G-BLOCKS (Castresana 2000). All data sets were analyzed by two different Maximum Likelihood (ML) methods, implemented in PhyML version 2.4 (Guindon and Gascuel 2003) and Treefinder (Jobb et al. 2004). The Whelan and Goldman (WAG) matrix of aa replacements (2001) corrected by the aa frequencies estimated from the data set and four discrete gamma distributed rates were used (WAG+F+Γ4). Bootstrap analyses with 100 replicates were performed with both methods to estimate the support for internal nodes. The CONSENSE option of the PHYLIP package (http://www.evolution.genetics.washington.edu/phylip.html) was used to generate the bootstrap consensus trees (100 replicates).

Acknowledgments

We thank William Martin (Düsseldorf) for the cDNA library as well as mRNA from *Euglena gracilis*, Woody Hastings (Harvard) for the cDNA library from *Lingulodinium polyedrum*, and the Kazusa DNA Research Institute (Japan) for providing *SBP* and *FBP* cDNA clones from *Porphyra yezoensis*. We are grateful to Sarina Scharbatke (Hanover) for excellent technical assistance and Naiara Rodriguez-Ezpeleta for helpful comments on the manuscript. The sequence data from *Galdieria sulphuraria* were provided by the Michigan State University *Galdieria* Database (http://www.genomics.msu.edu/galdieria_public). Major financial support, including a Ph.D. stipend for R.T., was received from the Deutsche Forschungsgemeinschaft (CE 1/27-2). D.B. is a postdoctoral researcher of the Fonds National de la Recherche Scientifique (FNRS) at the University of Liège (Belgium) and is gratefully indebted to the FNRS for the financial support of his stay at the University of Montréal. The authors also want to thank two anonymous reviewers for careful reading and constructive criticism of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.vacuum.2006.03.027.

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 Ø, 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

Archibald JM, Rogers MB, Toop M, Ishida K, Keeling PJ (2003) Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. Proc Natl Acad Sci USA 100: 7678-7683

Bachvaroff TR, Sanchez-Puerta MV, Delwiche CF (2005) Chlorophyll c-containing plastid relationships based on analyses of a multigene data set with all four chromalveolate lineages. Mol Biol Evol 22: 1772-1782

Bapteste E, Brinkmann H, Lee JA, Moore DV, Sensen CW, Gordon P, Durufle L, Gaasterland T, Lopez P, Müller M, Philippe H (2002) The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. Proc Natl Acad Sci USA 99: 1414-1419

Bhattacharya D, Medlin L (1995) The phylogeny of plastids: a review based on comparison of small-subunit ribosomal RNA coding regions. J Phycol 31: 489-498

Bhattacharya D, Helmchen T, Melkonian M (1995) Molecular evolutionary analyses of nuclear-encoded small subunit ribosomal RNA identify an independent rhizopod lineage containing the Euglyphina and the Chlorarachniophyta. J Eukaryot Microbiol 42: 65-69

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

Brinkmann H, van der Giezen M, Zhou Y, Poncelin de Raucourt G, Philippe H (2005) An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. Syst Biol 54: 743-757

Buchanan BB (1980) Role of light in the regulation of chloroplast enzymes. Annu Rev Plant Physiol 31: 341-374

Cai X, Fuller AL, McDougald LR, Zhu G (2003) Apicoplast genome of the coccidian *Eimeria tenella*. Gene 321: 39-46

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17: 540-552

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

Chesnick JM, Kooistra WH, Wellbrock U, Medlin L (1997) Ribosomal RNA analysis indicates a benthic pennate diatom ancestry for the endosymbionts of the dinoflagellates *Peridinium foliaceum* and *Peridinium balticium* (Pyrrhophyta). J Eukaryot Microbiol 44: 314-320

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

Dooijes D, Chaves I, Kieft R, Martin W, Borst P (2000) Base J originally found in kinetoplastida is also a minor constituent of nuclear DNA of *Euglena gracilis*. Nucleic Acid Res 28: 3017-3021

Douglas SE (1998) Plastid evolution: origins, diversity, trends. Curr Opin Genet Dev 8: 655-661

Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, Deng LT, Wu X, Reith M, Cavalier-Smith T, Maier UG (2001) The highly reduced genome of an enslaved algal nucleus. Nature 410: 1091-1096

Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, Wortman JR, Badger JH, Ren Q, Amedeo P, Jones KM, Tallon LJ, Delcher AL, Salzberg SL, Silva JC, Haas BJ, Majoros WH, Farzad M, Carlton JM, Smith RK, Garg J, Pearlman RE, Karrer KM, Sun L, Manning G, Elde NC, Turkewitz AP, Asai DJ, Wilkes DE, Wang Y, Cai H, Collins K, Stewart BA, Lee SR, Wilamowska K, Weinberg Z, Ruzzo WL, Wloga D, Gaertig J, Frankel J, Tsao CC, Gorovsky MA, Keeling PJ, Waller RF, Patron NJ, Cherry JM, Stover NA, Krieger CJ, Del Toro C, Ryder HF, Williamson SC, Barbeau RA, Hamilton EP, Orias E (2006) Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. PLoS Biol 4: 9

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

Felsenstein J (1978) Cases in which parsimony and compatibility methods will be positively misleading. Syst Zool 27: 401-410

Funes S, Davidson E, Reyes-Prieto A, Magallon S, Herion P, King M, Gonzalez-Halphen D (2002) A green algal apicoplast ancestor. Science 298: 2155

Funes S, Davidson E, Reyes-Prieto A, Magallon S, Herion P, King M, Gonzalez-Halphen D (2003) Response to comment on "A green algal apicoplast ancestor". Science 301: 49b

Gerbling KP, Steup M, Latzko E (1986) Fructose-1,6-bisphosphatase form B from *Synechococcus leopoliensis* hydrolyzes both fructose and sedoheptulose bisphosphate. Plant Physiol 80: 716-720

Gilson PR, McFadden Gl (1997) Good things in small packages: the tiny genomes of chlorarachniophyte endosymbionts. Bioessays 19: 167-173

Gilson PR, McFadden Gl (2002) Jam packed genomes — a preliminary, comparative analysis of nucleomorphs. Genetica 115: 13-28

Gray MW, Burger G, Lang BF (1999) Mitochondrial evolution. Science 283: 1476-1481

Gray MW, Lang BF, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Brossard N, Delage E, Littlejohn TG, Plante I, Rioux P, Saint-Louis D, Zhu Y, Burger G (1998) Genome structure and gene content in protist mitochondrial DNAs. Nucleic Acids Res 26: 865-878

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

Hackett JD, Maranda L, Yoon HS, Bhattacharya D (2003) Phylogenetic evidence for the cryptophyte origin of the plastid of Dinophysis (Dinophysiales, Dinophyceae). J Phycol 39: 440-448

Hannaert V, Saavedra E, Duffieux F, Szikora JP, Rigden DJ, Michels PA, Opperdoes FR (2003) Plant-like traits associated with metabolism of *Trypanosoma* parasites. Proc Natl Acad Sci USA 100: 1067-1071

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

Jobb G, von Haeseler A, Strimmer K (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol 4: 18

Lefebvre S, Lawson T, Zakhleniuk OV, Lloyd JC, Raines CA, Fryer M (2005) Increased sedoheptulose-1,7-bisphosphatase activity in transgenic tobacco plants stimulates photosynthesis and growth from an early stage in development. Plant Physiol 138: 451-460

Li S, Nosenko T, Hackett JD, Bhattacharya D (2006) Phylogenomic analysis identifies red algal genes of endosymbiotic origin in the chromalveolates. Mol Biol Evol 23: 663-674

Margulis L (1975) Symbiotic theory of the origin of eukaryotic organelles; criteria for proof. Symp Soc Exp Biol 29: 21-38

Marin B, Nowack EC, Melkonian M (2005) A plastid in the making: evidence for a second primary endosymbiosis. Protist 156: 425-432

Martin W, Borst P (2003) Secondary loss of chloroplasts in trypanosomes. Proc Natl Acad Sci USA 100: 765-767

Martin W, Herrmann RG (1998) Gene transfer from organelles to the nucleus: how much, what happens, and why? Plant Physiol 118: 9-17

Martin W, Kowallik KV (1999) Annotated English translation of Mereschkowsky's 1905 paper "Uber Natur und Ursprung der Chromatophoren im Pflanzenreiche.". Eur J Phycol 34: 287-295

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 AZ, 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

Maslov D, Yasuhira S, Simpson L (1999) Phylogenetic affinities of *Diplonema* within the Euglenozoa as inferred from the SSU rRNA gene and partial COI protein sequences. Protist 150: 33-42

McFadden Gl, Gilson PR, Hofmann CJ, Adcock GJ, Maier UG (1994) Evidence that an amoeba acquired a chloroplast by retaining part of an engulfed eukaryotic alga. Proc Natl Acad Sci USA 91: 3690-3694

Mereschkowsky C (1905) Uber Natur und Ursprung der Chromatophoren im Pflanzenreiche. Biol Zentbl 18: 593-604

Miyagawa Y, Tamoi M, Shigeoka S (2001) Overexpression of a cyanobacterial fructose-1,6-/sedoheptulose-1,7-bisphosphatase in tobacco enhances photosynthesis and growth. Nat Biotechnol 19: 965-969

Patron NJ, Rogers MB, Keeling PJ (2004) Gene replacement of fructose-1,6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. Eukaryot Cell 3: 1169-1175

Petersen J, Brinkmann H, Cerff R (2003) Origin, evolution, and metabolic role of a novel glycolytic GAPDH enzyme recruited by land plant plastids. J Mol Evol 57: 16-26

Petersen J, Teich R, Brinkmann H, Cerff R (2006) A "green" phosphoribulokinase in complex algae with red plastids: Evidence for a single secondary endosymbiosis leading to haptophytes, cryptophytes, heterokonts and dinoflagellates. J Mol Evol 62: 143-157

Philippe H (1997) MUST, a computer package of management utilities for sequences and trees. Nucleic Acids Res 2: 5264-5272

Philippe H, Zhou Y, Brinkmann H, Rodrigue N, Delsuc F (2005) Heterotachy and long-branch attraction in phylogenetics. BMC Evol Biol 5: 50

Plaxton WC (1996) The organization and regulation of plant glycolysis. Annu Rev Plant Physiol Plant Mol Biol 47: 185-214

Raines CA (2003) The Calvin cycle revisited. Photosynth Res 75: 1-10

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

Rogers M, Keeling PJ (2004) Lateral transfer and recompartmentalization of Calvin cycle enzymes of plants and algae. J Mol Evol 58: 367-375

Rogers MB, Gilson PR, Su V, McFadden Gl, Keeling PJ (2007) The complete chloroplast genome of the chlorarachniophyte *Bigelowiella natans:* Evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. Mol Biol Evol 24: 54-62

Sanchez-Puerta MV, Bachvaroff TR, Delwiche CF (2004) The complete mitochondrial genome sequence of the haptophyte *Emiliania huxleyi* and its relation to heterokonts. DNA Res 11: 1-10

Schnarrenberger C, Jacobshagen S, Müller B, Krüger I (1990) Evolution of isoenzymes of sugar phosphate metabolism in green algae. Prog Clin Biol Res 344: 743-764

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

Stoebe B, Maier UG (2002) One, two, three: nature's tool box for building plastids. Protoplasma 219: 123-130

Tengs T, Dahlberg O, Shalchian-Tabrizi K, Klaveness D, Rudi K, Delwiche C, Jakobsen K (2000) Phylogenetic analyses indicate that the 19'Hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. Mol Biol Evol 17: 718-729

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4876-4882

Van de Peer Y, De Wachter R (1997) Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site rate variation in 18S rRNA. J Mol Evol 45: 619-630

Waller RF, Keeling PJ, van Dooren GG, McFadden GI (2003) Comment on "A green algal apicoplast ancestor". Science 301: 49a

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18: 691-699

Yoon HS, Hackett JD, Pinto G, Bhattacharya D (2002) The single, ancient origin of chromist plastids. Proc Natl Acad Sci USA 99: 15507-15512

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