Nucleus-to-Nucleus Gene Transfer and Protein Retargeting into a Remnant Cytoplasm of Cryptophytes and Diatoms

Sven B. Gould,[1] Maik S. Sommer,* Peter G. Kroth,† Gillian H. Gile,‡ Patrick J. Keeling,‡ and Uwe-G. Maier**

*Philipps-Universität Marburg, Marburg, Germany; †Department of Biology, University of Konstanz, Konstanz, Germany; and ‡Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada

The complex plastid of the cryptophyte *Guillardia theta* and of the diatom *Phaeodactylum tricornutum* can both be traced back to an engulfed eukaryotic red alga. The eukaryotic origin of these plastids is most obvious in cryptophytes, where the organelle still possesses a remnant nucleus, the nucleomorph. The nucleomorph itself is embedded in the periplastid compartment (PPC), the remnant of the former red algal cytosol. In the cryptophyte and diatom, the complex plastid is surrounded by 4 membranes, the outer one being continuous with the host rough endoplasmatic reticulum. In a recent report, we have shown that a nuclear encoded PPC protein of *G. theta* expressed in *P. tricornutum* leads to a localization, recently described as being a "bloblike structure," which can be obtained by mutation of plastid protein–targeting sequences of the diatom itself.

Here we present further nucleus-encoded PPC proteins from *G. theta*, such as the eukaryotic translation elongation factor-1α, evidence for their nucleus-to-nucleus gene transfer, and retargeting of the proteins. We also investigated the first nuclear encoded PPC-targeted protein of *P. tricornutum* (Hsp70) and analyzed it for in vivo localization together with the identified *G. theta* PPC proteins. This revealed that all localize to the bloblike structures, which we suggest is the highly reduced PPC of *P. tricornutum*. Furthermore, the described cryptophyte PPC proteins possibly allow the elucidation of the processes by which proteins are involved in different levels of host control over its eukaryotic organelle.

Introduction

Cell evolution is frequently based on the reutilization and recombination of existing biological processes. This can be observed in secondary endosymbiosis, which created eukaryotic chimeras harboring complex plastids surrounded by 3 or 4 membranes. They originate from the enslavement and intracellular reduction of a eukaryotic and photoautotrophic cell by a heterotrophic host (Hjorth et al. 2005). Complex plastids are found in several algal groups, including the chromalveolates that are believed to be a monophyletic group with complex plastids of red algal origin (Cavalier-Smith 2003; Patron et al. 2004; Harper et al. 2005). An intermediate phenotype of secondary endosymbiosis can be observed in cryptophytes (Gibbs 1981). Their 4-membrane–bounded plastid still contains the remnant cytoplasm of the red alga (PPC) including a pigmy cell nucleus, the nucleomorph, which is located between the outer and the inner membrane pairs (fig. 1).

Sequencing of the nucleomorph genome of the cryptophyte *Guillardia theta* (Douglas et al. 2001) revealed a severely limited coding capacity and the absence of several genes necessary for enzymatic machineries specific for the PPC, for example, starch synthesis. Furthermore, early investigations on cryptophytes showed that the division of the nucleomorph precedes that of the cell nucleus (McKerracher and Gibbs 1982) and should be regulated by host factors. Hence, one has to postulate that several proteins need to be imported from the host cytosol to the PPC, which means crossing the outer 2 of the 4 membranes surrounding the complex plastid, which are the chloroplast endoplasmatic reticulum (cER) and periplastid membrane (PM).

Recently, we demonstrated that the uridine diphosphate–glucose–glycogen glucosyltransferase (UGGt) is encoded in the *G. theta* nucleus and synthesized as a preprotein with a bipartite topogenic signal (BTS), which enables the enzyme to cross only the 2 outer membranes (Gould et al. 2006). So far, a BTS has only been associated with nucleus-encoded plastid proteins that cross all 4 membranes of cryptophytes (Wastl and Maier 2000). The BTS of plastid and PPC proteins both begin with an N-terminal signal peptide (SP), which directs the preprotein cotranslationally into the ER lumen, most likely via a SEC61 complex. The SP is followed by a transit peptide (TP)–like sequence responsible for further translocation across the subsequent membranes of the complex plastids (Apt et al. 2002; Chaal and Green 2005). Closer analysis of the amino acid (aa) composition of the *G. theta* BTS sequences revealed a small but significant difference at the +1 position relative to the predicted cleavage site for a signal peptide peptidase (SiPP) (Gould et al. 2006). In plastid preproteins, this position was conserved for phenylalanine, whereas the UGGt protein had a serine. Substitution of the serine into a phenylalanine led to targeting of the green fluorescent protein (GFP) fusion protein into the plastid stroma when expressed in a diatom.

As the nucleomorph genome does not have the capacity to encode all proteins necessary for the maintenance of the PPC (Douglas et al. 2001), one can expect that the host not only provides missing factors but also exerts control over the endosymbiont. In order to learn more about the communication between host and endosymbiont, we screened our expressed sequence tag (EST) database for proteins with bipartite signal sequences, and putative PPC proteins were fused to GFP and expressed in *Phaeodactylum tricornutum*.

We chose the diatom *P. tricornutum* for in vivo localization of proteins for several reasons: 1) In contrast to cryptophytes, methods are available to integrate and express foreign genes (Apt et al. 1996); 2) the topogenic signals

[1] Present address: School of Botany, University of Melbourne, VIC-3010, Australia.

Key words: endosymbiosis, complex plastids, periplastid compartment, gene transfer, protein transport, chromists.

E-mail: maier@staff.uni-marburg.de.

Mol. Biol. Evol. 23(12):2413–2422. 2006
doi:10.1093/molbev/msl113
Advance Access publication September 13, 2006

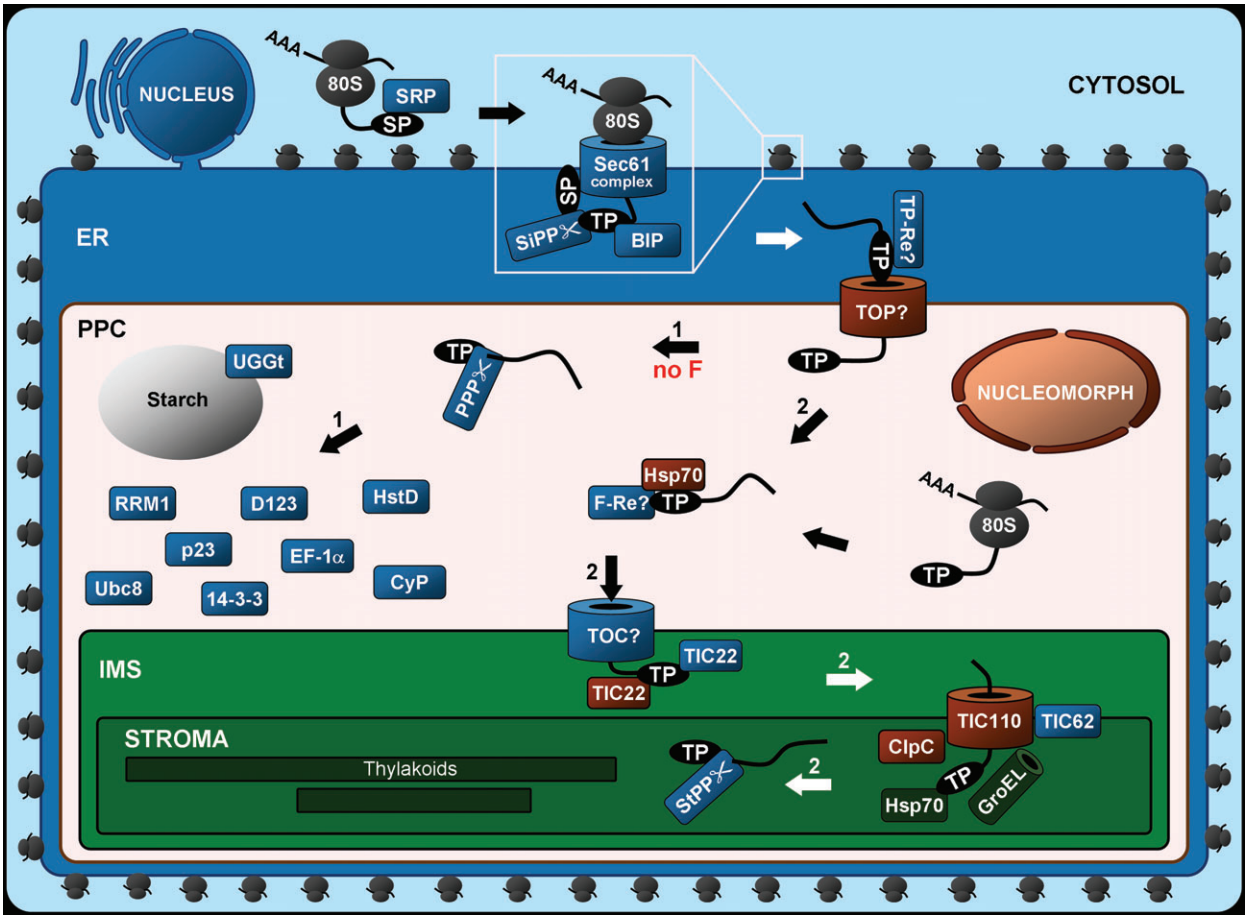

FIG. 1.—Schematic overview and postulated model of complex plastid protein targeting in cryptophytes. Nuclear encoded proteins are shown in blue, nucleomorph-encoded proteins are in brown, and plastid ones in green. All preproteins are first cotranslationally transported into the ER via the SEC61 complex after recognition of the SP by the signal recognition particle (SRP). In the ER lumen, the SP is cleaved by a SPP and the transit peptide (TP) recognized by endoplasmatic reticulum Hsp70-cognate binding protein and/or a transit peptide receptor (TP-Re). The nature of the next transport step across the second and third membrane remains uncertain, but translocators (TOP, translocator of the PM; TOC, translocator of the outer chloroplast membrane) seem most reasonable. Inside the PPC, the TP is most likely bound by Hsp70 together with an unknown factor recognizing the phenylalanine motif (F-Re, F-receptor). PPC-resident proteins do not have an F or other aromatic aa and are not recognized, and therefore, the TP is cleaved by a periplastid-processing peptidase (PPP). The nucleomorph-encoded TIC110 (TIC, translocator of the inner chloroplast membrane) might serve as a translocator for nucleus and nucleomorph-encoded plastid proteins. This transport step is assisted by a nuclear encoded TIC62 and TIC22 and a nucleomorph-encoded TIC22. Hypothetical proteins end with a question mark. For details see text.

for nucleus-encoded plastid proteins in the diatom are similar to those of *G. theta* (Kilian and Kroth 2005); and 3) the complex plastids of both are hypothesized to have a common origin and, with the exception of the nucleomorph, an identical morphology and membrane topology (Gould et al. 2006). We now show that these similarities are extended to the molecular level. We describe the first nuclear encoded protein (Hsp70, termed ptHsp70) from *P. tricornutum* that also encodes an N-terminal BTS leading to localization in the *P. tricornutum* PPC, which was recently described as a "bloblike structure" (Kilian and Kroth 2005). The same localization was observed for the predicted PPC proteins from *G. theta*. Until now this localization was only achieved by the mutation of nucleus-encoded plastid BTS sequences from *P. tricornutum* and the one BTS of the UGGt of *G. theta* (Gould et al. 2006). We also show that the *G. theta* host nucleus encodes a gene for translation elongation factor-1α (EF-1α) that is phylogenetically related to red algal homologues and is inferred to be derived from the nucleomorph (the host cytoplasm does not use EF-1α but uses the analogous EF-like protein (EFL) instead: Gile et al. 2006). This gene encodes a BTS, which directs PPC targeting in *P. tricornutum* and so overall represents the fate of many nucleomorph genes, which were transferred to the nucleus and their products are targeted to their original compartment using the BTS.

Material and Methods

EST Database and In silico Analyses

A normalized cDNA library enriched for full-length sequences from *G. theta* strain CCMP327 was generated by the Vertis Biotechnologie AG (Freising, Germany) and provided in Lambda ExCell phages. In vivo phagemid release was performed by the use of a standard protocol using *Escherichia coli* NP66 cells. In all, 9,000 single clones were 5′ end sequenced by MWG-Biotech (Ebersberg, Germany). Sequences were clustered and analyzed using the Sequencher software v4.5 from Genecodes (Ann Arbor, MI). Putative full-length clones were analyzed by software

analysis for topogenic signals by SignalP (http:www.cbs.dtu.dk/services/SignalP/) and TargetP (http:www.cbs.dtu.dk/services/TargetP/#submission).

In vivo Localization of GFP Fusion Proteins and Western Blot Analysis

All BTS-encoding regions were amplified by standard reverse transcription–polymerase chain reactions using specific oligonucleotides introducing 5′ *Eco*RI and 3′ *Nco*I restriction sites, using cDNA from *G. theta* as a template. Products were digested with *Eco*RI and *Nco*I, the plasmid pPhaT1 (Zaslavskaia et al. 2000) with *Eco*RI and *Hind*III, and the GFP-encoding fragment with *Nco*I and *Hind*III. All 3 fragments were ligated in one single step and the products transformed into *E. coli* XL1-Blue MRF′. Plasmids were isolated using MIDI plasmid purification Kit from Qiagen (Hilden, Germany) and inserts sequenced on a LI-COR LR4200 (LI-COR Biosciences, Bad Homburg, Germany) using the "Thermosequenase fluorescent labelled primer cycle sequencing Kit with 7-deaza-dGTP" from GE Healthcare (Munich, Germany). *Phaeodactylum tricornutum* CCMP632 was cultured and transformed according to Apt et al. (1996). Transgenic diatoms were analyzed with a confocal laser scanning microscope Leica TCS SP2 at room temperature in f/2 culture medium, using HCX PL APO 40×/1.25–0.75 oil CS or PL APO 63×/1.32–0.60 oil Ph3 CS objectives. GFP and chlorophyll fluorescence was excited at 488 nm, filtered with beam splitter TD 488/543/633, and detected by 2 different photomultiplier tubes with a bandwidth of 500–520 and 625–720 nm for GFP and chlorophyll fluorescence, respectively. Hoechst 33342 and MitoTracker Orange CMTMRos (both Molecular Probes) were used for DNA and mitochondrial staining, respectively. Images were processed with LCS-Lite from Leica (Wetzlar, Germany).

For western blot analysis of cells expressing GFP fusion protein constructs, protein extracts of *P. tricornutum* were separated on a 15% sodium dodecyl sulphate–polyacrylamide gel electrophoresis and blotted onto Hybond P membrane GE Healthcare, Buckinghamshire, United Kingdom at 0.8 mA/cm^2. Blots were analyzed using goat anti-GFP as a primary and rabbit anti-goat/horseradish peroxidase–coupled as a secondary antibody. Both antibodies were a generous gift from the AG Batschauer (Marburg, Germany).

Phylogenetic Analysis

New sequences were added to an alignment with a broad range of eukaryotic diversity (Keeling and Inagaki 2004) and analyzed by maximum likelihood (ML), distance, and Bayesian phylogenetic methods. ML and distance analyses assumed the Whelan–Goldman model of aa substitution (Whelan and Goldman 2001) with site-to-site rate variation modeled on a gamma distribution with 4 variable rate categories and an invariable category. The shape parameter alpha and the proportion of invariable sites were estimated from the data using PhyML 2.4.4 (Guindon and Gascuel 2003). ML trees and 1,000 ML bootstrap replicates were inferred using PhyML as above. Distance trees were inferred by weighted Neighbor-Joining using Weighbor 1.2 (Bruno et al. 2000) using distances calculated by Tree-Puzzle 5.2 (Schmidt et al. 2002). Bayesian analysis was carried out using MrBayes 3.0B4 (Ronquist and Huelsenbeck 2003) with the Jones–Taylor–Thornton model of aa substitution (Jones et al. 1992) and 4 equally probable gamma-distributed substitution rate categories and a proportion of invariable sites estimated from the data. One cold and 3 heated Markov chains were run for 1,000,000 generations, with a tree sampled every 100 generations. Log likelihoods stabilized after 2,000 generations, and the remaining trees were used to compute the 50% majority-rule consensus tree. Approximately unbiased (AU) tests (Shimodaira 2002) were performed using Consel 1.19 (Shimodaira and Hasegawa 2001) to test alternate positions in the tree for the *G. theta* PPC-targeted gene. An ML tree was inferred with the *G. theta* sequence excluded; the same clades were recovered, but, as with the other analyses, the branching order was unresolved. The *G. theta* PPC-targeted sequence was grafted to 33 alternate positions basal to and within resolved clades found in the ML topology. Site likelihoods for trees were computed using the –wsl option of Tree-Puzzle 5.1 using parameters from the original tree.

Results

Identification of Genes for PPC-Targeted Proteins in *G. theta*

From our current EST project on *G. theta*, we classified the entries to be located in different cellular compartments by standard Blast searches, as well as by bioinformatic prediction of putative topogenic signals. From out of 9,000 single read sequences, we got 4,288 independent clusters. We identified 92 clusters predicted to encode mitochondrial proteins due to a mitochondrial targeting peptide and sequence similarity to mitochondrial proteins of other organisms. In all, 220 were classified as proteins entering the secretory system due to a predicted N-terminal SP or signal anchor. Out of these, 139 also encoded a downstream sequence with characteristics of a transit peptide including the typical aromatic phenylalanine at the +1 position. In the case of the plastid phosphoglycerate kinase, we identified an aromatic tyrosine, which was sufficient for in vivo stromal targeting (not shown). This signature, phenylalanine, was recently described to be one indicator for nuclear encoded proteins that are transported across the 4 membranes of complex plastids in several chromalveolate groups and red algae (Kilian and Kroth 2005; Patron et al. 2005; Gould et al. 2006), as well as the envelope of plastids from *Cyanophora paradoxa* (Steiner et al. 2005). As most of these proteins furthermore show homology to known plastid proteins, we tentatively assigned them as nuclear encoded plastid proteins.

Assigning the location to cytosolic proteins is more complex in *G. theta* because 2 eukaryotic cytosolic compartments are present next to each other: the cytosol of the host cell and the reduced cytosol of the secondary endosymbiont, separated from each other by 2 membranes, the cER and PM (fig. 1). Both host and endosymbiont cytosolic compartments are supposed to contain housekeeping factors like DNA polymerases, cell cycle components, or translation factors, with one exception: in cryptophytes, starch is synthesized and deposited within the PPC only (Deschamps et al. 2006). Accordingly, we used a combination of

Table 1
List of Identified Proteins and Their In vivo Localization

Protein	Topogenic Signal[a]	N-Terminal Sequence[b]	Localization[c]
EFL	None	MAEKEHLSIVICGHVDSGKSTTTGRLLFEL	C
IscA	TP	MPRLLLLRRLPRCSASSLTSIPSIPLFSST	M
CatA	SP	*MAPWMAVVAMVAIGMQVALA***AV**PTHSRRRAGGGLAEPLAPGATDEITDLPG	ER
LHC	SP + TP (BTS)	*MIRACALLGLAASAA***AF**APSSLPIRANRASAVSKMSMQSNRFSYRS	P
EF-1	SP + TP (BTS)	*MAAAGQINLSVMKCVAVLFLMGLATTMGA***AG**RTFSAMGHRMNMIPSGTSPSMLRLRGGGK	PPC
UGGt	SP + TP (BTS)	*MRRSVLSAAAVLSLSLSLLSLTPSN***AS**TGVSSFSSSHRIPSLTRSALRLSKEGKGE	PPC
14-3-3	SP + TP (BTS)	*MLLRSGTTIILLMVALEILTVAS***SS**LSSPFPTKLLQSRPSRSTSLAAKVFALR	PPC
RRM1	SP + TP (BTS)	*MNTISWSSLVLLVVASLGSLEEVN***SD**RACRSISSPLRLRGGFRGDDNQYDQGYGR	PPC
CyP	SP + TP (BTS)	*MPRLILLLVVLGICSA***AN**YASRYAPMAGPRKFLGDGELPSVLNSREL	PPC
Ubc8	SP + TP (BTS)	*MKAASMRLKNLLLLLSLPLVLSSS***PS**PTDLVRLSSSRLRGALRRAVQEEETCIQR	PPC
Ubc14	SP + TP (BTS)	*MSLWSVDHCQLLLACFLSLALLPGLLA***AM**APAANSPTRLRSSQILRMQLRELQENPSE	Nd
D123	SP + TP (BTS)	*MHARLRPGTILLICTALVEA***AN**QLSWTKDIERVLPAMTSTVDIKKADPLAP	Nd
P23	SP + TP (BTS)	*MAGRKDGTRGGAMAGRLVFCMCVTFLCGTIV***SN**PMQGGEVARLNLRAYRPRAGHSLEGLRLR	Nd
HstD	SP + TP (BTS)	*MLALVPTMAATMAATIE***AA**TCSSTLFFLPCSPFSSLPYFHLSHHSLPP	Nd
PtHsp70	SP + TP (BTS)	*MVHLPSSSTLLTCVSVLLSGAHPAK***AS**WLARRTVEKPTLARIHEQRDSTDRKSRAP	PPC

[a] Topogenic signal predicted: TP, TP-like

[b] N-terminal sequences encoding the predicted topogenic signals: SP in italics, SP cleavage site in bold, and TP underlined.

[c] In vivo localization: C, cytosol; M, mitochondrion; ER, endoplasmatic reticulum; P, plastid stroma; PPC, periplastid compartment; and Nd, not determined.

topogenic signals and phylogeny to distinguish host and endosymbiont cytosolic proteins as well.

As already described, the UGGt has a BTS with attributes for targeting nucleus-encoded proteins across 2 of 4 membranes into the PPC of *G. theta* (Gould et al. 2006). The BTS of the UGGt was used as a "blueprint" to identify further nucleus-encoded proteins targeted to the PPC. In particular, the 31 sequences predicted to have both an SP and a TP, but which lacked an aromatic aa at the +1 position, were predicted to be putative PPC proteins. Nine of these putative PPC proteins with homology to soluble eukaryotic proteins were further analyzed (table 1): an RNA-binding protein (RNA recognition motif 1, RRM1), 2 regulating protein synthesis (D123; EF-1α), 3 for protein modification (cyclophilin, CyP; p23; histone deacetylase, HstD), 2 involved in protein degradation (ubiquitin-conjugating enzymes, Ubc8 and Ubc14), and a 14-3-3 protein, which is known to bind a phosphoserine motif and thereby regulate a variety of cellular processes.

Topogenic Signals in *G. theta* and Their Targeting Properties

To analyze the targeting properties of cryptophyte sequences predicted to be located in various compartments, presequences of proteins predicted to be targeted to various compartments within the cryptophyte cell (table 1) were expressed in the diatom *P. tricornutum* as GFP fusion proteins. As a representative cytosolic protein, the first 58 aa of EFL, which encodes no topogenic signal, were fused to GFP. This construct, as expected, led to localization of the GFP fluorescence in the cytosol and nucleus (fig. 5). The same is observed when GFP is expressed in *P. tricornutum* without any topogenic signal attached (fig. 1). As a representative ER-targeted protein, the 20 aa–predicted SP of cathepsin A was fused with GFP, resulting in fluorescence exclusively in the ER and nucleus envelope, which is consistent with previous observations with a diatom SP that was fused to GFP (Kilian and Kroth 2005). Fusion with the predicted mitochondrial targeting sequence of the iron–sulfur cluster biosynthesis enzyme IscA (Lill and Mühlenhoff 2005) lead to mitochondrial localization, as has recently been described for mitochondrial proteins of *P. tricornutum* (Apt et al. 2002) and further supported by parallel staining with MitoTracker (fig. 2). Lastly, the BTS from the well-characterized, plastid-targeted light-harvesting complex protein was fused with GFP, resulting in localization of the fusion protein to the plastid stroma, as expected (fig. 2). Thus, the representative topogenic signals from the cryptophyte directed subcellular localization in the diatom *P. tricornutum* as predicted.

The putative PPC proteins we have identified are generally predicted to be derived from the red algal endosymbiont and to have been transferred to the host nucleus during the reduction of the nucleomorph, leading to the prediction that these proteins are phylogenetically allied to red algae. One of the genes identified encodes translation EF-1α, the functional analogue of EFL, and a gene that is well sampled and has been used extensively for phylogenetic analysis (Hashimoto et al. 1995; Baldauf and Doolittle 1997; Roger et al. 1999; Gaucher et al. 2001). We conducted a phylogenetic analysis of this protein, which consistently showed the *G. theta* protein to be related to red algal homologues with very strong support (fig. 4), as predicted for an endosymbiont-derived gene. To determine whether the localization of the protein was consistent with its phylogenetic origin, we fused the BTS sequence from EF-1α to GFP and expressed the fusion construct in *P. tricornutum*, resulting in GFP fluorescence restricted to a single spot–like pattern, distinct from DNA staining (fig. 5). Because the EFL sequence and localization (fig. 5) is consistent with function in the host cytoplasm, it appears cryptophytes use a cytosolic EFL and PPC EF-1α in the same cell, and the EF-1α gene was transferred form the nucleomorph genome to that of the host nucleus and is now targeted back by a BTS.

We also fused the BTS-encoding regions from the 8 other putative PPC-targeted proteins (table 1) to GFP and expressed them in *P. tricornutum*. From 4 constructs, stable transformants expressing GFP were acquired: Ubc8, CyP, 14-3-3, and RRM1. All 4 showed the same characteristic single

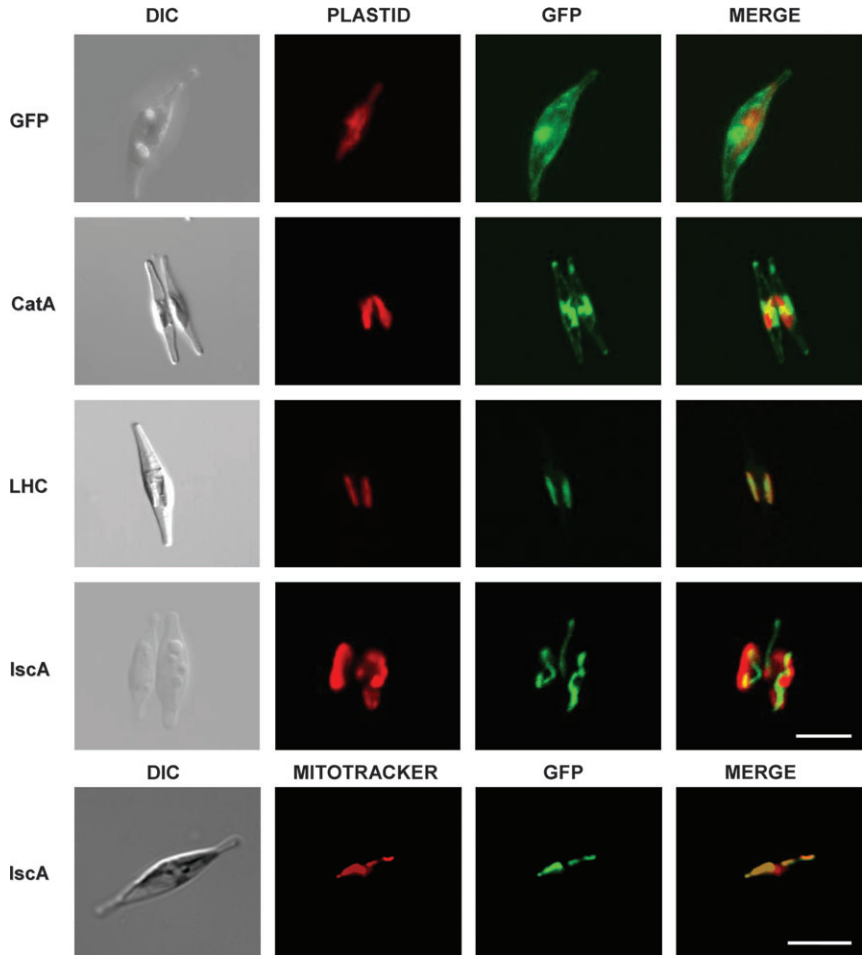

FIG. 2.—Targeting properties of *Guillardia theta* nucleus-encoded topogenic signals in *Phaeodactylum tricornutum*. Columns from left to right: light microscope images, red chlorophyll autofluorescence, GFP fluorescence, and merged green and red fluorescence. Cytosolic expressed GFP without a topogenic signal accumulates in the cytosol and nucleus. The SP of cathepsin A (CatA) leads to ER and nucleus envelope localization of GFP, and the plastid-targeting BTS of a light-harvesting protein (LHC) results in colocalization of GFP and the chlorophyll fluorescence. The mitochondrial targeting peptide of the iron–sulfur cluster assembly protein (IscA) targets GFP into mitochondria (also shown by the colocalization with MitoTracker). Scale bar represents 10 μm.

spot–like localization of the GFP signal that was observed with the EF-1α construct (figs. 5 and 6). In all these transformants, there was no additional GFP fluorescence in the cytosol or colocalizing with the autofluorescense of the plastid.

An Endogenous PPC-Targeted Protein of *P. tricornutum*

In addition to the putative PPC-targeted protein genes in *G. theta*, we detected in an EST database of *P. tricornutum* a nuclear encoded Hsp70 (PtHsp70; JGI Prot. ID28904), which, when compared with the mature Hsp70 domain (Pfam00012), included an N-terminal extension of around 90 aa. These 90 aa were predicted to encode an SP of 26 aa with the predicted SPP cleavage site between alanine and serine. The SP was followed by a predicted TP with a serine at the +1 position (table 1). This makes the PtHsp70 the first protein of *P. tricornutum* with a putative BTS lacking phenylalanine at the +1 position of the SPP cleavage site.

To test whether this protein is targeted to the PPC in vivo, we constructed a GFP fusion protein encoding the first 86 aa of PtHsp70 at the N terminus (fig. 3*A*) and transfected *P. tricornutum*. In 4 independent clones, the fusion protein showed the same single spot–like distribution as *P. tricornutum* cells expressing PPC–GFP fusion proteins of *G. theta*: somewhere within the plastid envelope but not colocalized with the plastid autofluorescence (fig. 3*B*). Western blot analysis of cells expressing this construct shows that the fusion protein is processed to 27 kDa, approximately the size of GFP (fig. 3*C*).

Discussion

Chromalveolates are believed to have a monophyletic origin, which emerged from a heterotrophic eukaryote that engulfed and integrated a red alga (Cavalier-Smith 1999, 2003; Fast et al. 2001; Patron et al. 2004; Harper et al. 2005). The complex plastid architecture in cryptophytes and diatoms, located within the host ER and surrounded by 4 membranes, may reflect this ancestry. This morphology provides several problems still unresolved, such as protein transport into the organelle across the multiple membranes and communication between host and symbiont.

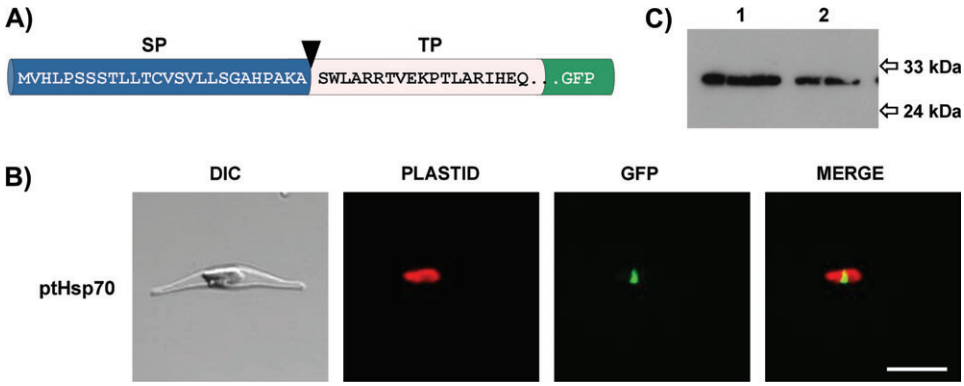

FIG. 3.—PPC localization of a nuclear encoded Hsp70 of *Phaeodactylum tricornutum*. (*A*) The identified Hsp70 of *P. tricornutum* (ptHsp70) has an N-terminal extension of around 90 aa, which is predicted to be a BTS, comprising an SP and a TP. The arrowhead indicates the predicted SP cleavage site. (*B*) Fused to GFP, this BTS targets the fusion protein to the PPC of *P. tricornutum*. Finally, in (*C*) western blot analysis with protein extracts of cells expressing the fusion construct (lane 1) and as a control GFP without a topogenic signal (lane 2) revealed that the BTS of the ptHsp70 is processed, resulting in a size corresponding to the predicted mass of GFP of around 27 kDa. Scale bar represents 10 μm.

The Highly Reduced PPC of *P. tricornutum*

It is apparent that the space between the second and third outermost membranes of diatoms is far more reduced than in cryptophytes. Nevertheless, it is homologous to the cytoplasm of the cryptophyte endosymbiont, and it may be postulated that several proteins needed for maintaining the plastid and managing transport processes are located within the space between the second and third membrane in diatoms as well as in cryptophytes.

As shown in figure 2, the predicted targeting properties of *G. theta* presequences from proteins for various compartments were all consistent with their observed localization in the diatom. Hence, not only are the plastid topogenic signals exchangeable between cryptophytes and diatoms, but also for other compartments. There is no obvious reason why this should not also hold true for the PPC. Indeed, we describe a diatom Hsp70 that is localized in this compartment (fig. 3*B*) and includes a similar BTS as those of cryptophytes.

It is implausible that the observed localization of PPC proteins in a structure, which was recently termed "bloblike structure" (Kilian and Kroth 2005), is an artifact due to GFP overexpression for several reasons: 1) The fluorescence was always observed at the same single spot–like position with all 6 independent fusion protein constructs; 2) it was never observed with any topogenic signal other than a BTS, and finally iii) the *P. tricornutum* PtHsp70, encoding an N-terminal BTS of around 90 aa, showed the same localization. It is most reasonable that the recently described bloblike structure in *P. tricornutum* (Kilian and Kroth 2005) is located between the inner and outer membrane pair of the complex plastid and is homologous to the cryptophyte PPC. The ptHsp70 transit peptide seems to be cleaved inside this compartment (fig. 3), which is also true for the transit peptide of the UGGt in the PPC of *G. theta* (Deschamps et al. 2006).

Implications for a Transport Model

Our findings significantly increase the number of identified proteins that are transported across 2 or 4 membranes of the complex plastid in *G. theta*. There seems to be no functional difference between SPs used for import into the endosymbiont and those of ER-resident proteins. Experiments focusing on the first transport step of nucleus-encoded plastid proteins, whether in cryptophytes, diatoms, or apicomplexans agree that the preproteins enter the ER cotranslationally through the Sec61 complex, where the SP is then cleaved (Wastl and Maier 2000; Chaal and Green 2005; Kilian and Kroth 2005). Thus, discrimination between secretory proteins and those intended to enter the complex plastid occurs inside the ER, most likely by recognition of the transit peptide, which is absent in secreted proteins. In most plastid-targeting experiments in chromalveolates, including ours, only the BTS was fused to GFP. These constructs led to correct targeting of the fusion proteins into the complex plastids. This allows the conclusion that no further topogenic signal other than the transit peptide is required for correct targeting, unless of course the GFP by coincidence encodes a not yet recognized topogenic signal also present in all nuclear encoded complex plastid proteins.

Protein import of nucleus-encoded proteins into secondary plastids of red algal origin is only poorly understood, especially how these proteins cross the second and third membranes, and only a few factors involved have been identified (McFadden and van Dooren 2004). It has been a matter of debate whether vesicles or subsequent translocators are responsible for preprotein translocation across these membranes (Gibbs 1979; van Dooren et al. 2001; Cavalier-Smith 2003; Bodyl 2004; Kilian and Kroth 2005). As transit peptides are found on plastid and PPC proteins of cryptophytes and on the ptHsp70, and all these leaders have similar characteristics, we suggest a model (fig. 6) where 1 common pathway is responsible for import of all preproteins into the PPC. In particular, for PPC preproteins, a vesicle transport seems unlikely as these preproteins would have to be released somehow from the transport vesicles before they fuse with the third membrane. Furthermore, the vesicle model cannot explain how nucleomorph-encoded plastid proteins are transported across the envelope into isolated plastids of *G. theta* (Wastl and Maier 2000).

Plastid proteins, whether nucleus- or nucleomorph-encoded, are most likely recognized by an unknown receptor inside the PPC, and the highly conserved N-terminal

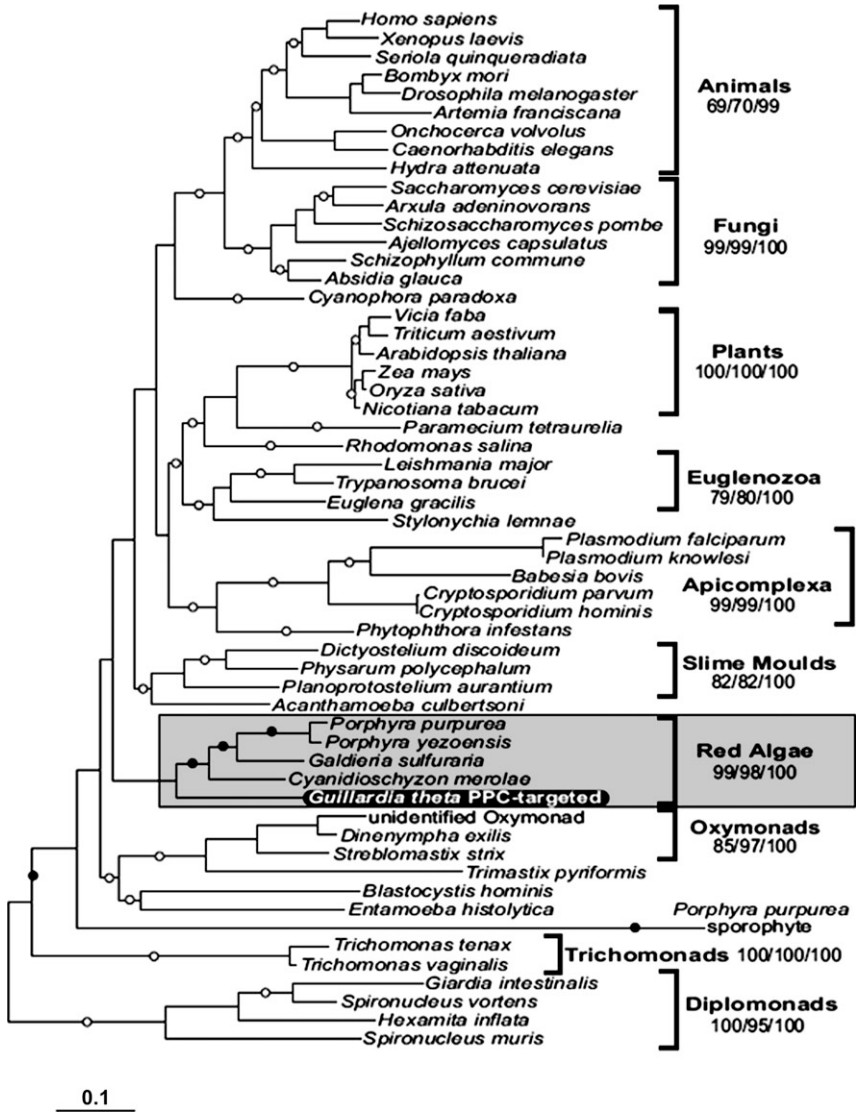

FIG. 4.—ML phylogeny of EF-1α. Bracketed clades were recovered in all phylogenetic analyses, but the branching order of these and the non-bracketed taxa was unresolved. Numbers indicate support for that clade in the order, ML bootstrap, weighted Neighbor-Joining bootstrap, and Bayesian posterior probability. Open circles indicate positions where the *Guillardia theta* PPC-targeted sequence was rejected by AU tests, whereas closed circles indicate positions that could not be rejected.

phenylalanine undoubtedly plays a role in this recognition. In very few cases, other aromatic acids were detected (tyrosine and tryptophan) instead of the phenylalanine, and the phenylalanine is not as highly conserved in other systems (e.g., glaucophyte cyanelles or plastids of dinoflagellates and haptophytes: Patron et al. 2005, 2006; Steiner et al. 2005). We believe an unknown factor (maybe the receptor itself) recognizes the aromatic aa, and therefore, Hsp70 can bind the transit peptides keeping the plastid proteins' transport competent. If PPC proteins systematically lack this phenylalanine, they would not be recognized, leading to the cleavage of the transit peptide inside the PPC. That TP cleavage occurs inside the PPC of the cryptophyte and the diatom was shown by our sequencing of the mature N terminus of the UGGt (Deschamps et al. 2006) and the western blot analysis of the ptHsp70 (fig. 2). Altogether, these results strongly support a receptor-mediated import through a protein-based translocon, as also suggested by others (Cavalier-Smith 1999, 2003; van Dooren et al. 2001; Bodyl 2004).

Host Control

Lateral gene transfer from a symbiont genome to the nucleus of the host is an ongoing process and well examined especially in land plants (Leister 2005). Such transfers have almost exclusively been observed for sequences moving from a prokaryotic organelle to the eukaryotic nucleus of the host. There are reports of nuclear encoded genes, potentially originating from the red algal nucleomorph (Deane et al. 2000; Stibitz et al. 2000), and nucleus-to-nucleus transfers of organellar genes are also known (Kroth et al. 2005; Patron et al. 2006), but none of these proteins are targeted back to their source compartment.

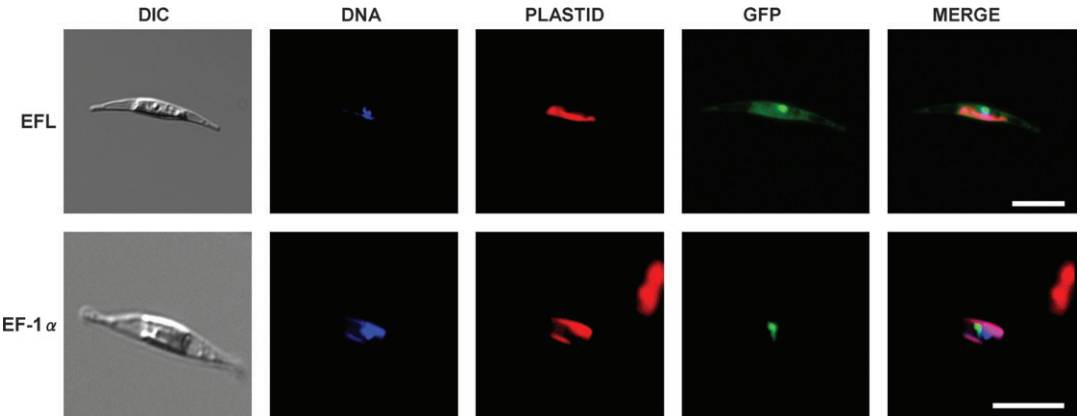

FIG. 5.—Localization of EFL and EF-1α GFP constructs in *Phaeodactylum tricornutum*. When fusing the N-terminal region of the EFL gene of *Guillardia theta* to GFP, the fluorescence is observed inside the cytosol and nucleus, just like when GFP is expressed alone (fig. 2). In contrast, the BTS of the EF-1α results in a single spot–like localization, clearly separated from the nucleus, as shown by Hoechst 33342 staining of the DNA. Scale bar represents 10 μm.

Phylogenetic analyses on the EFL protein revealed an unusual punctate distribution amongst dinoflagellates, haptophytes, cercozoa, green algae, choanoflagellates, fungi (Keeling and Inagaki 2004), and now in *G. theta* and other cryptomonads (Gile et al. 2006; this study). It was shown that within these genomes only one gene or the other is typically present, either the canonical EF-1α or the intruded EFL gene, and it was concluded that in the course of time EFL has replaced EF-1α in several lineages. In *G. theta*, however, we find for the first time unambiguous evidence that both proteins are encoded in the nucleus of one organism, but even here the proteins do not operate in the same compartment. Phylogenetic analysis of the protein sequences revealed that the EF-1α is of red algal origin and was, therefore, most

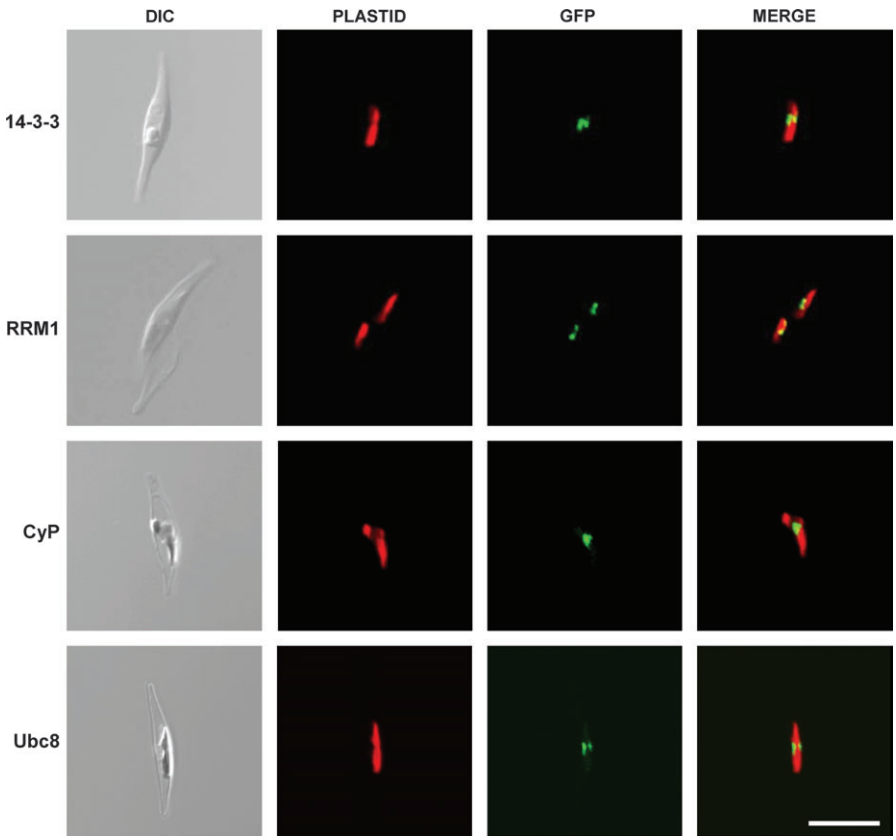

FIG. 6.—Localization of nuclear encoded PPC proteins of *Guillardia theta*. GFP fusion proteins of various PPC proteins resulted in GFP localization within the highly reduced PPC of *Phaeodactylum tricornutum*: Ubc8, ubiquitin-conjugating enzyme; CyP, Cyclophilin; 14-3-3 and RRM1, RNA recognition motif 1. The merge of plastid autofluorescence and the GFP shows the tight association but not colocalization of both fluorescences. The same was observed with the PtHsp70 of *P. tricornutum* (fig. 3). Scale bar represents 10 μm.

certainly transferred from the nucleomorph to the nucleus. Once transferred, the gene was equipped inter alia with a BTS-encoding sequence, which allows the protein to be targeted back to the PPC of the cryptophyte. Hence, in *G. theta* there is not one copy per genome, but rather one copy per eukaryotic cytosol bearing 80S ribosomes.

As is the case with prokaryotic endosymbionts, the cryptophyte host cell apparently effectively controls important cellular functions of the eukaryotic endosymbiont, including the maturation of proteins by CyP, an otherwise cytoplasmic cyclophilin and the posttranslational regulation by a 14-3-3 protein. Furthermore, our findings, which for the first time inspect eukaryote-over-eukaryote control, indicate that the host can affect protein turnover to a certain degree by encoding PPC-targeted Ubc8 and Ubc14, enzymes involved in the specific ubiquitination of proteins. Translation elongation within the endosymbiont is regulated by the earlier described EF-1α. Additionally, we detected that the host can participate in the nucleotide metabolism of the symbiont, as seen by the PPC localization of a GFP fusion protein expressing the topogenic signal of a protein with an RRM1–RNA-binding domain.

Concluding Remarks

Our study offers important insight into the coevolution of 2 eukaryotic cells. For the first time, a nucleus-to-nucleus gene transfer of a nonplastid protein and its retargeting is documented. Of this protein, 2 phylogenetically independent copies of a eukaryotic housekeeping factor coexist next to each other in one genetic compartment. In addition, we show the first example of a heterokontophyte protein inserted into the PPC-like space between the second and third outermost membrane in a diatom, which is significant given the vast reduction of complexity in this space. A model for targeting signals has also allowed us to identify a number of cryptophyte proteins potentially involved in different levels of host control over its eukaryotic organelle.

Supplementary Material

New and relevant EST data were deposited under AM183803–AM183820.

Acknowledgments

The authors thank A. Gruber for help in the assembly of the ptHsp70 gene of *P. tricornutum*. S.B.G., M.S.S., and U.G.M. are supported by the Deutsche Forschungsgemeinschaft (DFG) within the SFB 593. P.G.K. is supported by the DFG (Project KR1661/3-3). P.J.K. and G.H.G. are supported by an operating grant (227301) from the Natural Sciences and Engineering Research Council of Canada. P.J.K. is a Fellow of the Canadian Institute for Advanced Research and a New Investigator of the Canadian Institutes for Health Research and Michael Smith Foundation for Health Research.

Literature Cited

Apt KE, Kroth-Pancic PG, Grossman AR. 1996. Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. Mol Gen Genet. 25:572–579.

Apt KE, Zaslavkaia L, Lippmeier JC, Lang M, Kilian O, Wetherbee R, Grossman AR, Kroth PG. 2002. *In vivo* characterization of diatom multipartite plastid targeting signals. J Cell Sci. 115:4061–4069.

Baldauf SL, Doolittle WF. 1997. Origin and evolution of the slime molds (Mycetozoa). Proc Natl Acad Sci USA. 94:12007–12012.

Bodyl A. 2004. Evolutionary origin of a pre-protein translocase in the periplastid membrane of complex plastids: a hypothesis. Plant Biol. 6:513–518.

Bruno WJ, Socci ND, Halpern AL. 2000. Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. Mol Biol Evol. 17:189–197.

Cavalier-Smith T. 2003. Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae). Philos Trans R Soc Lond B Biol Sci. 358:109–133.

Chaal BK, Green BR. 2005. Protein import pathways in 'complex' chloroplasts derived from secondary endosymbiosis involving a red algal ancestor. Plant Mol Biol. 57:333–342.

Deane JA, Fraunholz M, Su V, Maier U-G, Martin W, Durnford DG, McFadden GI. 2000. Evidence for nucleomorph to host nucleus gene transfer: light-harvesting complex proteins from cryptomonads and chlorarachniophytes. Protist. 151:239–252.

Deschamps P, Haferkamp I, Dauvillee D, et al. (12 co-authors). 2006. The nature of the periplastid pathway of amylose synthesis in the cryptophyte *Guillardia theta*. Eukaryotic Cell. 5:954–963.

Douglas E, 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.

Fast NM, Kissinger JC, Roos DS, Keeling PJ. 2001. Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. Mol Biol Evol. 18:418–426.

Gaucher EA, Miyamoto MM, Benner SA. 2001. Function-structure analysis of proteins using covarion-based evolutionary approaches: elongation factors. Proc Natl Acad Sci USA. 98:548–552.

Gibbs SP. 1979. The route of entry of cytoplasmically synthesized proteins into chloroplasts of algae possessing chloroplast ER. J Cell Sci. 35:253–266.

Gibbs SP. 1981. The chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae. Ann NY Acad Sci. 361:193–208.

Gile GH, Patron NJ, Keeling PJ. Forthcoming 2006. EFL GTPase in cryptomonads and the distribution of EFL and EF-1α in chromalveolates. Protist.

Gould SB, Sommer MS, Hadfi K, Zauner S, Maier UG. 2006. Protein targeting into the complex plastid of cryptophytes. J Mol Evol. 62:674–681.

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

Harper JT, Waanders E, Keeling PJ. 2005. On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. Int J Syst Evol Microbiol. 55:487–496.

Hashimoto T, Nakamura Y, Kamaishi T, Nakamura F, Adachi J, Okamoto K, Hasegawa M. 1995. Phylogenetic place of mitochondrion- lacking protozoan, Giardia lamblia, inferred from amino acid sequences of elongation factor 2. Mol Biol Evol. 12:782–793.

Hjorth E, Hadfi K, Gould SB, Kawach O, Sommer MS, Zauner S, Maier UG. 2005. Zero, one, two, three, and perhaps four. Endosymbiosis and the gain and loss of plastids. Endocytobiosis Cell Res. 15:459–468.

Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 8:275–282.

Keeling PJ, Inagaki Y. 2004. A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1α. Proc Natl Acad Sci USA. 101:15380–15385.

Kilian O, Kroth PG. 2005. Identification and characterization of a new conserved motif within the pre-sequence of proteins targeted into complex diatom plastids. Plant J. 4:175–183.

Kroth PG, Schroers Y, Kilian O. 2005. The peculiar distribution of class I and class II aldolases in diatoms and in red algae. Curr Genet. 48:389–400.

Leister D. 2005. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. Trends Genet. 21:655–663.

Lill R, Mühlenhoff U. 2005. Iron-sulfur-protein biogenesis in eukaryotes. Trends Biochem Sci. 30:133–141.

McFadden GI, van Dooren GG. 2004. Evolution: red algal genome affirms a common origin of all plastids. Curr Biol. 14:514–516.

McKerracher L, Gibbs S. 1982. Cell and nucleomorph division in the alga *Cryptomonas*. Can J Bot. 60:2440–2452.

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

Patron NJ, Waller RF, Archibald JM, Keeling PJ. 2005. Complex protein targeting to dinoflagellate plastids. J Mol Biol. 348: 1015–1024.

Patron NJ, Waller RF, Keeling PJ. 2006. A tertiary plastid uses genes from two endosymbionts. J Mol Biol. 357: 1373–1382.

Roger AJ, Sandbloom O, Doolittle WF, Philippe H. 1999. An evaluation of elongation factor 1 alpha as a phylogenetic marker for eukaryotes. Mol Biol Evol. 16:218–233.

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

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. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics. 17:1246–1247.

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

Stibitz TB, Keeling PJ, Bhattacharya D. 2000. Symbiotic origin of a novel actin gene in the cryptophyte *Pyrenomonas helgolandii*. Mol Biol Evol. 17:1731–1738.

Van Dooren G, Schwartzbach SD, Osafune D, McFadden GI. 2001. Translocation of proteins across the multiple membranes of complex plastids. Biochim Biophys Acta. 1541:34–53.

Wastl J, Maier UG. 2000. Transport of proteins into cryptomonads complex plastids. J Biol Chem. 275:23194–23198.

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.

Zaslavskaia LA, Lippmeier JC, Kroth PG, Grossman AR, Apt KE. 2000. Additional selectable marker and reporter genes for use in transforming the diatom *Phaeodactylum tricornutum*. J Phycol. 36:379–386.

Geoffrey McFadden, Associate Editor

Accepted September 7, 2006