Localization and Evolution of Putative Triose Phosphate Translocators in the Diatom *Phaeodactylum tricornutum*

Daniel Moog[1,6], Stefan A. Rensing[2], John M. Archibald[3,4], Uwe G. Maier[1,5,*], and Kristian K. Ullrich[2]

[1]LOEWE Centre for Synthetic Microbiology (SYNMIKRO), Philipps University Marburg, Germany

[2]Plant Cell Biology, Philipps University Marburg, Germany

[3]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada

[4]Program in Integrated Microbial Biodiversity, Canadian Institute for Advanced Research, Toronto, Ontario, Canada

[5]Laboratory for Cell Biology, Philipps University Marburg, Germany

[6]Present address: Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada.

*Corresponding author: E-mail: maier@biologie.uni-marburg.de.

Accepted: September 29, 2015

Abstract

The establishment of a metabolic connection between host and symbiont is a crucial step in the evolution of an obligate endosymbiotic relationship. Such was the case in the evolution of mitochondria and plastids. Whereas the mechanisms of metabolite shuttling between the plastid and host cytosol are relatively well studied in Archaeplastida—organisms that acquired photosynthesis through primary endosymbiosis—little is known about this process in organisms with complex plastids. Here, we focus on the presence, localization, and phylogeny of putative triose phosphate translocators (TPTs) in the complex plastid of diatoms. These proteins are thought to play an essential role in connecting endosymbiont and host metabolism via transport of carbohydrates generated by the photosynthesis machinery of the endosymbiont. We show that the complex plastid localized TPTs are monophyletic and present a model for how the initial metabolic link between host and endosymbiont might have been established in diatoms and other algae with complex red plastids and discuss its implications on the evolution of those lineages.

Key words: complex plastid, secondary endosymbiosis, metabolic coupling, protein targeting, diatoms.

Introduction

In endosymbioses between a heterotrophic host cell and a phototrophic endosymbiont, the host profits from energy-rich carbohydrates synthesized by the endosymbiont's photosynthetic machinery. In 1905, Mereschkowsky postulated that the exchange of carbohydrates might have been a crucial factor in the establishment of a stable host-endosymbiont relationship during the evolution of photosynthetic eukaryotes. However, it is still not known if carbohydrates were the initial driving force or whether other factors or metabolites might have played an important role during stabilization of endosymbioses involving photosynthetic endosymbionts.

To gain access to symbiont-produced metabolites, membrane-spanning molecules evolved, which enabled directional transport from the endosymbiont (which became the organelle) to the host (McFadden 2014). Research over the past decade has revealed several transporters with the ability to transport metabolites out of and into the plastid (chloroplast) of land plants, which in the case of carbohydrate transport are predominantly plastidic phosphate translocators (pPTs) (Flügge 1999; Weber et al. 2005). This group of typically dimeric membrane channels includes triose phosphate translocators (TPTs), phosphoenolpyruvate translocators (PPTs) as well as glucose 6-phosphate (GPTs)/xylulose 5-phosphate (XPTs) translocators, all of which function as strict antiporters for inorganic phosphate (Weber and Linka 2011). All classes of pPTs are present in land plant plastids (green lineage) whereas in red algae only functional TPTs and PPTs have thus far been identified (Weber and Linka 2011). Glaucophytes seem to lack typical pPTs. However, translocators of the bacterial UhpC-type, known to transport hexose phosphates, were recently identified in the genome of the glaucophyte *Cyanophora paradoxa* (Price et al. 2012). These UhpC-type transporters are also found in red and green algae and are, besides the nucleotide sugar transporter-derived pPTs in these two lineages (Weber et al. 2006), currently thought to be a pivotal

factor in the establishment of a metabolic connection between host and endosymbiont in the progenitor of all Archaeplastida (green algae, their land plant descendants, red algae, and glaucophytes) (Karkar et al. 2015). Thus, carbohydrate transport out of the primary plastids of Archaeplastida is in many respects well understood.

In contrast, relatively little is known about the process of carbohydrate transport across the multitude of cellular membranes surrounding the plastids of secondarily evolved eukaryotic phototrophs (for review see Weber and Linka 2011). Secondary plastids are the result of endosymbioses in which a red or green algal endosymbiont was reduced to a so-called complex plastid surrounded by three or four membranes (Gould et al. 2008; Archibald 2009; Keeling 2013). This cellular architecture makes metabolite exchange between stroma and cytoplasm of the host more complicated. In organisms with a red algal-derived plastid surrounded by four membranes (cryptophytes, apicomplexans/chromerids, stramenopiles, and haptophytes; CASH (Petersen et al. 2014)) convincing homologs of PPTs and GPTs/XPTs cannot generally be identified in sequenced genomes. TPTs have, however, been identified in several CASH lineages and appear to have a monophyletic origin (Haferkamp et al. 2006; Weber et al. 2006).

Studies of TPTs in the non-photosynthetic complex plastid (or "apicoplast") of medically important apicomplexans revealed the presence of two TPTs in *Plasmodium falciparum* (Mullin et al. 2006) and at least one (in addition to a possible PPT homolog) in *Plasmodium berghei* (Banerjee et al. 2012), whereas only one TPT was found in *Toxoplasma gondii* (Fleige et al. 2007). While the *P. falciparum* TPTs localize to the outer (PfoTPT) and inner (PfiTPT) complex plastid membrane, respectively (Mullin et al. 2006), the TPT of *T. gondii* (TgAPT), which is essential for parasite survival (Brooks et al. 2010), localizes to multiple complex plastid membranes (Fleige et al. 2007; Karnataki et al. 2007). Another interesting feature of the translocators of both organisms is their similar substrate specificity: they transport mainly triose phosphate (dihydroxyacetone phosphate, DHAP) as well as phosphoenolpyruvate (PEP) in exchange for inorganic phosphate (Brooks et al. 2010; Lim et al. 2010). The *P. berghei* TPT similarly localizes to the complex plastid and was shown to be essential for parasite survival (Banerjee et al. 2012). Because no photosynthesis, and therefore carbon (CO_2) fixation, takes place in the parasite's complex plastid, it is assumed that transport occurs only from the host cytosol into the complex plastid to fuel local biosynthetic pathways such as isoprenoid and fatty acid biosynthesis, and to provide an energy source for organellar functions (Mullin et al. 2006; Brooks et al. 2010).

The presence of complex plastid TPTs was also reported for the cryptophyte *Guillardia theta* (Haferkamp et al. 2006), an organism that retains the vestigial nucleus of the red algal endosymbiont (nucleomorph) within the periplastidal compartment (PPC, reduced endosymbiont cytoplasm) of its complex plastid, where starch is stored (Douglas et al. 2001; Deschamps et al. 2006). Haferkamp et al. (2006) identified two TPTs from expressed sequence tag (EST) data, which are able to catalyze the exchange of DHAP and PEP with inorganic phosphate, as shown by transport assays with isolated complex plastid membranes. Gene expression analysis demonstrated diurnal regulation for TPT1 (dark) and TPT2 (light), which correlates with starch metabolism (Haferkamp et al. 2006). Haferkamp et al. (2006) concluded that the TPT1 protein may be localized to the plastid surrounding endoplasmic reticulum ("chloroplast ER," cER) membrane (outermost membrane of the complex plastid), whereas TPT2 probably resides in one of the two innermost plastid membranes (third or fourth membrane).

Diatoms are another group of ecologically important secondarily evolved phototrophs and have become an important model system for the study of complex plastids (Stork et al. 2013). With respect to intracellular carbohydrate transport, the plastid (stroma) localization of a TPT presequence-green fluorescent protein (GFP) fusion protein has been reported for one TPT in the diatom *Phaeodactylum tricornutum* (Kilian and Kroth 2005). In addition, we recently reported on the localization of a further putative TPT (TPT2) in the periplastidal membrane (PPM, second outermost) of *P. tricornutum* (Lau et al. 2015). Here, we investigate the presence, localization, and phylogeny of TPTs in diatoms in more detail to complete the puzzle of TPT function and evolution in this ecologically important group of organisms. We show that, besides the PPM-localized TPT2, three additional putative TPTs of red algal origin are present in the complex plastid envelope membranes of *P. tricornutum*. We confirm that the endosymbiont-derived TPTs are integral membrane proteins and provide a model for the connection of endosymbiont and host metabolic pathways, which is based on the exchange of phosphorylated C3-compounds. Our data reveal how the diatom host cell integrated endosymbiont-derived components into its own compartments, thereby utilizing the symbiont's ability to generate chemically stored energy. The use of "old" translocators at new cellular destinations facilitated the evolution of new cellular entities.

Materials and Methods

Domain Architecture and Sequence Feature Prediction

To identify putative TPTs in the genome of the diatom *P. tricornutum,* we performed Basic Local Alignment Search Tool (BLAST) analyses using the amino acid sequence of the complex plastid (apicoplast) localized TPT from *T. gondii* (NCBI accession number: ABU49222) as query (Fleige et al. 2007; Karnataki et al. 2007). Detected sequences then were used for BLAST searches against the *P. tricornutum* genome database (http://genome.jgi-psf.org/Phatr2/Phatr2.home.html, last accessed October 19, 2015) to identify additional homologs. BLAST analyses were further complemented by hmmer

searches (version 3.1b1) (Finn et al. 2011) using existing protein family Pfam domains (TPT, PF03151; Nuc_sug_transp, PF04142; UAA, PF08449; and EamA, PF00892) and the *P. tricornutum* genome (Phatr2_chromosomes_geneModels_FilteredModels2_aa.fasta.gz). Two detection thresholds were applied to cope with hmmer profile similarities between the used Pfam domains and to uncover domain architecture. Here, either the existing gathering cutoff [–cut_ga] was used as a stringent threshold or an *e*-value of 100 [-E 100] as a very loose detection threshold, only retaining domain hits with domain scores higher or equal to 10 in the resulting domtblout file [–domtblout] (supplementary table S1, Supplementary Material online). Gene models of thereby detected proteins TPT1 (model name Phatr2: estExt_Phatr1_ua_kg.C_chr_20080), TPT2 (estExt_gwp_gw1.C_chr_100108), TPT4a (estExt_Phatr1_ua_kg.C_chr_10107), TPT4b (estExt_Genewise1.C_chr_10395), TPT6 (estExt_fgenesh1_pg.C_chr_300063) TPT12 (e_gw1.1.514.1), and TPT13 (estExt_fgenesh1_pg.C_chr_70295) were completely covered by EST data, whereas for TPT5 (estExt_fgenesh1_pg.C_chr_60290), TPT7 (fgenesh1_pg.C_chr_24000019), TPT8 (estExt_fgenesh1_pg.C_chr_20047), TPT9 (fgenesh1_pg.C_chr_16000283), TPT10 (estExt_fgenesh1_pg.C_chr_230136), and TPT11 (estExt_fgenesh1_pg.C_chr_10932) only partial EST data was present.

Protein sequences were inspected with SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP-3.0/, last accessed October 19, 2015) (Bendtsen et al. 2004), TargetP 1.1 (http://www.cbs.dtu.dk/services/TargetP/, last accessed October 19, 2015) (Emanuelsson et al. 2000), and blastp (http://blast.ncbi.nlm.nih.gov/Blast.cgi, last accessed October 19, 2015) to search for the presence of putative N-terminal targeting sequences (supplementary table S2, Supplementary Material online) as well as conserved protein sections and functional domains (Moog et al. 2011, 2015; Stork et al. 2012). In addition to the tools aforementioned, sequences were analyzed using SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/, last accessed October 19, 2015) (Petersen et al. 2011), Predotar (https://urgi.versailles.inra.fr/predotar/predotar.html, last accessed October 19, 2015) (Small et al. 2004), PredSL (http://aias.biol.uoa.gr/PredSL/input.html, last accessed October 19, 2015) (Petsalaki et al. 2006), WoLF PSORT (http://www.genscript.com/psort/wolf_psort.html, last accessed October 19, 2015) (Horton et al. 2007), PSORTII (http://psort.hgc.jp/form2.html, last accessed October 19, 2015) (Nakai and Horton 1999), PTS1 predictor (http://mendel.imp.ac.at/mendeljsp/sat/pts1/PTS1predictor.jsp, last accessed October 19, 2015) (Neuberger et al. 2003), and Target signal predictor (peroxisomal target signals (PTSs) predictor; http://216.92.14.62/Target_signal.php, last accessed October 19, 2015) (Schlüter et al. 2007). Searches for potential transmembrane domains (TMDs) within the primary sequences were conducted using TMHMM (http://www.cbs.dtu.dk/services/TMHMM-2.0/, last accessed October 19, 2015) (Krogh et al. 2001) (supplementary table S2, Supplementary Material online).

Sequence Amplification and Plasmid Construction

For bipartite targeting sequence (BTS) and full-length eGFP localization studies, all relevant TPT-encoding sequences were amplified from *P. tricornutum* cDNA using primers usually equipped with terminal restriction sites for *Eco*RI and *Bam*HI (supplementary table S3, Supplementary Material online). Amplified sequences were cloned together with *egfp* (*Bam*HI/*Hind*III) into the light-inducible pPha-T1 (NCBI accession number: AF219942) vector and transformed into *P. tricornutum* as described previously (Zaslavskaia et al. 2000; Sommer et al. 2007). For the self-assembling GFP assay, TPT-encoding sequences were cloned together with the large *gfp* fragment (GFP1-10) into the second multiple cloning site (MCSII) of the pPha_DUAL_2xNR (NCBI accession number: JN180664) vector using *Eco*RI, *Bam*HI, and *Hind*III restriction sites. The small *gfp* fragment (GFP11) was fused to the 3′-end (*Bam*HI) of respective gene sequences encoding marker proteins of the complex plastid subcompartments in the MCSI using *Spe*I and *Sac*II restriction sites. The marker proteins included the full-length sequence of the ER-resident protein disulphide isomerase (PDI, Phatr2 protein identification number (ID): 44937) (Hempel et al. 2009), the BTS of the PPC-localized heat shock protein 70 (sHsp70, 55890) (Hempel et al. 2009), the full-length sequence of the intermembrane space (IMS) located monogalactosyl diacylglycerol synthase 1 (MGD1, 21505) (Bullmann et al. 2010) and the BTS of the plastid localized ATPase gamma subunit AtpC (50958) (Hempel et al. 2009). After transformation, protein expression was induced by culturing positive clones in liquid f/2 medium containing 0.89 mM nitrate (NO_3^-) as sole nitrogen source for 6 h (Hempel et al. 2009) prior analysis via confocal microscopy (see later). Negative clones—that is, clones for which no fluorescence signal was detectable—were investigated for correct genomic integration of the fusion sequences using colony-polymerase chain reaction (PCR).

Confocal Laser Scanning Microscopy

Transformed *P. tricornutum* cells were screened for GFP fluorescence using a Leica TCS SP2 confocal laser scanning microscope with an HCX PL APO 40×/1.25–0.75 Oil CS objective. Excitation of eGFP and chlorophyll fluorescence occurred at 488 nm with a 65 mW Argon laser, whereas fluorescence emission was detected at a bandwidth of 500–520 nm for eGFP and 625–720 nm for chlorophyll (plastid autofluorescence), respectively.

Cell Subfractionation, Western Blot, and Immunodetection

To investigate the membrane integration of overexpressed TPT-eGFP fusion proteins, 150 ml of culture was harvested

(1,500 × g, 22 °C, 5 min) and resuspended in 4 ml solubilization buffer (50 mM NaCl, 50 mM imidazole/HCl ph 7.0, 2 mM 6-aminohexanoic acid, 1 mM ethylenediaminetetraacetic acid (EDTA), 8.5% (w/v) sucrose, protease inhibitor cocktail 1:200). Cell disruption via one passage in a French press followed with an applied pressure of 1,000 psi, whereupon intact cells were removed by centrifugation (1,500 × g, 4 °C, 5 min). Cell subfractionation with the remaining protein extract was then performed by ultracentrifugation (120,000 × g, 4 °C, 1 h) to separate soluble, associated and membrane proteins. The supernatant containing soluble proteins was removed, and associated and membrane proteins were again separated by ultracentrifugation (see earlier) after incubation for 30 min in carbonate buffer (100 mM $NaHCO_3$, pH 11.5, protease inhibitor cocktail 1:200). The resulting membrane pellet was dissolved in solubilization buffer and protein fractions were precipitated with trichloroacetic acid (10% v/v). All fractions were loaded on a sodium dodecyl sulfate (SDS)-gel to equal amounts, separated, and analyzed by Western blot and immunodetection using antibodies against GFP (detection of TPT-GFP fusion proteins, 1:3,000), RbcL (control soluble, 1:7,500), and PsbD (control membrane integral, 1:7,500).

Ortholog Prediction

To predict *P. tricornutum* TPT orthologs among the 41 annotated genomes listed in supplementary table S4, Supplementary Material online, all proteins were screened with hmmer in an identical procedure to that described earlier for the *P. tricornutum* TPTs. To further reduce the set of putative TPT orthologs additional sorting criteria were applied on the filtered domtblout file. Only proteins that showed a TPT domain model coverage higher than 50% and a Nuc_sug_transp domain model coverage lower than 50% were retained. The resulting proteins were used to build a BLAST database and an all-against-all BLAST search was performed using blastp+ version 2.2.30 (Camacho et al. 2009) with an e-value of 10 by using the BLOSUM62 matrix. The resulting blastp results were then filtered to account for the twilight zone of ortholog prediction by applying a changed version of formula (2) as indicated by (Rost 1999) with an additional coverage cutoff of 50% (query/subject length divided by alignment length). These filtered blastp results were then used with proteinortho version 5.11 (Lechner et al. 2011) to detect coorthologs within and between these species by using the following options [-e=1e-05; -identity=11; -cov=50; -conn=0.1; -sim=0.75; -selfblast]. The proteinortho results were filtered for coortholog groups containing putative *P. tricornutum* TPT proteins (supplementary table S1, Supplementary Material online). To account for gene prediction artifacts the resulting proteins were filtered for highly similar sequences with cd-hit version 4.6.1-2012-08-27 (Li and Godzik 2006) using a cutoff of 95% and only one of the similar proteins was kept for phylogenetic analysis.

Phylogenetic Analysis

Multiple sequence alignments (MSAs) were calculated for the full-length *P. tricornutum* proteins listed in supplementary table S1, Supplementary Material online (MSA1.fa), for the full-length protein set of TPT coorthologs from 41 annotated genomes (MSA2.fa) and for a subset of them (MSA3.fa) with MAFFT E-INS-i version 7.037b (Katoh and Standley 2013) with default settings (for MSA data and all *P. tricornutum* sequences used in phylogenetics see supplementary file, Supplementary Material online). The individual MSA was visualized and manually curated using Jalview version 2.8 (Waterhouse et al. 2009). For all curated MSAs the LG + G4 + I + F model was selected as the best fitting amino acid substitution model according to Bayesian Information Criterion in ProtTest version 3.3 (Darriba et al. 2011). To reconstruct the phylogeny we used the Neighbor-Joining (NJ) method implemented in MEGA6 (Tamura et al. 2013), Bayesian phylogenetic inference implemented in MrBayes version 3.2.4 (Ronquist and Huelsenbeck 2003) and the maximum-likelihood (ML) method implemented in RAxML version 8.1.15 (Stamatakis 2014) and IQ-TREE version 1.34 (Nguyen et al. 2015).

For the NJ method, an optimal tree was initially computed using the JTT matrix-based method by modeling rate variation among sites with a gamma distribution and a shape parameter of one removing all ambiguous positions in a pairwise fashion. A bootstrap consensus tree was inferred by 1,000 replicates and used to show the percentage of replicate trees in which the associated taxa from the optimal tree clustered together.

For MrBayes, two runs of four Markov-chain Monte Carlo (MCMC) chains of 2×10^7 generations each from a random starting tree, sampling every 1,000 generations [settings: rates = invgamma, ngammacat = 4, aamodelpr = LG] were initiated. A 25% burn-in was chosen and convergence was assessed by standard deviation of split frequencies falling below 0.01.

For RAxML, initially the best ML tree among 20 generated trees with the best likelihoods was determined using the PROTGAMMAILGF model [settings: -d -# 20]. Next, we conducted 1,000 non-parametric bootstrap inferences with the rapid hill-climbing mode using the PROTGAMMAILGF model [settings: -d -b -#1000]. The bootstrap replicates were used to draw bipartitions on the best ML tree aforementioned and to build a consensus tree applying the extended-majority rule option (-m PROTGAMMAILGF -J MRE).

For IQ-TREE, a ML tree was generated and the branch support values were assessed with the ultrafast bootstrap approximation (Minh et al. 2013) and with single branch tests (Adachi and Hasegawa 1996; Guindon et al. 2010) using

1,000 replicates each [settings: -bb 1000 -alrt 1000 -lbp 1000].

Phylogenetic trees were rooted either by midpoint or by the sequences fgenesh1_pg.C_scaffold_62000062 from *Phytophthora ramorum* and estExt_fgenesh1_pg.C_110114 from *Phytophthora sojae*, both of which are related to diatoms but lack a plastid. Figtree version 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/, last accessed October 19, 2015) was used to visualize and Adobe Illustrator CS5 to postprocess the tree images.

Topology Tests

To determine whether trees were significantly different from each other, AU (Shimodaira 2002) and SH tests (Shimodaira and Hasegawa 1999) were performed with CONSEL v0.20 (Shimodaira and Hasegawa 2001). Site likelihoods for each tree were calculated with tree-puzzle version 5.2 (Schmidt et al. 2002) with the options to estimate amino acid frequencies from the data and using Gamma distributed rates with four categories and alpha parameter estimated from the data set (supplementary table S5, Supplementary Material online).

Results

Bioinformatic Analysis

Nine putative TPT genes (*TPT1, 2, 4a, 4b, 5, 6, 8, 10, 11*) were identified in the genome of the diatom *P. tricornutum* by BLAST searches using the *T. gondii* TgAPT as an initial query. Additional genes encoding TPT-family domains were detected via hmmer searches and were grouped by subsequent phylogenetic analysis resulting in 13 putative TPT genes (*TPT1-13*, supplementary fig. S1 and table S1, Supplementary Material online). Bioinformatic analyses relvealed the presence of six to ten potential TMDs for all TPTs and clear prediction of putative N-terminal targeting signals for TPT1, TPT2, TPT4a, TPT4b, TPT8, and TPT10 (see Materials and Methods and supplementary table S2, Supplementary Material online). Whereas a potential complex plastid targeting signal—a BTS consisting of a classical signal peptide (SP, mediates ER import) and a transit peptide-like sequence (TPL, mediates transport across the PPM and beyond) (Stork et al. 2013)—could be detected for three of the proteins (TPT2, TPT4a, TPT4b), only a classical SP could be predicted for TPT1, TPT8, and TPT10 (table 1 and supplementary table S2, Supplementary Material online). In diatoms, a plastid-specific BTS is characterized by the presence of an ASA-FAP motif, located at the SP/TPL-border, which is typical for proteins translocated across all four plastid-surrounding membranes in these organisms (Gruber et al. 2007). Such a motif was found in TPT4a and TPT4b, while in TPT2 a non-ASA-FAP BTS type was detected, suggesting targeting of TPT2 to the PPC/PPM (Gould et al. 2006; Gruber et al. 2007). For the latter, the prediction is correct, as shown in a recent study by *in vivo* expression and localization of a full-length TPT2-GFP fusion protein (Lau et al. 2015). With much lower probability, the presence of potential targeting sequences was furthermore predicted for TPT9, TPT11 (potential SP, respectively), and TPT13 (potential signal anchor or SP), whereas for TPT5-7 and TPT12 no characteristic N-terminal targeting sequences could be detected (supplementary table S2, Supplementary Material online).

A set of selected transporters and their corresponding Pfam domains were used to screen the *P. tricornutum* genome and to highlight their domain architecture under different detection limits (TPT, PF03151; Nuc_sug_transp, PF04142; UAA, PF08449; and EamA, PF00892; see Materials and Methods, supplementary table S1 and fig. S1*H*, Supplementary Material online). All putative translocators exhibit a conserved TPT or TPT-like domain (PF03151), whereas an additional EamA or EamA-like domain (PF00892) was found upstream of, or overlapping with, the TPT-family domain (TPT1-7, TPT9-13) and an additional UAA or UAA-like domain (PF08449) was found either overlapping (TPT1, TPT4a, TPT4b, TPT9-11, TPT13) or spanning the TPT and EamA domains (TPT2, TPT7; see supplementary fig. S1*H*, Supplementary Material online). The variety of domain architectures only partially reflects the phylogenetic grouping most likely due to Pfam domain profile similarity with EamA or EamA-like domains present in most of the putative transporters investigated.

Localization Studies

To further elucidate the subcellular localization of the *P. tricornutum* TPTs, *in vivo* localization studies with the nine initially detected candidates (see earlier) were performed using presequence-only and full-length TPT-eGFP fusion proteins. Figure 1 shows the results for TPT1, TPT2, TPT4a, TPT4b, and TPT8, which appear to be targeted to the complex plastid of *P. tricornutum*, whereas the TPT5-6 and TPT10-11 constructs resulted in localizations distinctly outside the complex plastid (supplementary fig. S2, Supplementary Material online). No localization studies were performed with TPT7, TPT9, and TPT12-13, for which, like TPT5-6 and TPT11, evidence for the presence of cleavable N-terminal targeting signals (SP or BTS) was sparse (supplementary table S2, Supplementary Material online).

For TPT1, expression of both the presequence (SP) and full-length eGFP fusion protein resulted in a localization in the (c)ER or (c)ER membrane (including the nuclear envelope), respectively. A similar result was obtained for the presequence of TPT8 fused to eGFP (expression of two different lengths with 38 and 63 amino acids presequence fused to eGFP, see also supplementary fig. S3, Supplementary Material online), whereas the full-length protein directed GFP into the so-called "blob-like structure," the hallmark of a PPC or PPM localization (Kilian and Kroth 2005). It is thus not clear whether TPT8 localizes to the (c)ER or PPM. A typical blob-like GFP localization structure was also observed upon expression of the

Table 1

Predicted TPT Presequences in *Phaeodactylum tricornutum*

Name[a]	Putative Presequence[b]
TPT1	MSSSTAKDSGAASPLKLFVLVVCWYAGNT*FYNIY*...
TPT2	**MRFAAWLVILTGTFVEA**DRVQSTQP*QSASN*...
TPT4a	**MMMKRALVVLTLSVGVSARASA**FAPGAAVKNHAGATQSAIHKQTPFPTTELEKLRPQ*TSLAL*...
TPT4b	**MKVATTLTLAFICCASAFGL**NGQTTSVMKKVGFDAGSKPMVQAIDVQGNRLGSNMQPLKSAVANEDAP*RGGAT*...
TPT8	**MLVIAILLTFVVHFSTA**FPPTV*SFGSS*...
TPT10	**MAQTKASQSSTLWLLVWMVNNIGVTLL***NKAAF*...

[a]Name refers to given names of the TPT proteins studied here. For model names and IDs see Materials and Methods and supplementary table S2, Supplementary Material online.

[b]bold, SP prediction by hidden Markov models (HMM) (SignalP 3.0); underlined, SP prediction by neural networks (NN) (SignalP 3.0); unchanged letters, putative TPL sequence (N-terminal extension prior to conserved protein part, blastp); italics, conserved mature protein; highlighted in gray,

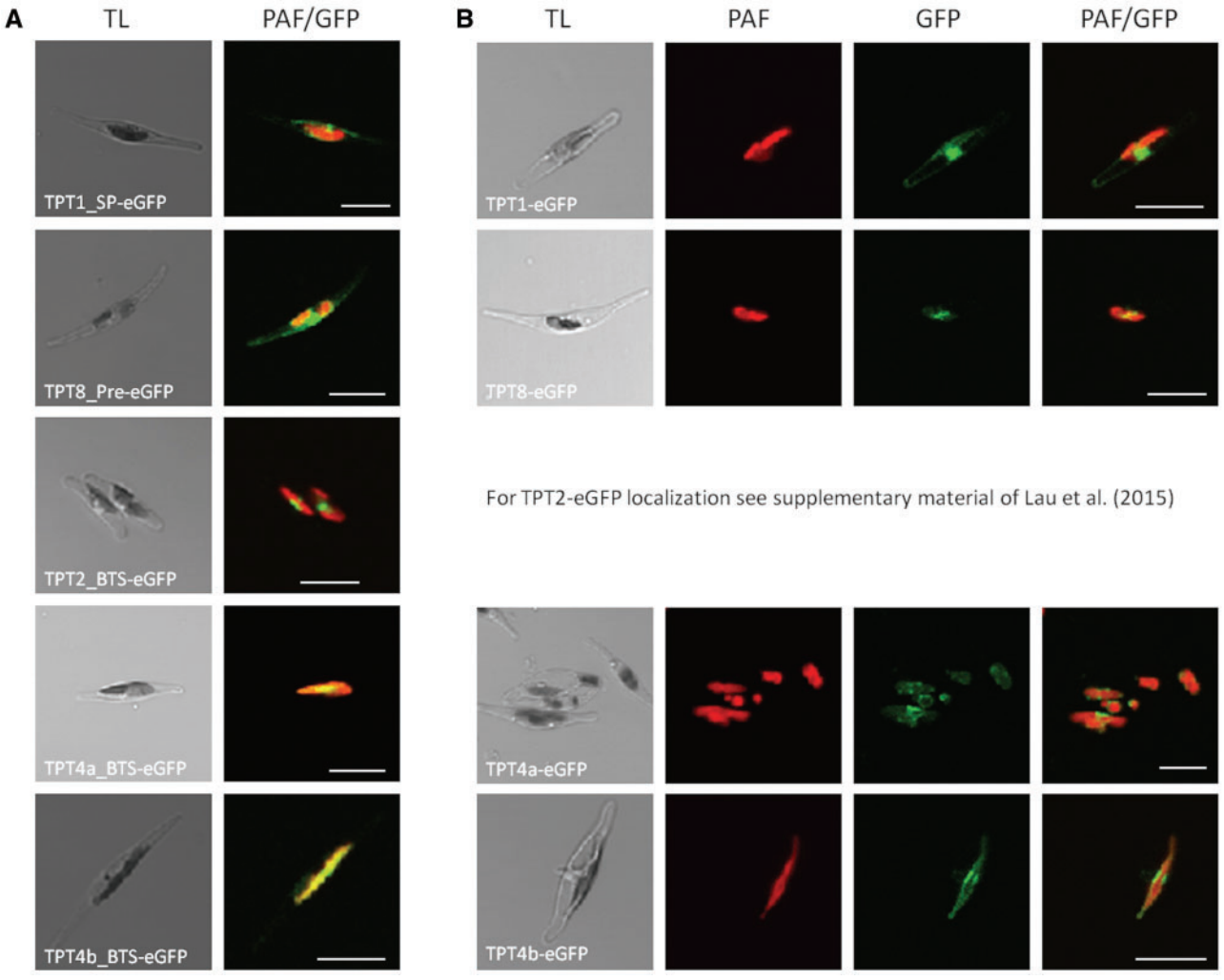

Fig. 1.—Homologous expression of presequence (*A*) and full-length (*B*) TPT-eGFP fusion proteins in *Phaeodactylum tricornutum*. eGFP fusion protein expressions indicated TPT localizations in the (c)ER membrane (TPT1), the PPM/PPC (TPT2; see also (Lau et al. 2015)), as well as the plastid OEM/IEM (TPT4a, TPT4b). Localization of TPT8 fusions led to indistinct results ((c)ER or PPM). TL, transmitted light; PAF, plastid autofluorescence; GFP, enhanced green fluorescent protein; PAF/GFP, overlay of plastid and GFP fluorescence; SP, signal peptide; BTS, bipartite targeting signal; Pre, presequence; scale bar represents 10 μm.

putative BTS as well as the full-length sequence of TPT2 fused to eGFP, as shown here for the TPT2_BTS-eGFP fusion (fig. 1; for full-length TPT2-eGFP PPM localization see Lau et al. (2015)). For two TPTs, TPT4a and TPT4b, expression of the BTS-eGFP fusion sequences containing the plastid targeting determinating conserved ASA-FAP motif at the SP/TPL-border (Gruber et al. 2007) (table 1) resulted in a stromal GFP localization structure that was consistent with autofluorescence of the plastid. As expected, expression of the full-length eGFP fusion proteins of TPT4a and TPT4b resulted in more or less concentrated fluorescence signals directly surrounding the plastid autofluorescence (fig. 1). Although the TPT4a and TPT4b data are consistent with a plastid localization for both full-length fusion proteins, the precise membrane targeted (i.e., inner or outer plastid envelope membrane, IEM/OEM) cannot be inferred from these results.

Analysis of Target Membranes and Orientations of the TPTs

To determine whether TPT4a and TPT4b are IEM- or OEM-proteins, we took advantage of the self-assembling GFP system (Cabantous et al. 2005). This assay, previously established for *P. tricornutum* (Hempel et al. 2009; Bullmann et al. 2010; Lau et al. 2015), is based on the discrete expression of two GFP fragments (β strand 1–10 and β strand 11) that in isolation are not able to fluoresce. If localized to the same compartment, however, the high affinity of both fragments for each other results in reconstitution of the fluorophore and detectable fluorescence.

To test the practicality of the self-assembling GFP system for the complex plastid TPTs of *P. tricornutum*, we first investigated the (c)ER-localized TPT1 (fig. 1). A fusion protein consisting of the full-length sequence of TPT1 and the small self-assembling GFP fragment GFP_S11 was simultaneously expressed with a cytosolically localized large GFP_S1-10 fragment. As shown in figure 2, this resulted in a (c)ER-characteristic GFP fluorescence, indicating a C-terminal orientation of the fusion protein toward the cytosplasm of the host cell. As a negative control the same TPT1 fusion protein was expressed with an ER-targeted PDI-GFP_S1-10 (PDI) fusion construct, resulting in absence of a detectable fluorescence signal (fig. 2). As further proof of principle, we deleted the last predicted TMD (C-terminus) of TPT1 (TPT1_TMD-) and again expressed the truncated construct as a fusion with C-terminal GFP_S11, together with the ER-marker GFP_S1-10 fusion protein (PDI-GFP_S1-10), which did not fluoresce with the full-length TPT1 version. As shown in figure 2, this combination induced a recovery of fluorescence, indicating that the system is useful for studying TPT orientation and topology.

To prove the localization and orientation of the two plastidal TPTs, either TPT4a or TPT4b was expressed with a C-terminal GFP_S1-10 fusion together with a plastid stroma targeted GFP_S11, as well as an IMS (space between OEM and IEM) directed GFP_S11, respectively (see Materials and Methods). Whereas simultaneously expression of both individual fusion proteins with the stromal marker fused to GFP_S11 resulted in GFP fluorescence surrounding the plastid autofluorescence, no signal was detectable in combination with the IMS marker (fig. 2). This shows that TPT4a and TPT4b are localized in the IEM with the C-terminus of both proteins facing the plastid stroma.

Similar analyses were also performed for TPT2 and TPT8. However, due to limitations of the system, especially concerning topological studies of membrane spanning PPM proteins (Hempel et al. 2009), it was not possible to determine the exact orientation (and localization in case of TPT8) of the putative translocators. Fluorescence signals were obtained upon expression of the full-length GFP_S11 fusion proteins with both the ER- and the PPC-marker fused to GFP_S1-10. A similar observation was reported in an earlier study attempted to determine the topology of the PPM-localized, symbiontic sDer1 proteins (truncated versions) (Hempel et al. 2009).

Analysis of Membrane Integration

To investigate whether the complex plastid localized TPT-eGFP fusion proteins are actually integral membrane proteins, cellular subfractionations including carbonate extraction were performed (see Materials and Methods). As shown in figure 3, the majority of the expressed TPT-eGFP fusion proteins (TPT1/4a/4b) could be detected in the integral membrane fraction. Nevertheless, signal could also be detected in the soluble fraction, which might be due to the light induced overexpression of the fusion proteins. A similar observation was made for the abundant control protein of the integral fraction, the thylakoid membrane localized photosystem II component PsbD (fig. 3). The membrane integration of the PPM-localized TPT2-eGFP was confirmed previously in Lau et al. (2015).

Phylogeny

Molecular phylogenetics has been used to elucidate the origin and evolution of putative plastid targeted TPTs in organisms with complex red plastids (Weber et al. 2006; Price et al. 2012; Karkar et al. 2015). Such analyses have pointed to a monophyletic, endosymbiont-derived origin of the TPTs, predicted to be plastid localized in all investigated CASH lineages. However, due to the lack of experimental evidence for the localization of the majority of proteins included in these data sets, with exception of individual apicomplexan TPTs, and the lack of reliable targeting signal predictions in these organisms, the actual target membrane for most of the predicted complex plastid TPTs remained unclear. Additionally, previous published data sets often included only a subset of putative TPTs, in particular for diatom sequences (Weber et al. 2006; Price et al. 2012; Karkar et al. 2015).

With the goal of elucidating the evolutionary origin of the *P. tricornutum* TPTs localized in this study, we performed

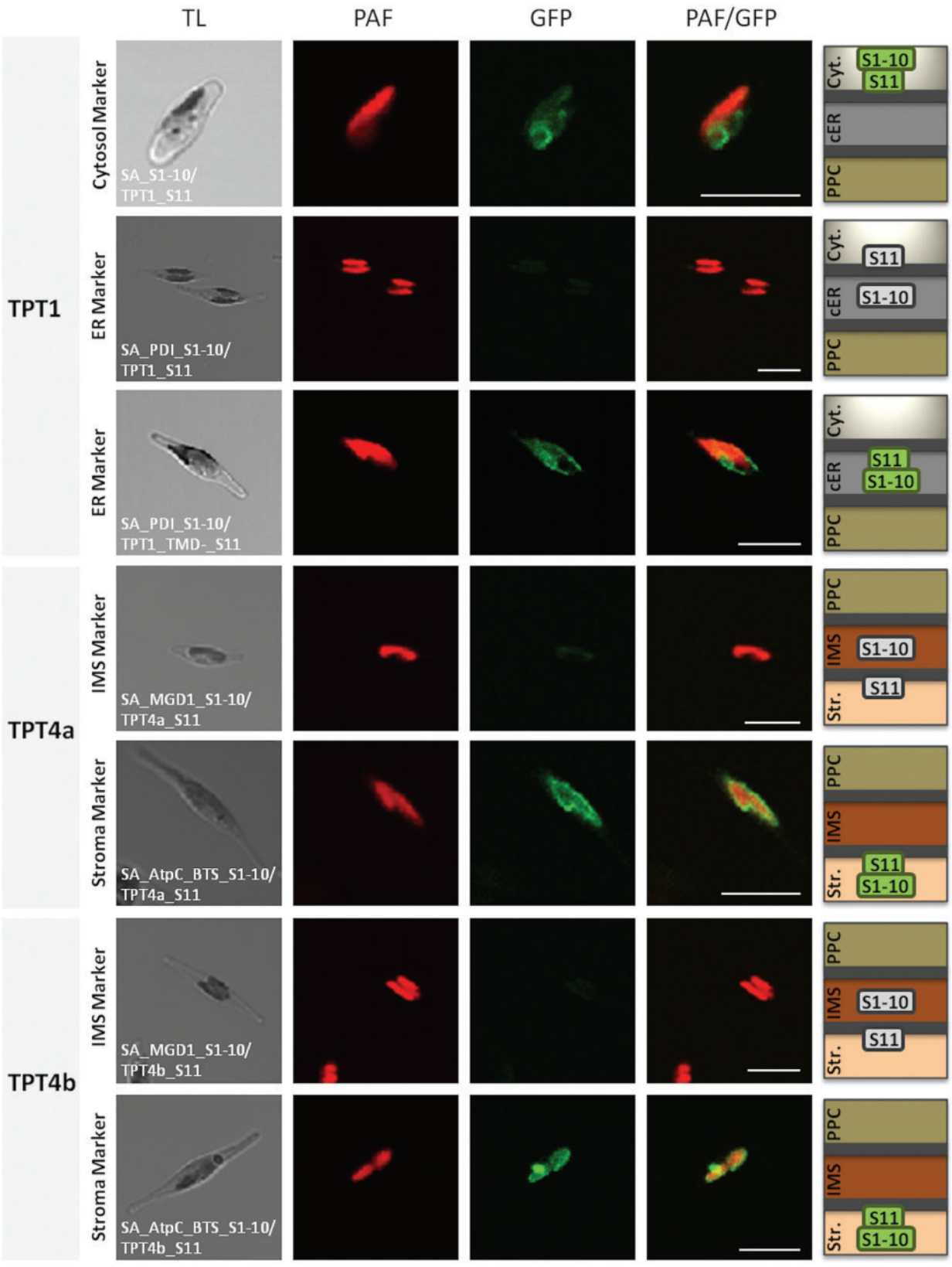

Fig. 2.—Self-assembling GFP analysis of complex plastid TPTs in *Phaeodactylum tricornutum*. Simultaneous expression of TPT1 fused to GFP_S11 with a cytosolic GFP_S1-10 led to a (c)ER-characteristic fluorescence pattern. No fluorescence could be observed after expression of TPT1_S11 with the ER marker

(continued)

phylogenetic analyses using distance, Bayesian and ML methods. Our analyses consistently place the 13 TPT-family domain-containing proteins of *P. tricornutum* (supplementary fig. S1 and table S1, Supplementary Material online) into four different groups (fig. 4*A*). However, consistent with previous results (Weber et al. 2006; Price et al. 2012), only one of the four putative TPT groups in *P. tricornutum* branches with the pPTs of photosynthetic organisms (fig. 4, supplementary figs. S4 and S5, Supplementary Material online). TPT1, TPT2, TPT4a, and TPT4b, the putative TPTs which reside in the membranes surrounding the *P. tricornutum* complex plastid, cluster with homologous sequences from all other CASH lineages including the complex plastid localized TPTs of the apicomplexans *P. falciparum* and *T. gondii* (fig. 4). In agreement with previous studies, the tree suggests a monophyletic origin of these proteins with the TPTs of CASH lineages and red algal plastid localized TPTs, which is supported by three different phylogeny reconstruction methods (fig. 4, supplementary table S5, Supplementary Material online). Additionally, TPT5-8 is also present in this group (fig. 4). However, in contrast to TPT1, TPT2, TPT4a, and TPT4b, the most likely non-complex plastid localized TPT5-7 (supplementary fig. S2 and table S2, Supplementary Material online) and the (c)ER membrane or PPM-targeted TPT8 (fig. 1) are located at the very base of this group. TPT5-8 tends to group together with PPT1/PPT2 proteins of red algae, plants, and green algae, lacking any homologs in other CASH lineages except for related diatoms (fig. 4, supplementary figs. S4 and S5, Supplementary Material online).

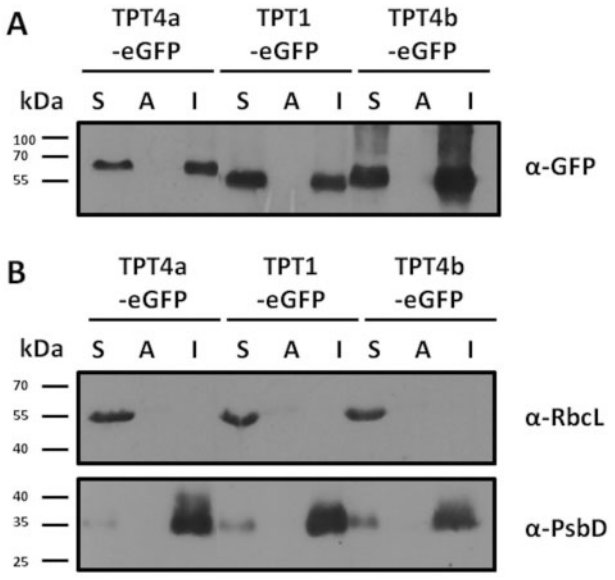

Fig. 3.—Analysis of the membrane integration of homologous expressed TPT-eGFP fusion proteins in *Phaeodactylum tricornutum*. (*A*) Probing over-expressed TPT1/4a/4b-eGFP fusion proteins with a GFP-antibody after cell subfractionation and carbonate extraction led to detection of dominant signals in the integral protein fraction, respectively. However, a weaker signal could additionally be detected in the soluble fraction, which was also the case for the control integral membrane protein (see *B*). (*B*) As a control for soluble and integral membrane proteins, protein extracts of TPT1/4a/4b-eGFP expressing cultures were probed with antibodies against the large subunit of RuBisCO (RbcL, soluble) as well as the thylakoid membrane localized photosystem II component PsbD (integral), respectively. kDa, molecular mass in kilodalton; S, soluble fraction; A, associated fraction ; I, integral fraction.

Discussion

The transformation of a free-living cell into an obligate endosymbiont or organelle involves manipulation of the integrated organism via factors evolved by the host. With respect to photosynthetically active endosymbionts or organelles (i.e., plastids), transport of energy-rich carbohydrates produced by the endosymbiont into host compartments is one of the major targets for manipulation. Current evidence suggests that in a common ancestor of red and green algae, a nucleotide sugar translocator gene of host origin was duplicated, acquired the ability to encode the relevant targeting signals, and the resulting protein evolved into the pPT (Weber et al. 2006) providing the host cell with access to the energy produced by the endosymbiont's photosynthesis machinery. However, this essential invention might not have been sufficient, as recent findings suggest that genes of bacterial origin might also have contributed significantly to the buildup of the initial metabolic connection during primary endosymbiosis (Karkar et al. 2015).

In the case of secondary plastids, the initial steps toward establishing a metabolic connection between host and endosymbiont were presumably more complex, as the cytoplasm of the host cell and the stromal compartment of the plastid are separated from one another by one or two additional membranes. How did complex plastid-bearing eukaryotes solve this problem? Evolution of the "tools" for metabolic coupling, that is, targeting of transporters into both endosymbiont

Fig. 2.—Continued

(PDI) fused to GFP_S1-10. A truncated version of TPT1, lacking the last predicted C-terminal TMD (TPT1_TMD-), fused to GFP_S11 and expressed with the ER-localized PDI-GFP_S1-10 resulted in a (c)ER-characteristic GFP fluorescence. Both, TPT4a and TPT4b, fused to GFP_S11 did not show a fluorescence signal when expressed with the IMS marker MGD1 fused to GFP_S1-10, whereas a clear GFP signal circling the plastid autofluorescence could be observed upon expression with the stromal-targeted AtpC_BTS-GFP_S1-10, respectively. TL, transmitted light; PAF, plastid autofluorescence; GFP, enhanced green fluorescent protein; PAF/GFP, overlay of plastid and GFP fluorescence; BTS, bipartite targeting signal; Cyt., cytosol; cER, chloroplast endoplasmic reticulum; PPC, periplastidal compartment; IMS, intermembrane space; Str., plastid stroma; scale bar represents 10 μm.

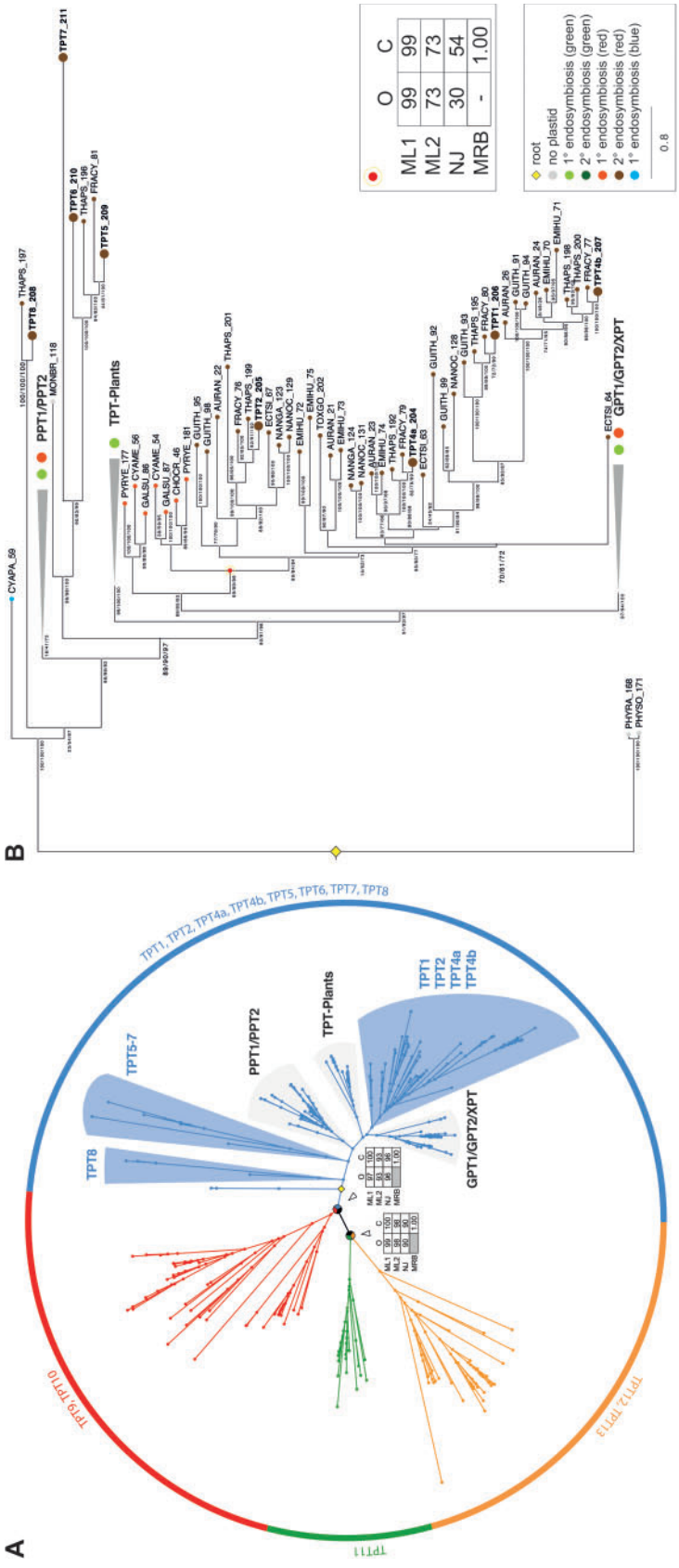

Fig. 4.—Phylogeny of eukaryotic TPTs with focus on pPTs. (*A*) The tree shown is a rooted consensus tree inferred by ML (IQ-TREE) and 1,000 bootstrap replicates separating putative *Phaeodactylum tricornutum* TPT coorthologs into four subgroups (blue, red, green, and orange ring). Major splits are highlighted and branch support values are given for two ML implementations (ML1: IQ-TREE; ML2: RAxML [percentage bootstrap support]), one NJ method [percentage bootstrap support] and one Bayesian inference method (MRB: MrBayes [posterior probabilities]) for either the optimal (O) or consensus (C) tree. Subgroups of the blue group show the placement of the individual *P. tricornutum* putative TPT sequences (light-blue area) and the plant related subgroups PPT1/PPT2, TPT-plants, and GPT1/GPT2/XPT (light-gray area). (*B*) Rooted optimal ML tree (IQ-TREE) of the blue group. All four complex plastid localized TPTs (TPT1/2/4a/4b) of *P. tricornutum* cluster with translocators of other CASH lineages and have a monophyletic origin with red algal TPTs with the split highlighted and branch support values given separately (yellow-red ring, abbreviations as in (*A*), for sequence aliases see supplementary table S6, Supplementary Material online). Numbers at nodes represent bootstrap support values and two independent single branch support test values (bb/alrt/lbp; see Materials and Methods). The phylogenetic plastid affiliation per taxon is indicated with different circle colors (gray, no plastid; blue, glaucophytes (primary endosymbiosis); light-green, green algae (primary endosymbiosis); green, organisms resulting from secondary endosymbiosis involving a green alga; orange, red algae (primary endosymbiosis); brown, organisms resulting from secondary endosymbiosis involving a red alga). Note that the whole tree shown here is a subgroup (blue group) of a larger tree (see (*A*) and supplementary fig. S4, Supplementary Material online). Scale bar indicates inferred number of amino acid substitutions per site. For both (*A*) and (*B*) two sequences from the genus *Phytophthora* (plastid-lacking stramenopiles) were used to root the trees (yellow diamonds). See supplementary material, Supplementary Material online for complete trees as well as additional trees inferred using different methods.

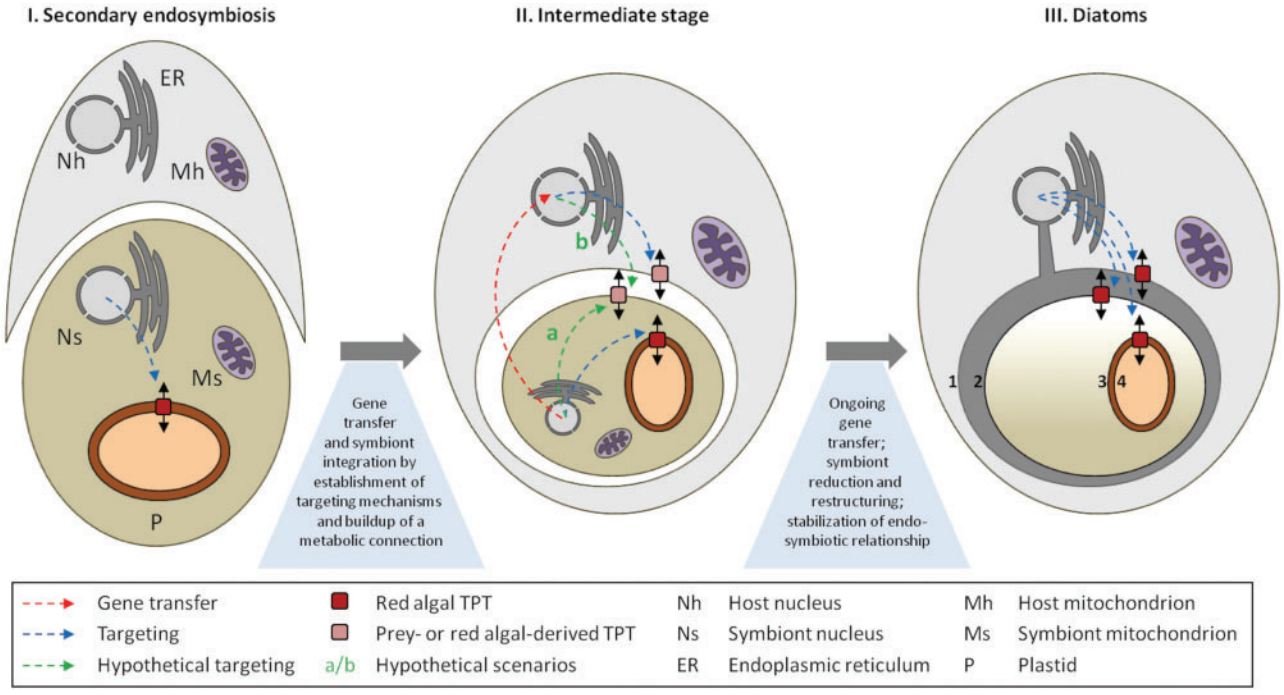

Fig. 5.—Model for the development of an initial metabolic connection of host and red algal endosymbiont during secondary endosymbiosis leading to the evolution of extant diatoms (stramenopiles). (*I*) A single-celled red alga was phagocytosed by a eukaryotic host cell but not digested. (*II*) To profit from the photosynthetic capacity of the engulfed autotrophic cell—the later endosymbiont—the host had to find a way to access profitable metabolites (e.g., energy-rich carbohydrates). To achieve this, two hypothetical scenarios are discussed that putatively have occurred in an intermediate stage of endosymbiosis: (*a*) Besides a transporter (TPT) encoded in the symbiont's nucleus and targeted to the symbiont's primary plastid, the endosymbiont transiently targeted a similar, probably duplicated factor, to its plasma membrane, maybe via its secretory system, which was not reduced and still fully functional at this time. Gained via endosymbiotic or earlier HGT, a prey- or red algal-derived transporter was targeted into the phagotrophic/perialgal vacuole via the hosts secretory system, thereby providing the host access to the metabolites produced by the endosymbiont. (*b*) Considering exactly the same situation for the phagotrophic/perialgal vacuole as described in scenario (*a*), in addition to that, the host cell already found a way to target an additional, probably duplicated, prey- or red algal-derived transporter beyond the phagotrophic/perialgal vacuole into the endosymbionts plasma membrane—most likely enabled by the early evolution of a protein transport system across the second outermost membrane surrounding the symbiont. It is possible that scenarios (*a*) and (*b*) occurred as a sequence of connected events. (*III*) Gene transfer (and duplication) as well as reduction and restructuring of the endosymbiont led to the situation found in extant diatoms. All TPTs located in the symbiont-surrounding membranes are of red algal origin and encoded in the nucleus of the host cell. Numbers from 1–4 refer to the individual membranes of the complex plastid. For discussions on the origin of the complex plastid membranes see Gould et al. (2015).

and host-derived plastid-surrounding membranes concurrently, is a complicated process and is unlikely to have originated several times independently. New insights could provide evidence in support of a monophyletic origin of secondarily evolved organisms or, contrarily, for their polyphyletic origin or their evolution via serial endosymbiosis.

To address the puzzle of complex plastid evolution, which is still hotly debated (Petersen et al. 2014; Stiller et al. 2014; Gould et al. 2015), we have provided data on the subcellular localization and evolution of triose phosphate transporters (TPTs) essential for directional transport of energy-rich C-compounds out of photosynthetically active or into secondarily non-photosynthetic plastids. Using diatoms as a model system, we show that the TPTs, located either in the endosymbiont-derived inner membrane (IEM) or in one of the two outer membranes of the diatom's complex plastid, are of red algal origin (TPT1/2/4a/4b; fig. 4); all of them are most likely the product of gene duplications starting with the red algal (endosymbiont) transporter gene. In addition, our studies suggest that the TPT phylogeny as seen in diatoms is similar for all investigated organisms with a red algal-derived secondary plastid, including that of apicomplexan parasites. With respect to the subcellular localization, we present evidence for the presence of TPTs in three of four complex plastid-surrounding membranes with the exception of the OEM (figs. 1–3; see Discussion). Considering these results, we suggest that TPTs of a red algal endosymbiont were reused and integrated into membranes of both host and symbiont origin.

Given the presence and subcellular localization of red algal-derived TPTs in diatoms (and most likely all other CASH lineages), we propose the "minimal action model," which posits that gene transfer (and duplication) of the relevant TPT-

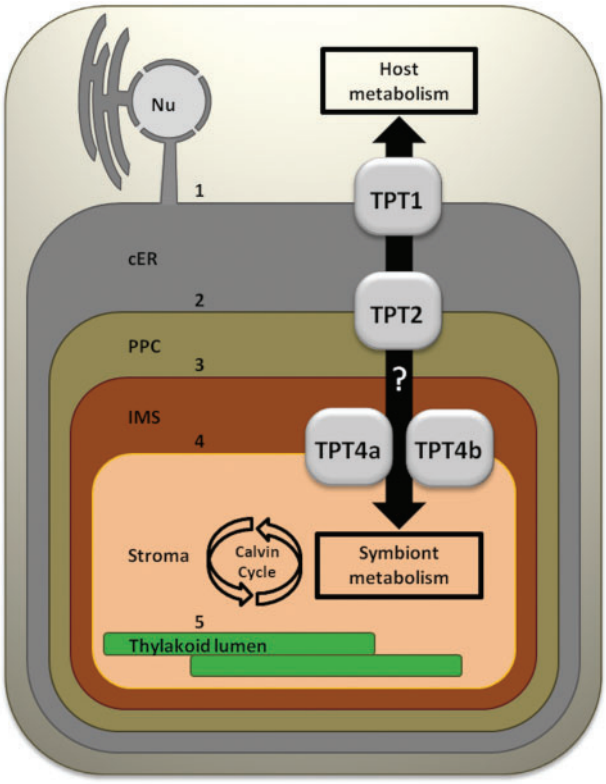

Fig. 6.—Model for the connection of host and endosymbiont metabolic pathways in *Phaeodactylum tricornutum*. Depicted is a schematic of a *P. tricornutum* cell with focus on the four membrane-bound complex plastid. With exception of the OEM (membrane 3), putative TPTs are present in all plastid membranes including the cER-membrane (1, TPT1), the PPM (2, TPT2) as well as the IEM (4, TPT4a and TPT4b). Energy-rich metabolites generated by photosynthesis (most likely in form of triose phosphates) can be directionally transported from the stroma to the cytoplasm of the host (or vice versa) to fuel host metabolic pathways or to supply energy storage. Carbohydrates might cross the OEM either by diffusion or transport through a hitherto undetected tranclocator (indicated by a question mark). The presence of two TPTs in the IEM might indicate a regulatory mechanism for substrate exit/import from the stroma. Note that the cER-membrane or PPM-localized TPT8 was not included into the model as its potential involvement into this process is unclear yet. Nu, nucleus; cER, chloroplast endoplasmic reticulum; PPC, periplastidal compartment; IMS, intermembrane space.

encoding genes from the endosymbiont nucleus to the host nucleus, as well as the evolution of a protein translocation system (at least in the PPM) occurred prior to the origin of a stable metabolic link between host and endosymbiont. What evidence speaks for and against this hypothesis, and are there alternative explanations?

One scenario to explain the evolution of TPTs in diatoms involves horizontal gene transfer (HGT) or predator-prey interactions (Larkum et al. 2007). Gained via HGT, a (endosymbiont-derived) transporter might have been present in the host before the endosymbiosis was established and (re-)direction to the endosymbiont could have contributed to triose phosphate leakage in the early phase of the symbiotic relationship, from which the host benefited. In that case, and if carbohydrate exchange was the main driving force for stabilizing the symbiosis, transport systems for proteins encoded by transferred endosymbiont genes probably had time to evolve later. However, considering the present-day membrane structure of complex plastids in diatoms and many other CASH lineages (four membranes separating the stroma from the host cytoplasm), it is not clear how proteins from the host might have been targeted beyond the outermost endosymbiont-surrounding membrane without an established protein transfer system connecting the two partners. This gap might have been bridged by the endosymbiont, for example, by providing translocators in an early stage of endosymbiosis (as it probably was the case for TPTs in the IEM) perhaps by occasional "wrong" dual targeting (Martin 2010; Gile et al. 2015), that is, before a protein transport system in the PPM was established, followed by subsequent replacement of endosymbiont genome-encoded factors by host-encoded ones as soon as host proteins were able to insert/traverse the PPM. Alternatively, the host directly found a way to target a symbiont-derived translocator beyond the outermost membrane most likely by developing a protein transfer system prior to the stabilization of the metabolic link (fig. 5). Thus, in the case of diatoms, the development of an efficient transport-system for symbiont-derived genes expressed by the host even might have been an essential prerequisite to the establishment of a metabolic connection.

Similar ideas were put forward in the so-called targeting-ratchet model (TRM (Keeling 2013)). The TRM postulates that a heterotrophic predator captured and digested algae, and at an early stage of association learned to retain the prey for longer periods and to profit from its photosynthetic capacity. Transporters evolved, and were targeted and integrated into the transient intracellular phototrophs. Finally, extensive gene transfer to the host nucleus and (re-)targeting of products to the endosymbiont eventually led to stabilization of the endosymbiotic relationship. As the source of the prey might have been phylogenetically diverse, transporters (and other molecules) could have had different evolutionary origins not ultimately related to that of the endosymbiont (Keeling 2013). Although the TRM does not provide an explanation for the establishment of the targeting mechanisms for proteins to cross, or to get inserted into, the individual membranes surrounding the endosymbiont, it is generally consistent with the situation found in diatoms. However, notably in the case of the diatom TPT1/2/4a/4b, the transporters and the symbiont both are of red algal origin (fig. 4), and thus, appear to come from the same source.

The localization of the diatom TPTs with respect to the origin of their target membranes speaks to the evolution of organisms with complex red plastids. Independent of the

different models of the origin of the plastid-surrounding membranes (Cavalier-Smith 2000; Melkonian 2001; Gould et al. 2015), the outermost membrane is generally thought to be a host-derived (i.e., non-endosymbiont) membrane. The origin of the PPM, the second outermost membrane, is still controversial; some suggest it is of endosymbiont origin (i.e., derived from the endosymbiont's plasma membrane (Cavalier-Smith 2000)), while others propose it is a host-derived ER (Zimorski et al. 2014; Gould et al. 2015) or autophagosomal-like membrane (Melkonian 2001). In diatoms, the outermost membrane, which researchers generally agree is of host origin, harbors a symbiont-derived TPT (TPT1; figs. 1 and 2). Given that TPTs were already present in the common ancestor of green and red algae (Weber et al. 2006), it would appear that the ancestor of diatoms (initial host cell) made use of already established transporters. What is special about this is not to recycle a transporter, but to recycle a transporter from a eukaryotic endosymbiont and integrate it into another eukaryotic phylogenetically different host and its endomembrane system. Therefore, the reuse of red algal endosymbiont-derived TPTs demonstrates that nature worked here in a parsimonious way and equipped a recycled protein with specific targeting signals to direct the product into (at least) one host membrane instead of making new inventions, which ultimately might have been redundant.

Although this "minimal action" scenario might help explain the origin of diatom plastids via secondary endosymbiosis, it does not formally exclude more complex evolutionary mechanisms such as tertiary or even quaternary endosymbiosis (Petersen et al. 2014; Stiller et al. 2014). However, due to massive structural reduction (membrane losses) and genomic reorganization of the endosymbiont, which would have been the consequences of such complex mechanisms, a repeated transfer of the original red algal TPT gene(s) to the individual tertiary or quaternary host genome as well as targeting signal acquisition would have been necessary. Furthermore, during each endosymbiotic event, regardless whether secondary, tertiary, or quaternary, the individual host cell would have been confronted with the same issue again and again: the metabolite transport across the one or two additional membranes separating the host and the endosymbiont cytoplasmic compartments during the initial stages of endosymbiotic association. This involves membranes like the plasma membrane of the endosymbiont and, as often thought, a host-derived phagotrophic/perialgal vacuole that were not been capable of transporting the relevant metabolites before. According to the situation found in diatoms, the host obviously would have solved this issue the same way consistently, namely by gene duplication, adaptation, and targeting of the pre-existing red algal endosymbiont TPT gene(s) and its/their product(s) (fig. 5). From an evolutionary perspective, this hypothetical sequence of events would represent a prime example of parsimony; however, it might be considered exceedingly rare. Would it nevertheless have had the potential to occur independently on several occasions, that is, during the evolution of other CASH lineages as well (assuming that they are not the product of a single secondary endosymbiosis)?

As demonstrated by several recent phylogenetic studies (Weber et al. 2006; Price et al. 2012; Karkar et al. 2015) including the analysis presented herein (fig. 4), complex plastid localized TPTs of CASH lineages and the plastidal TPT of red algae appear to have a monophyletic origin. Thus, not only diatoms but also all other complex algae with red algal-derived plastids included into these analyses possess symbiont-derived TPTs, which most likely function to connect the metabolic pathways of host and endosymbiont (see also studies on apicomplexans and cryptophytes; (Haferkamp et al. 2006; Mullin et al. 2006; Fleige et al. 2007; Karnataki et al. 2007; Brooks et al. 2010; Banerjee et al. 2012)). With respect to the mechanistic complexity and necessary sequence of events underlying the establishment and maintenance of a metabolic host-symbiont connection in these organisms, especially during an early coevolutionary state (e.g., gene transfer, evolution of targeting systems; see issues discussed earlier for the diatom model), the explanation of these results might favor a single secondary endosymbiotic origin of plastids in CASH lineages over a more complex evolutionary history involving several independent tertiary or even quaternary endosymbioses. However, although more difficult to explain and seemingly rather unlikely, the possibility that a gene encoding the symbiont's red TPT was uniformly transferred, similarly (re-) targeted, and therefore, efficiently reused several times independently, that is, in case the evolution of CASH lineages involved more complex scenarios than a single secondary endosymbiotic event, cannot be excluded. Based on our model, it would, however, have required the presence of a functional mechanism for the insertion of a transporter into the second outermost endosymbiont membrane. In case of all CASH lineages with complex plastids surrounded by four membranes, this transport system could have been provided through the earlier evolution of the symbiont-derived protein translocator in the PPM. This translocator, known as symbiont-specific ER-associated degradation (ERAD)-like machinery (SELMA), is itself a recycled system (Bolte et al. 2011) and a widely distributed hallmark feature of the aforementioned CASH lineages (Sommer et al. 2007; Stork et al. 2012). Taken together, the most straightforward interpretation of our data favors a single secondary endosymbiotic origin of plastids in CASH lineages with operation of a functional protein translocation system in the endosymbiont-surrounding membranes prior the buildup of a stable metabolic connection between host and endosymbiont.

Regardless of the actual evolutionary scenario of the establishment of a metabolic connection between host and endosymbiont, the data presented here show that at least four of the transporters (TPT1/2/4a/4b) in *P. tricornutum* can be localized to three out of the four complex plastid-surrounding membranes. This was demonstrated by complementary

presequence, full-length and self-assembling GFP localization studies (figs. 1 and 2) as well as cellular subfractionation (fig. 3). TPT1, TPT2, TPT4a, and TPT4b might represent a complete set of transporters for exit of phosphorylated carbohydrates from the plastid in extant diatoms (fig. 6). This assumption takes into account that the passage of these metabolites through the OEM is facilitated by other mechanisms (e.g., involving beta-barrel proteins not identified so far), most likely equivalent to carbohydrate transport through the outer plastid envelope membrane of primary plastids (Breuers et al. 2011). In contrast to the (c)ER-membrane and PPM, for which only one specific TPT was detected (whether TPT8 plays an additional role is unclear), two TPTs are localized in the IEM. This might enable a more controlled exit of substrates from the stroma, as both TPTs could possess different substrate specificities (e.g., DHAP and/or PEP), whereas TPT2 and TPT1 might be less selective. Furthermore, it is possible that one of the two IEM-localized TPTs acts as an "exporter," whereas the other one could function as an "importer" of C3-compounds (Mullin et al. 2006). To clarify this, it will be important to investigate the substrate specificities of the TPTs of the complex plastid membranes of *P. tricornutum* in future studies. However, it seems reasonable to assume that these translocators have similar characteristics to their homologs in cryptophytes and apicomplexans (Haferkamp et al. 2006; Mullin et al. 2006; Brooks et al. 2010), which connect symbiont and host metabolism by antiport of DHAP and PEP against inorganic phosphate.

Supplementary Material

Supplementary file, tables S1–S6, and figures S1–S5 are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was supported by the LOEWE program of the state of Hesse and the International Max Planck Research School for Environmental, Cellular, and Molecular Microbiology (IMPRS-Mic), Marburg. Daniel Moog and John Archibald acknowledge support from the Canadian Institutes of Health Research, the Canadian Institute for Advanced Research and Dalhousie's Centre for Comparative Genomics and Evolutionary Bioinformatics. The authors thank Franziska Hempel for kindly providing self-assembling GFP vectors.

Literature Cited

Adachi J, Hasegawa M. 1996. MOLPHY version 2.3 – Programs for molecular phylogenetics based on maximum likelihood. Computer Science Monographs, No. 28. Tokyo (Japan): Institute of Statistical Mathematics.

Archibald JM. 2009. The puzzle of plastid evolution. Curr Biol. 19:R81–R88.

Banerjee T, Jaijyan DK, Surolia N, Singh AP, Surolia A. 2012. Apicoplast triose phosphate transporter (TPT) gene knockout is lethal for *Plasmodium*. Mol Biochem Parasitol. 186:44–50.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 340:783–795.

Bolte K, et al. 2011. Making new out of old: recycling and modification of an ancient protein translocation system during eukaryotic evolution. Mechanistic comparison and phylogenetic analysis of ERAD, SELMA and the peroxisomal importomer. Bioessays 33:368–376.

Breuers FK, Brautigam A, Weber AP. 2011. The plastid outer envelope - a highly dynamic interface between plastid and cytoplasm. Front Plant Sci. 2:97.

Brooks CF, et al. 2010. The *Toxoplasma* apicoplast phosphate translocator links cytosolic and apicoplast metabolism and is essential for parasite survival. Cell Host Microbe. 7:62–73.

Bullmann L, et al. 2010. Filling the gap, evolutionarily conserved Omp85 in plastids of chromalveolates. J Biol Chem. 285:6848–6856.

Cabantous S, Terwilliger TC, Waldo GS. 2005. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nat Biotechnol. 23:102–107.

Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Cavalier-Smith T. 2000. Membrane heredity and early chloroplast evolution. Trends Plant Sci. 5:174–182.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165.

Deschamps P, et al. 2006. Nature of the periplastidial pathway of starch synthesis in the cryptophyte *Guillardia theta*. Eukaryot Cell. 5:954–963.

Douglas S, et al. 2001. The highly reduced genome of an enslaved algal nucleus. Nature 410:1091–1096.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol. 300:1005–1016.

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39:W29–W37.

Fleige T, Fischer K, Ferguson DJ, Gross U, Bohne W. 2007. Carbohydrate metabolism in the *Toxoplasma gondii* apicoplast: localization of three glycolytic isoenzymes, the single pyruvate dehydrogenase complex, and a plastid phosphate translocator. Eukaryot Cell. 6:984–996.

Flügge UI. 1999. Phosphate translocators in plastids. Annu Rev Plant Physiol Plant Mol Biol. 50:27–45.

Gile GH, Moog D, Slamovits CH, Maier UG, Archibald JM. 2015. Dual organellar targeting of aminoacyl-tRNA synthetases in diatoms and cryptophytes. Genome Biol Evol. 7:1728–1742.

Gould SB, Maier UG, Martin WF. 2015. Protein import and the origin of red complex plastids. Curr Biol. 25:R515–R521.

Gould SB, et al. 2006. Protein targeting into the complex plastid of cryptophytes. J Mol Evol. 62:674–681.

Gould SB, Waller RF, McFadden GI. 2008. Plastid evolution. Annu Rev Plant Biol. 59:491–517.

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

Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59:307–321.

Haferkamp I, et al. 2006. Molecular and biochemical analysis of periplastidial starch metabolism in the cryptophyte *Guillardia theta*. Eukaryot Cell. 5:964–971.

Hempel F, Bullmann L, Lau J, Zauner S, Maier UG. 2009. ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. Mol Biol Evol. 26:1781–1790.

Horton P, et al. 2007. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 35:W585–W587.

Karkar S, Facchinelli F, Price DC, Weber AP, Bhattacharya D. 2015. Metabolic connectivity as a driver of host and endosymbiont integration. Proc Natl Acad Sci U S A. 112:10208–10215.

Karnataki A, et al. 2007. Cell cycle-regulated vesicular trafficking of *Toxoplasma* APT1, a protein localized to multiple apicoplast membranes. Mol Microbiol. 63:1653–1668.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30:772–780.
Keeling PJ. 2013. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. Annu Rev Plant Biol. 64:583–607.
Kilian O, Kroth PG. 2005. Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. Plant J. 41:175–183.
Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 305:567–580.
Larkum AW, Lockhart PJ, Howe CJ. 2007. Shopping for plastids. Trends Plant Sci. 12:189–195.
Lau JB, Stork S, Moog D, Sommer MS, Maier UG. 2015. N-terminal lysines are essential for protein translocation via a modified ERAD system in complex plastids. Mol Microbiol. 96:609–620.
Lechner M, et al. 2011. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics 12:124.
Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659.
Lim L, Linka M, Mullin KA, Weber AP, McFadden GI. 2010. The carbon and energy sources of the non-photosynthetic plastid in the malaria parasite. FEBS Lett. 584:549–554.
Martin W. 2010. Evolutionary origins of metabolic compartmentalization in eukaryotes. Philos Trans R Soc Lond B Biol Sci. 365:847–855.
McFadden GI. 2014. Origin and evolution of plastids and photosynthesis in eukaryotes. Cold Spring Harb Perspect Biol. 6:a016105.
Melkonian M. 2001. Systematics and evolution of the algae. I. Genomics meets phylogeny. In: Esser K, Lüttge U, Kadereit JW, Beyschlag W, editors. Progress in Botany. Berlin (Germany): Springer. p. 340–382.
Mereschkowsky C. 1905. Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. Biol Centralbl. 25:593–604.
English translation in Martin W, Kowallik KV. 1999. Annotated English translation of Mereschkowsky's 1905 paper 'Über Natur und Ursprung der Chromatophoren im Pflanzenreiche'. Eur J Phycol. 1934: 1287–1295.
Minh BQ, Nguyen MA, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 30:1188–1195.
Moog D, Stork S, Reislohner S, Grosche C, Maier UG. 2015. In vivo localization studies in the stramenopile alga *Nannochloropsis oceanica*. Protist 166:161–171.
Moog D, Stork S, Zauner S, Maier UG. 2011. In silico and in vivo investigations of proteins of a minimized eukaryotic cytoplasm. Genome Biol Evol. 3:375–382.
Mullin KA, et al. 2006. Membrane transporters in the relict plastid of malaria parasites. Proc Natl Acad Sci U S A. 103:9572–9577.
Nakai K, Horton P. 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem Sci. 24:34–36.
Neuberger G, Maurer-Stroh S, Eisenhaber B, Hartig A, Eisenhaber F. 2003. Prediction of peroxisomal targeting signal 1 containing proteins from amino acid sequence. J Mol Biol. 328:581–592.
Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32:268–274.
Petersen J, et al. 2014. *Chromera velia*, endosymbioses and the rhodoplex hypothesis–plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). Genome Biol Evol. 6:666–684.
Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 8:785–786.
Petsalaki EI, Bagos PG, Litou ZI, Hamodrakas SJ. 2006. PredSL: a tool for the N-terminal sequence-based prediction of protein subcellular localization. Genomics Proteomics Bioinformatics 4:48–55.
Price DC, et al. 2012. *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. Science 335:843–847.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.
Rost B. 1999. Twilight zone of protein sequence alignments. Protein Eng. 12:85–94.
Schlüter A, et al. 2007. PeroxisomeDB: a database for the peroxisomal proteome, functional genomics and disease. Nucleic Acids Res. 35:D815–D822.
Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504.
Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst Biol. 51:492–508.
Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol. 16:1114–1116.
Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17:1246–1247.
Small I, Peeters N, Legeai F, Lurin C. 2004. Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics 4:1581–1590.
Sommer MS, et al. 2007. Der1-mediated preprotein import into the periplastid compartment of chromalveolates? Mol Biol Evol 24:918–928.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.
Stiller JW, et al. 2014. The evolution of photosynthesis in chromist algae through serial endosymbioses. Nat Commun. 5:5764.
Stork S, Lau J, Moog D, Maier UG. 2013. Three old and one new: protein import into red algal-derived plastids surrounded by four membranes. Protoplasma 250:1013–1023.
Stork S, et al. 2012. Distribution of the SELMA translocon in secondary plastids of red algal origin and predicted uncoupling of ubiquitin-dependent translocation from degradation. Eukaryot Cell. 11:1472–1481.
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 30:2725–2729.
Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191.
Weber AP, Linka M, Bhattacharya D. 2006. Single, ancient origin of a plastid metabolite translocator family in Plantae from an endomembrane-derived ancestor. Eukaryot Cell. 5:609–612.
Weber AP, Linka N. 2011. Connecting the plastid: transporters of the plastid envelope and their role in linking plastidial with cytosolic metabolism. Annu Rev Plant Biol. 62:53–77.
Weber AP, Schwacke R, Flügge UI. 2005. Solute transporters of the plastid envelope membrane. Annu Rev Plant Biol. 56:133–164.
Zaslavskaia LA, Lippmeier JC, Kroth PG, Grossman AR, Apt KE. 2000. Transformation of the diatom *Phaeodactylum tricornutum* (Bacillariophyceae) with a variety of selectable marker and reporter genes. J Phycol. 36:379–386.
Zimorski V, Ku C, Martin WF, Gould SB. 2014. Endosymbiotic theory for organelle origins. Curr Opin Microbiol. 22:38–48.

Associate editor: Bill Martin