Research Article:

Chimeric Plastid Proteome in the Florida “Red Tide” Dinoflagellate

Karenia brevis

Tetyana Nosenko, Kristy L. Lidie¹, Frances M. Van Dolah¹, Erika Lindquist², Jan-Fang Cheng³, US Department of Energy - Joint Genome Institute⁴, and Debashish Bhattacharya*

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

¹Marine Biotoxins Program, NOAA National Ocean Service, Center for Coastal Environmental Health and Biomolecular Research, Charleston, South Carolina

²DOE Joint Genome Institute and Lawrence Berkeley National Laboratory, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

³Genomics Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, MS 84-171, Berkeley, CA 94720, USA

⁴Corporate author, see Supplementary Material

*Corresponding author: University of Iowa, Department of Biological Sciences and the Roy J. Carver Center for Comparative Genomics, 446 Biology Building, Iowa City, Iowa 52242-1324, USA, Telephone: (319) 335-1977, Fax: (319) 335-1069, Email: debashi-bhattacharya@uiowa.edu

Key words: endosymbiosis, endosymbiotic gene transfer, Karenia brevis, proteome, red tide

Running head: Karenia plastid proteome
Abstract

Current understanding of the plastid proteome comes almost exclusively from studies of plants and red algae. The proteome in these taxa has a relatively simple origin via integration of proteins from a single cyanobacterial primary endosymbiont and the host. However, the most successful algae in marine environments are the chlorophyll c-containing chromalveolates such as diatoms and dinoflagellates that contain a plastid of red algal origin derived via secondary or tertiary endosymbiosis. Virtually nothing is known about the plastid proteome in these taxa. We analyzed expressed sequence tag data from the toxic “Florida red tide” dinoflagellate *Karenia brevis* that has undergone a tertiary plastid endosymbiosis. Comparative analyses identified 30 nuclear encoded plastid-targeted proteins in this chromalveolate that originated via endosymbiotic or horizontal gene transfer from multiple different sources. We identify a fundamental divide between plant/red algal and chromalveolate plastid proteomes that reflects a history of mixotrophy in the latter group resulting in a highly chimeric proteome. Loss of phagocytosis in the “red” and “green” clades effectively froze their proteomes, whereas chromalveolate lineages retain the ability to engulf prey allowing them to continually recruit new, potentially adaptive genes through subsequent endosymbioses and horizontal gene transfers. One of these genes is an electron transfer protein (plastocyanin) of green algal origin in *K. brevis* that likely allows this species to thrive under conditions of iron depletion.
Introduction

Oxygenic photosynthesis, the ability to harvest solar energy to fix carbon dioxide and produce organic material evolved in cyanobacteria 2.5-3 billion years ago (Schopf 1993; Hedges et al. 2001). The advent of photosynthesis was arguably the most important event in the evolution of life. It changed the redox balance permitting the development and further metabolic and structural diversification of eukaryotic life forms (Blankenship 2001). Eukaryotes acquired the ability to photosynthesize about 1.5 billion years ago (Hedges et al. 2004; Yoon et al. 2004; Hackett et al. 2006) through the capture and retention of a cyanobacterium in the “host” cytoplasm. This primary endosymbiosis gave rise to the plastid (e.g., chloroplast) in the common ancestor of three extant primary lineages: green algae and land plants (Viridiplantae), red algae (Rhodophyta), and Glaucophyta (Bhattacharya and Medlin 1995; Cavalier-Smith 2004), together known as the Plantae and more recently, the Archaeplastida (Adl et al. 2005). Thereafter, seven other groups of eukaryotes gained their plastid through secondary endosymbiosis, whereby a non-photosynthetic protist engulfed an existing alga and retained its plastid (Bhattacharya, Yoon, and Hackett 2004). Separate secondary endosymbioses with green algae are believed to have given rise to the plastid in euglenoids and chlorarachniophytes. Red algae contributed the plastid to a variety of protist lineages putatively united in the supergroup Chromalveolata that includes alveolates (apicomplexans, ciliates, and dinoflagellates) and chromists (cryptophytes, haptophytes, and stramenopiles) (Cavalier-Smith 1999). The number of red algal secondary endosymbiotic events that gave rise to the chromalveolate plastid is under debate (Bhattacharya, Yoon, and Hackett 2004) although recent data suggest a single origin (Harper and Keeling 2004; Yoon et al. 2005; Li et al. 2006; Weber, Linka, and Bhattacharya 2006). The ancestral red algal plastid was however not maintained in all chromalveolates with its loss in ciliates, its diminution (apicoplast) to a remnant genome in the obligate parasites, apicomplexans, and its replacement with a plastid from a different alga on at least five separate occasions in dinoflagellates (Saldarriaga 2001; Ishida and Green 2002; Bhattacharya, Yoon, and Hackett 2004; Hackett et al. 2006). In *Lepidodinium viride* the broadly distributed peridinin containing dinoflagellate plastid of red algal origin was replaced by one of green-algal origin. In other species the
ancestral plastid was replaced with one from a cryptophyte, stramenopile, or haptophyte through tertiary endosymbiosis (i.e., capture of a plastid of secondary endosymbiotic origin (Hackett et al. 2004a; Yoon et al. 2005).

The dinoflagellates that acquired their plastid from a haptophyte alga are known as fucoxanthin dinoflagellates because of the presence of this light harvesting protein in the endosymbiont. Given this complex series of events, and the potential for large-scale transfer of endosymbiont genes from all of its genomes to the host nucleus (endosymbiotic gene transfer, EGT) (Martin et al. 2002), the complement of plastid proteins (proteome) in fucoxanthin dinoflagellates could quite literally be composed of a menagerie of proteins of red and haptophyte (and potentially other) origins. Much of what we know about plastid proteomes comes from analysis of the green and red lineages that have a significantly simpler evolutionary history with a single cyanobacterial endosymbiosis and no known cases of phagotrophy. Dinoflagellates on the other hand have undergone plastid replacements and are important grazers in the marine environment. The capacity to engulf different algae and bacteria as prey provides the opportunity to accumulate foreign genes in the dinoflagellate nucleus. These horizontally transferred genes could potentially add to or replace genes involved in plastid function (Archibald et al. 2003; Hackett et al. 2004b; Li et al. 2006).

To determine the origins of the plastid proteome in a tertiary endosymbiosis, we reconstructed the evolutionary history of nuclear encoded plastid targeted proteins in the fucoxanthin “Florida red tide” dinoflagellate *Karenia brevis*. The plastid-targeted proteins were identified in two extensive EST libraries from this species. The phylogenetic approach allowed us to reconstruct the chain of endosymbiotic gene transfer events that resulted in the complex plastid proteome in *K. brevis*. We also demonstrate a significant contribution of green algal-derived genes to the chromalveolate plastid proteome. Our data underline the great plasticity of the plastid biochemical machinery and provide insights into the molecular mechanisms of plastid establishment.

**Materials and Methods**
Generating the *K. brevis* ESTs

Two cDNA libraries were generated from clonal *K. brevis* Wilson cells cultured under 25 ± 1°C on a 16:8-hour light-dark cycle. The Dark library was constructed using log phase cells grown on seawater enriched with f/2 medium and during the dark phase of the diel cycle. The cDNA clones were sequenced from the 5’ end. A total of 7,001 ESTs from the Dark library are available from the NCBI expressed sequence tag database (ESTdb and at [www.marinegenomics.org](http://www.marinegenomics.org)), which comprise 5,280 non-redundant gene clusters. For complete details about Dark library generation, sequencing, and processing see Lidie et al. (2005). The non-normalized Stress library was constructed using *K. brevis* Wilson cells grown under five different culture conditions: 1) nitrate-depletion, 2) phosphate-depletion, 3) in log phase under replete conditions, harvested during the light phase 4) in the presence of oxidative metals, 5) undergoing heat stress. Total RNA from each treatment was purified with Trireagent (Molecular Research Center, Cincinnati, OH) and pooled to construct a cDNA library in the Lambda Zap II vector system (Strategene, La Jolla, CA) as previously described. A total of 22,272 ESTs from the Stress library and an additional 9,984 ESTs from the Dark library were sequenced from both 5’- and 3’-termini. This work was done by JGI.

The EST data were processed using Trace2dbest V2.1 (Parkinson and Blaxter 2004). Low complexity regions of the DNA sequences were masked using lower case and RepeatMasker V3.15 ([http://www.repeatmasker.org/](http://www.repeatmasker.org/)). Clustering and assembly of the EST data was done with TGICL ([http://www.tigr.org/tdb/tgi/software/](http://www.tigr.org/tdb/tgi/software/)) (Pertea et al. 2003) and resulted in 9,784 non-redundant gene clusters.

Identification of Plastid Proteins and Phylogenetic Analysis

Plastid-targeting proteins were identified using sequence similarity search (BLAST) against the GenBank non-redundant database (nr). The primary structure of identified sequences has been analyzed to confirm nuclear localization of the gene (presence of poly-A tail, 3’ untranslated region [UTR] and plastid localization of its product [i.e., presence of a 5’ plastid targeting leader peptide]). The 5’ leader peptides
were analyzed using the following protein prediction programs: SignalP (http://www.cbs.dtu.dk/services/SignalP/), TargetP (plant version) (http://www.cbs.dtu.dk/services/TargetP), PSORT (http://www.psort.org/), and PATS (http://gecco.org.chemie.uni-frankfurt.de/pats/pats-index.php). Protein precursors in chromalveolates are considered to be plastid targeted if they contain an N-terminal extension composed of a signal peptide (SignalP probability ≥ 0.7) and transit peptide. Any amino acid sequence separating the signal peptide and mature protein is considered a transit peptide. The length of the N-terminal extension predicted by the computer programs was verified using the protein alignments of the *K. brevis* sequence with its homologs from other taxa. Identified *K. brevis* sequences encoding plastid-targeted proteins have been deposited in GenBank under accession numbers DQ531572 – DQ531601 that are listed in Table 1.


Amino acid sequences for each protein were aligned using ClustalW and manually refined under BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). The 6-protein concatenated protein alignment was assembled using BioEdit. For each data set, a phylogeny was reconstructed under maximum likelihood (ML) using the PHYML V2.4.3 computer program (Guindon and Gascuel 2003; Guindon et al. 2005) using the WAG + Γ + I evolutionary model and tree optimization. The alpha values for the gamma
distribution were calculated using eight rate categories. To assess the stability of monophyletic groups in the ML trees, we calculated PHYML bootstrap (100 replicates) support values (Felsenstein 1985). In addition, we calculated bootstrap values (500 replications) using the neighbor joining (NJ) method with JTT + Γ distance matrices (PHYLIPV3.63, http://evolution.genetics.washington.edu/phylip.html). The NJ analysis was done with randomized taxon addition. We calculated Bayesian posterior probabilities for nodes in the ML tree using MrBayes V3.0b4 (Huelsenbeck and Ronquist 2001) and the WAG + Γ model. The Metropolis-coupled Markov chain Monte Carlo from a random starting tree was run for 1,000,000 generations with trees sampled each 1,000 cycles. The initial 20,000 cycles (200 trees) were discarded as the “burn in.” A consensus tree was made with the remaining 800 phylogenies to determine the posterior probabilities at the different nodes. Manual inspection of the trees produce by the MCMC tree search suggested convergence by 200,000 generations.

Results

We identified sequences encoding 30 different plastid-targeted proteins from two EST libraries of K. brevis (see Table 1). The evolutionary origins of 22 of these genes have been resolved. The sequences encoding K. brevis light harvesting proteins were counted as one protein and were omitted from the phylogenetic analysis due to the presence of multiple paralogs in different photosynthetic lineages. Our phylogenetic analyses of the remaining genes show that most of the K. brevis plastid-targeted proteins are shared by different lineages of chromalveolates. Thirteen of them were derived by the chromalveolate ancestor from its red algal endosymbiont, six proteins originate from green algae, and three plastid-targeted proteins have arisen through the duplication of genes encoding cytosolic proteins in the ancestral chromalveolate (Table 1). After the replacement of the peridinin-containing plastid with the haptophyte-derived plastid, 12 of these genes were acquired by fucoxanthin dinoflagellates from their tertiary haptophyte endosymbiont and one gene was derived from the genome of the dinoflagellate ancestor. Origins of the remaining nine proteins in the tertiary plastid proteome were unresolved. Here we classify
proteins by their origin in the secondary plastid (i.e., red algal-derived) proteome that is typical for most chromalveolates.

Red Algal Contribution to the *K. Brevis* Proteome through Secondary and Tertiary Endosymbioses

The 13 proteins acquired by chromalveolates from their secondary red algal endosymbiont trace their origins to the primary endosymbiosis: i.e., 11 “cyanobacterial” proteins and two proteins of pro-mitochondrial origin recruited for a plastid function before the Archaeplastida divergence. We resolved unambiguously the phylogeny of six of the “cyanobacterial” proteins: ATP synthase gamma subunit (encoded by *atpC*), oxygen-evolving-enhancer 1 protein (*psbO*), cytochrome b6 (*petC*), photosystem II 12kD extrinsic protein (*psbU*), phosphoglycerate mutase (PGAM), and beta-ketoacyl-ACP reductase (*fabG*) (see Fig. S1 in the Supplementary Material). These genes were transferred from a cyanobacterium to a red alga (primary endosymbiosis), from the red alga to the chromalveolate ancestor (secondary endosymbiosis), and finally from a haptophyte alga to fucoxanthin dinoflagellates (tertiary endosymbiosis). To robustly position *K. brevis* within a tree of plastid-targeted nuclear proteins, we generated a 6-protein alignment using these proteins. The resulting tree strongly supports the monophyly of the *K. brevis* and haptophyte sequences and the expected monophyly of red algae and chromalveolates (bootstrap proportions, maximum likelihood, BPml = 100%; neighbor joining, BPnj = 100%; Bayesian posterior probability, BPP = 1.0 for both nodes; see Fig. 1). The red algal-derived plastid clade includes representatives of both groups of chromalveolates, the chromists represented by haptophytes and stramenopiles and the alveolates represented by peridinin dinoflagellates. This tree of nuclear-encoded proteins provides significant support for the origin of all chromalveolates from a single common ancestor that all initially shared a red algal secondary endosymbiont (Bhattacharya, Yoon, and Hackett 2004; Li et al. 2006; Weber, Linka, and Bhattacharya 2006).

The phylogeny of translation elongation factor G (EF-G encoded by *fusA*) is similar to the tree shown in Fig. 1, but is a host contribution to the primary plastid proteome. EF-G catalyzes the
translocation step of protein synthesis in prokaryotic ribosomes including plastids and mitochondria (Breitenberger and Spremulli 1980). Analysis of organellar EF-Gs in photosynthetic algae using immunological techniques revealed that the two proteins have distinct structures (Breitenberger and Spremulli 1980). Surprisingly, plastid EF-G, unlike the mitochondrial protein, has a significant structural similarity to *Escherichia coli* EF-G. This is believed to be a case of convergent evolution (Breitenberger and Spremulli 1980). Our analysis provides a phylogenetic perspective on this issue. The phylogenetic analyses support the origin of plastid *fusA* from alpha proteobacteria (BPml = 99%, BPnj = 100, BPP = 1.0; Fig. 2). It is believed that a *Rickettsia*-like proteobacterium gave rise to the mitochondrion in eukaryotes (Gray 1998). The inclusion of mitochondrial-targeted sequences into the *fusA* tree shows that they form a separate clade that is closely related to delta-proteobacteria, spirochaetes, and planctomycetes (BPml = 100%, BPnj = 100%, BPP = 1.0). Our interpretation of these results is that a *fusA* sequence derived by HGT from a non-photosynthetic bacterium was recruited for protein translation in mitochondria and later replaced the original mitochondrial *fusA*. This recruitment occurred prior to the divergence of plants, animals, and fungi (see Fig. 2). Thereafter, the original pro-mitochondrial *fusA* retained in the host nuclear genome replaced its homolog from the cyanobacterial endosymbiont and gave rise to the plastid *fusA* in the common ancestor of Archaeplastida. Chromalveolates acquired the gene encoding plastid EF-G from the red-algal secondary endosymbiont. In turn, fucoxanthin dinoflagellates gained this gene from the haptophyte tertiary endosymbiont (BPml = 94%, BPnj = 95%, BPP = 1.0).

The evolutionary origin in *K. brevis* of plastid fructose-1,6-bisphosphatase (FBPase) that is a core enzyme of the Calvin cycle and glycolysis, parallels *fusA*. Plastid FBPase arose through the duplication of the primary host cytosolic FBPase that occurred early in eukaryotic evolution. Similarity of the cytosolic FBPase of eukaryotes to its proteobacterial homolog supports the pro-mitochondrial origin of this enzyme (Martin and Schnarrenberger 1997). Here we show that plastid FBPase of chromalveolates was acquired from the red algal secondary endosymbiont (BPml = 68%, BPnj = 69, BPP = 1.0; Fig. 3). The origin of FBPase in tertiary endosymbiosis remains however unresolved.
Green Algal Contribution to the Tertiary Plastid Proteome

Six *K. brevis* plastid-targeted proteins originated in chromalveolates from green algae. These proteins represent components of the primary plastid proteome: four of them are of a cyanobacterial origin and two are of host origin (Table 1, Fig. 4). These “green genes” will be dealt with separately.

Glutamate 1-semialdehyde 2,1-aminomutase (GSA-AT) is an enzyme of the class-III pyridoxal-phosphate-dependent aminotransferase family that is involved in chlorophyll and heme biosynthesis in the plastid stroma in plants and algae (Kannangara and Gough 1978). Phylogenetic analysis shows this protein to have a cyanobacterial provenance (BPml = 86 %, BPP = 1.0) and has originated in chromalveolates from a green algal lateral transfer (BPml = 64%, BPnj = 73%, BPP = 1.0, Fig. 4A). Chromists and alveolates form a monophyletic group (BPml = 81 %, BPnj = 100%, BPP = 1.0) suggesting that acquisition of the green gene encoding GSA-AT by chromalveolates predates the divergence of these two lineages. The position of *K. brevis* relative to haptophytes and peridinin dinoflagellates is however unresolved.

Plastid ferredoxin NADP(H) reductase (FNR encoded by *petH*). FNRs are flavoprotenoids that catalyze the reversible electron transfer between NADP(H) and electron carrier proteins ferredoxin or flavodoxin (Arakaki, Ceccarelli, and Carrillo 1997). Plants contain two tissue-specific types of plastidic FNRs. One of these is present only in plastids from photosynthetic leaf tissue and the second isoform is present in plastids from non-photosynthetic organs such as root, fruits, and petals (Ceccarelli et al. 2004). The two isoforms have arisen through duplication of a cyanobacterial gene that occurred early in the evolution of the Archaeplastida. After the divergence of the three primary photosynthetic lineages, the ancestral gene was apparently lost in the red algae and the glaucophytes, whereas the green algae lost the “leaf” type of *petH*. We find that chromalveolates have only one isoform of *petH* (Fig. 4B). The phylogenetic analysis strongly supports (BPml = 100%, BPnj = 89%, BPP = 1.0) the monophyly of chromists and the “green” clade of the “root” (green-algal) type. *PetH* in alveolates, apicomplexan, and peridinin dinoflagellates belong to the “root” type. However these sequences are highly divergent and we are unable to resolve the position of alveolates in the *petH* tree. The substitution of the red-algal gene with
one of green-algal origin presumably occurred in the common ancestor of chromalveolates (Fig. 4B). The petH tree topology supports the monophyly of fucoxanthin dinoflagellates and haptophytes implying that fucoxanthin dinoflagellates acquired their “green” petH from the tertiary endosymbiont.

Plastocyanin (Pc encoded by petE) is a small (ca. 100–150 aa) thylakoid lumen copper-binding protein that is an electron carrier between Photosystems II and I (Gross 1993). Plastocyanin is believed to be unique to plants, green algae, and some cyanobacteria (Ho and Krogmann 1984; Gross 1993). In red algae and in chromalveolates that contain a red algal-derived plastid, electron transport from cytochrome b6/f to Photosystem I is achieved via the iron binding analog of plastocyanin, cytochrome C6 (CytC6) (Sandmann et al. 1983; Price et al. 1991). The complete genome sequences of Cyanidioschyzon merolae, Galdieria sulfuraria, and Thalassiosira pseudonana confirm the absence of petE encoding plastocyanin from the genomes of red algae and stramenopiles. Both, petE and CytC6 have a cyanobacterial origin. Many cyanobacteria and green algae still retain both genes. In these organisms, CytC6 is the functional backup of plastocyanin that is activated under copper deficiency (Wood 1978; Ho, Krogmann 1984).

We identified 18 clones encoding the petE sequence in the K. brevis EST libraries; i.e., two clones from the “dark” library and 16 clones from the “stress” library (see Materials and Methods). These ESTs are encoded by five closely related genes that have arisen through recent gene duplications. BLAST searches against NCBI identified petE sequences from K. micrum (ABA55542.1). In addition, we found two petE encoding ESTs in the haptophyte alga Emiliania huxleyi (CX773129 and CX773088). These two sequences have identical 3'UTRs indicating that they are derived from a single locus. The K. brevis and E. huxleyi petE sequences are polyadenylated, encode a complete petE ORF, have a stop codon in the position conserved among all photosynthetic eukaryotes, and a 3'UTR. The E. huxleyi petE sequences do not include an N-terminal extension, whereas K. micrum petE encodes an N-terminal extension that has a structure typical for dinoflagellate plastid targeting signals; i.e., it contains a cleavable 19 aa signal peptide that directs the protein precursor to the endoplasmic reticulum (ER) followed by a plastid targeting transit peptide (Fig S2A). Interestingly, the N-terminal extension of all K. brevis petE sequences differ from the canonical dinoflagellate plastid-targeting peptides (Fig S2B). The N-terminal 48 aa form a
highly hydrophobic, putatively uncleaved signal peptide (P = 0.979 according to SignalP) followed by a 92 aa sequence. The C-terminal thylakoid-spanning motif of the *K. brevis* plastocyanin leader peptide shares significant similarity with the corresponding sequence in *K. micrum* plastocyanin. According to PSORT (http://www.psort.org/) and TMpred (http://www.ch.embnet.org/software/TMPRED_form.html), a 48 aa segment can be classified as a type II signal anchor peptide that orients the protein in the N’-cytoplasm/C’-ER lumen topology in the ER membrane. We propose that plastocyanin targeting to the plastid involves some alternative mechanism in *K. brevis*. Study of the prion protein, a brain glycoprotein involved in various neurodegenerative diseases, provides an example of an alternative protein trafficking mechanism (Lopez et al. 1990). The prion protein precursor contains an uncleaved N-terminal segment that functions as a membrane-anchor motif in cell-free systems. However in vivo prion protein precursors are translocated across the ER membrane with the assistance of cytosolic factors and trafficked to the cell surface.

The distribution of petE suggests that haptophytes and fucoxanthin dinoflagellates acquired this gene either from a green alga or a cyanobacterium. Phylogenetic analysis of the protein provides strong support (BPml = 89%, BPP = 1.0) for the divergence of haptophyte and dinoflagellate petE within the green clade (Fig. 4C). Based on this tree topology, we suggest that the haptophyte ancestor acquired petE from a green alga. Thereafter fucoxanthin dinoflagellates captured this gene from the haptophyte tertiary endosymbiont. CytC6 has not been found in *K. brevis*, but it was identified in the EST library of *K. micrum* (PEPdb accession number KME00008061; http://amoebidia.bcm.umontreal.ca/public/pepdb/welcome.php), in different species of haptophytes including *E. huxleyi* (GenBank accession number CX774985), and in peridinin dinoflagellates (*Alexandrium tamarense, Heterocapsa triquetra, Lingulodinium polyedrum*, and *Amphidinium carterae*). Thus, fucoxanthin dinoflagellates, like green algae, possess both electron carrier proteins. The results of a recent microarray analysis show that *K. brevis* petE is expressed under both normal and iron-limited culture conditions (KL Lidie and F Van Dolah unpublished data). PetE transcript accumulation is apparently not affected by a change in iron concentration in the medium. Based on these observations, we
hypothesize that in fucoxanthin dinoflagellates, like in green algae, copper-binding plastocyanin is used to transfer electrons from the cytochrome b6/f to Photosystem I dominant over CytC6. The expression pattern of CytC6 in fucoxanthin dinoflagellates remains to be determined.

Gamma-tocopherol O-methyltransferase (\(\gamma\)-Tmt) is the final enzyme of the \(\gamma\)-tocopherol (vitamin E) synthesis pathway that is found only in photosynthetic organisms (Sattler et al. 2003). It is believed that tocopherol protects membrane fatty acids from oxidative degradation by scavenging reactive oxygen species produced by photosynthesis (Fryer 1993). In plants and algae, tocopherol biosynthesis occurs on the plastid membrane. Until now, proteins involved in tocopherol biosynthesis have been reported only in plants, green and red algae, and cyanobacteria. We found an EST encoding the entire \(\gamma\)-Tmt coding region in \(K. \)brevis. Partial \(\gamma\)-Tmt sequences have been identified in the EST libraries of \(K. \)micrum and a haptophyte alga Isochrysis galbana (PEPdb accession numbers KME00005944 and ISE00003655, respectively). The bootstrap (BPml = 87\%, BPnj = 91\%) and Bayesian analyses (BPP = 0.99) provide support for the monophyly of green algae, plants, haptophytes, and fucoxanthin dinoflagellates suggesting the same scenario for the origin of \(\gamma\)-Tmt in these genomes as described above for plastocyanin (Fig. 4D).

Soluble inorganic pyrophosphatases (sPPases) are ubiquitous enzymes that are responsible for the removal of inorganic pyrophosphate produced by a variety of vital biosynthetic reactions (Perez-Castineira et al. 2001). Eukaryotic cells contain several isoforms of sPPase located in the cytosol and mitochondrion of animals and fungi, in the plastid and mitochondrion of photosynthetic eukaryotes, and only in the mitochondrion of heterotrophic protists. All of these isoforms, with the exception of plant and green algal mitochondrial sPPase, arose though the duplication of the eukaryotic gene, which replaced the bacterial homologs of the proteobacterial and cyanobacterial endosymbionts. Analysis of the plastid sPPase distribution shows this protein to be present in both photosynthetic plastid-bearing chromalveolates as well as non-photosynthetic apicoplast-bearing apicomplexans. We however excluded the highly divergent apicoplast targeted proteins from the analysis (results not shown). The sPPase phylogenetic tree strongly supports the monophyly of chromalveolates and the origin of this gene from a
green alga (BPml = 96%, BPP = 0.96; Fig. 4E). Similar to GSA-AT and petH, “green” sPPase originated in chromalveolates prior to the split of chromists and alveolates. The strongly supported (BPml = 100% BPnj = 100, BPP = 1.0) position of K. brevis within the haptophyte clade indicates that this gene originated in fucoxanthin dinoflagellates from the tertiary endosymbiont.

Serine protease IV (sppA) is a thylakoid-bound stroma-exposed protease involved in the light-induced plastid protein degradation in plants (Lensch, Herrmann, and Sokolenko 2001). Phylogenetic analyses show that sppA encoding plastid-targeted proteins from all photosynthetic taxa form a monophyletic group (BPml = 100%, BPnj = 100 %, BPP = 1.0; Fig. 4F) suggesting a common ancestry. The plastid-targeted sppA genes presumably have arisen in Archaeplastida through the duplication of mitochondrial sppA (BPml = 95%, BPnj = 94%, BPP = 1.0). However, an unresolved phylogeny within the mitochondrial/eubacterial clade leaves open the possibility for an origin of plastid sppA through HGT from a non-cyanobacterium. The nested position of the K. brevis plastid sppA inside of the green algal/plant clade (BPml = 93%, BPnj = 99%, BPP = 1.0) suggests a green algal origin of the gene. Absence of plastid sppA homologs from other chromalveolate taxa does not allow us to infer the time of entry of this “green” gene into the chromalveolate lineage.

The Secondary Host Contribution to the Tertiary Plastid Proteome

In chromalveolates, plastid targeted fructose-1,6-bisphosphate aldolase (FAB II) and glyceraldehyde-3-phosphahte dehydrogenase (GAPDH) arose through the duplication of the secondary host cytosolic isoforms that replaced homologs of plastid-targeted proteins of the secondary endosymbiont (Harper and Keeling 2003; Patron, Rogers, and Keeling 2004). The results of our phylogenetic analysis of K. brevis and K. micrum FAB II support the haptophyte origin of this gene in fucoxanthin dinoflagellates that is consistent with previously reported FAB II phylogenies (Patron, Waller, and Keeling 2006).

The cytosolic GAPDH that is the precursor of chromalveolate plastidic GAPDH belongs to the most common form of these enzymes, NADH dependent GAPDH (NAD-GAPDH) that is involved in glycolytic and gluconeogenic pathways in all organisms (Forthergill-Gilmore and Michels 1993). In
plants, plastid NAD-GAPDH, which originated through the duplication of the gene encoding a plant cytosolic NAD-GAPDH, is involved in glycolysis that takes place in non-photosynthetic plastids and in the dark in chloroplasts (Plaxton 1996). Photosynthetic CO₂ assimilation in plant, green algal, and red algal plastids is catalyzed by the cyanobacterial-derived NADPH dependent GAPDH (Martin et al. 1993). In chromalveolate plastids that lack the NADPH dependent GAPDH, NAD-GAPDH presumably is involved in both the Calvin cycle and glycolysis.

Several NAD-GAPDH phylogenies including fucoxanthin dinoflagellates have been previously published (Takishita, Ishida, and Maruyama 2004; Yoon et al. 2005). These data show the dinoflagellate nuclear genome to encode two cytosolic forms. One of them (gapC3) has been derived either from spirochaetes or from euglenoids by horizontal gene transfer, whereas the second (gapC2) is a normal vertically inherited isoform of NAD-GAPDH. Duplication of the gapC2 gave rise to the gene encoding the plastid-targeting protein. The plastid-targeted NAD-GAPDH found in fucoxanthin dinoflagellates has been derived from the haptophyte tertiary endosymbiont. Analyses of the *K. brevis* EST libraries allowed us to identify four distinct isoforms of GAPDH: two cytosolic (gapC2 and gapC3) and two plastid-targeted (gapC1-pd and gapC1-fd) isoforms (Fig. S3). One of the plastid isoforms is similar to the described previously gapC1-fd (Yoon et al. 2005) and clusters with plastid-targeted sequences of the haptophyte algae, whereas another isoform, gapC1-pd, is positioned within the clade that includes plastid-targeted proteins from peridinin dinoflagellates (Fig. 5). These results demonstrate for the first time that *K. brevis* retains both tertiary endosymbiont haptophyte-derived and secondary endosymbiont red algal-derived forms of NAD-GAPDH.

Discussion

**Plastid Establishment and Endosymbiotic Gene Transfer**

The evolution of the fucoxanthin dinoflagellate plastid proteome was shaped by primary, secondary, and tertiary endosymbioses (Yoon et al. 2005). Study of the green and red algal plastids that resulted from primary endosymbiosis have until now received the greatest attention. Using these models, it was
demonstrated that establishment of the primary plastid proteome involved the integration of the molecular machineries of both the endosymbiont and the host (Martin et al. 2002; Richly and Leister 2004). This process was facilitated by large-scale gene movement from the cyanobacterial endosymbiont to the host nuclear genome, with the subsequent re-targeting of these gene products to the plastid (Martin et al. 1998). Using present estimates, ca. 900 – 1200 plastid-targeted proteins encoded by the *Arabidopsis thaliana* nuclear genome and 676 proteins in the thermoacidophilic red alga *Cyanidioschyzon merolae* are of cyanobacterial origin (Reumann, Inoue, and Keegstra 2005; Sato et al. 2005). For *Arabidopsis*, these numbers comprise up to 50% of its nuclear-encoded plastid-targeted proteins. The remaining proteins (around 1200) are derived from the host nuclear or pro-mitochondrial genomes, or from various sources via HGT. Phagocytosis is considered to be the major mechanism in free living eukaryotes for the acquisition of new genes via HGT (Doolittle 1998). The inability of virtually all red and green algae to carry out phagocytosis has most certainly limited the influx of new genetic material through HGT into these lineages. However, it is likely that HGT played an important role in the early formation of the Archaeplastida primary plastid proteome (Martin and Schnarrenberger 1997) because phagotrophy was an ancestral character in this group that most likely allowed the capture of the primary plastid.

Following primary endosymbiosis red and green plastids were acquired by other protist lineages through secondary endosymbiosis (Bhattacharya, Yoon, and Hackett 2004). In secondary endosymbiosis, most of the genes encoding plastid-targeted proteins were transferred from the endosymbiont nucleus to the new host nucleus and reused for plastid function and regulation. The results of our phylogenetic analysis of the plastid-targeted proteins confirm this prediction with 13 out of 30 proteins in *K. brevis* being derived by chromalveolates from a red alga through secondary endosymbiosis (Table 1). This number includes 11 proteins of cyanobacterial origin and two proteins (*fusA* and FBPase) derived from the pro-mitochondrial genome of the primary host (Figs. 6A, B).

Despite the large-scale influx of endosymbiont genes into the host nuclear genome, the secondary plastid proteome should not be considered simply as a copy of that in the primary plastid. Phylogenetic analyses show significant changes in the proteome of the red algal-derived plastid that occurred in the
chromalveolates during the establishment of secondary endosymbiosis. These changes can be classified into two major types. The first involves the substitution of cyanobacterial genes with host paralogs. Two examples are GAPDH and FBA (Harper and Keeling 2003; Patron, Rogers, and Keeling 2004). Both of these substitutions occurred prior to the divergence of the chromalveolate lineages (Fig. 6B). The second type of change in the red algal-derived plastid proteome is the substitution of genes derived from the red algal endosymbiont with green algal or bacterial homologs. This category also includes the acquisition of “foreign” genes, homologs of which are absent from red algae. There are two possible, but not mutually exclusive explanations for the occurrence of “foreign” genes in chromalveolates, 1) continuous HGT over chromalveolate evolution, or 2) an ancient green algal endosymbiosis potentially preceding the red algal-derived plastid in the chromalveolate ancestor. The latter idea was proposed by Funes et al. (2002) as an explanation for the occurrence of the green algal derived mitochondrial protein COXII in apicomplexans.

The red algal secondary endosymbiosis that gave rise to the chromalveolate plastid occurred about 1.3 billion years ago (Yoon et al. 2004). However, unlike Archaeplastida, many extant photosynthetic chromalveolates still retain the ability for phagocytosis (Kugrens and Lee 1990; Kawachi et al. 1991; Wilcox and Wedemayer 1991; Jones, Leadbeater, and Green 1994) that makes possible the influx of new genes via HGT. If continuous HGT is a major mechanism for chromalveolate plastid proteome evolution, we would expect to see the substitutions of red algal-derived genes occurring at different time points in the chromalveolate tree. Three examples of such substitutions have been reported so far. One of them is plastid protein delta-aminolevulinic acid dehydrogenase ($\text{hemB}$) that has a green origin in dinoflagellates and a red origin in stramenopiles (Hackett et al. 2004b). A second example is the substitution of the normal plastid encoded RuBisCo with its proteobacterial homolog that is nuclear encoded in dinoflagellates (Palmer 1996). And finally, a plastid-targeted protein encoded by fused genes for the shikimate biosynthetic enzyme $AroB$ and an O-methyltransferase (OMT) derived by HGT from a cyanobacterium has been found in dinoflagellates (Waller, Slamovits, and Keeling 2006). These examples provide evidence for the importance of intra- and interdomain HGT in chromalveolate plastid evolution. Other reported cases of plastid proteins of non-red algal origin in chromalveolates include two “green”
genes, chlorophyll A synthase (*chlG*, Li et al. 2006) and phosphoribulokinase (PRK, Li et al. 2006; Petersen et al. 2006) that are shared by different chromalveolate lineages. Red algal derived homologs of these genes have not yet been found in chromalveolates.

Analysis of the *K. brevis* EST libraries identified six proteins of green algal origin: GSA-AT, *petH*, and sPPase that are shared by chromists and alveolates, *petE* and γ-Tmt that are shared by haptophytes and fucoxanthin dinoflagellates, and *sppA*, homologs of which have not been found in other lineages of chromalveolates (Fig. 6B, C). Phylogenetic analyses of *petE* and γ-Tmt show that tertiary EGT from the haptophyte endosymbiont is the most plausible explanation for the presence of these genes in fucoxanthin dinoflagellates (Figs. 6C, D). There are two possible scenarios for the occurrence of these genes in haptophytes. Under the first scenario, haptophytes gained *petE* and γ-Tmt through an independent HGT. The observation that many extant species of photosynthetic haptophytes still retain the ability for phagotrophy (Kawachi et al. 1991; Jones, Leadbeater, and Green 1994) makes plausible this scenario. Under the second scenario, these genes entered the chromalveolate nuclear genome prior to the divergence of the constituent lineages and have subsequently been lost from stramenopiles and alveolates after the divergence of the cryptophyte-haptophyte branch (for the chromalveolate phylogeny see Hackett et al. 2006). Limited molecular data from cryptophyte algae preclude us from reaching a decisive conclusion on the time of *petE* and γ-Tmt recruitment by chromalveolates. Phylogenies of GSA-AT, *petH*, sPPase, PRK, and *chlG* however suggest that a major influx of green genes occurred early in the evolution of chromalveolates prior to the divergence of the chromists and alveolates.

It remains a daunting task to distinguish between the scenarios that the “green” genes are derived by EGT from a single endosymbiont prior to the establishment of the “red” plastid or alternatively via HGT from multiple green algal donors after this event. One possible approach to resolving this issue is to compare the phylogenetic composition of foreign genes acquired by Chromalveolata before and after the split of its constituent lineages. The present data show that the list of organisms that donated genes to the chromalveolate lineages after their divergence from each other includes green algae, proteobacteria, and
cyanobacteria (see above). However all foreign genes acquired by chromalveolates before their split are derived from a single donor lineage, the green algae. One reasonable explanation for these data is the presence of a green algal endosymbiont in the chromalveolate ancestor prior to the establishment of the red algal-derived plastid. Unfortunately, the red algal endosymbiosis likely resulted in the substitution of most green algal nuclear genes of plastid function with red algal homologs, effectively erasing the major evidence for the previous endosymbiosis.

The results of our study show that secondary endosymbiosis brought together components of highly diverged red and green algal photosynthetic machineries. Examples are \textit{psbU}, that is unique in red algae and green algal plastocyanin, structural homologs of which are absent from the red algal lineage. A significant flexibility of the light-harvesting complex (LHC) in plastids has been previously demonstrated (Grabowski, Cunningham, and Gantt 2001), showing that LHC proteins from a red alga, \textit{Porphyridium cruentum}, could functionally bind pigments that had evolved separately in different evolutionary lineages (e.g., plants, stramenopiles, dinoflagellates). The successful incorporation of green algal-derived plastid proteins into the proteome of the red plastid demonstrated by our study is consistent with these results (Grabowski, Cunningham, and Gantt 2001). These data support the significant biochemical plasticity of key components of plastid energy metabolism, such as the electron transport chain and the pyrophosphate utilization machinery.

Tertiary endosymbiosis introduced an additional level of complexity to the process of plastid regulation. In this case, the genomes of both the endosymbiont and the host contain genes of plastid function. The discovery of nuclear-encoded plastid-targeted PSBO in fucoxanthin dinoflagellates that originated from the haptophyte tertiary endosymbiont led Ishida and Green (Ishida and Green 2002) to propose the replacement of ancestral red algal derived genes in these photosynthetic taxa with the haptophyte endosymbiont homologs. Until recently, this was the leading hypothesis for the origin of the tertiary plastid proteome. Patron, Waller, and Keeling (2006) however reported several plastid-targeted proteins of the ancestral peridinin-dinoflagellate-type that were retained in the nuclear genome of the fucoxanthin dinoflagellate \textit{K. micrum}. These proteins include thylakoid-bound ascorbate peroxidase and
phosphoribulokinase. Based on their involvement in processes unrelated to photosynthesis, Patron, Waller, and Keeling (2006) suggested that the fucoxanthin dinoflagellate ancestor contained a non-photosynthetic (apicoplast-like) plastid that required a reduced set of genes to express its function. Thus, genes of the haptophyte tertiary endosymbiont may not have replaced, but rather reestablished the photosynthetic machinery with genes that were lost prior to this endosymbiotic event. Analysis of the *K. brevis* EST data turned up 13 genes encoding plastid-targeted proteins of haptophyte origin: *atpC*, *psbO*, *petC*, *psbU*, PGAM, *fabG*, *petH*, *petE*, γ-Tmt, *fusA*, sPPase, FBA II, and *gapC1-fd* (Fig. 6C).

In addition to haptophyte-derived GAPDH encoded by *gapC1-fd*, we found the ancestral red algal derived *gapC1-pd* sequence in *K. brevis*. This the first report of both ancestral and tertiary endosymbiont-derived genes encoding homologous plastid-targeted proteins in fucoxanthin dinoflagellates. This suggests that fucoxanthin dinoflagellates likely retain two functional sets of other, yet undetected nuclear genes encoding plastid-targeted proteins. However, the clear dominance of expressed genes of haptophyte provenance suggests that most ancestral genes encoding plastid-targeted proteins have been lost or replaced after the establishment of the tertiary endosymbiont. This situation may mirror the more ancient substitution of green algal genes with genes derived from the red algal endosymbiont postulated above.

### Serial Endosymbiosis and HGT as Mechanisms for Adaptation in Chromalveolates

Chromalveolate algae are the most abundant group of eukaryotic marine phytoplankton. It is however unclear why these taxa have risen to pre-eminence. Grzebyk et al. (2003) posited as explanation the presence of a larger set of genes involved in photosynthesis and energy transduction in chromalveolate plastids than in reds and greens (plastid portability hypothesis) and structural traits of the chromalveolate host cell such as an armored cell wall as potential explanations.

We propose that the ability to recruit a new genetic material through serial endosymbiosis and HGT is a key chromalveolate trait that provides the ability for rapid adaptation to changing environmental conditions, thereby increasing their evolutionary fitness. For example, *Karenia brevis* is remarkably tolerant to low iron conditions (10⁻⁹M) that will not support the growth of diatoms (A Neeley, G DiTullio,
FM Van Dolah unpublished data). Green algal-derived petE is an example of a gene that is absent in red algae and provides a novel mechanism in fucoxanthin dinoflagellates that may contribute to this trait. Like green algae and some cyanobacteria, haptophytes and fucoxanthin dinoflagellates possess two types of functionally homologous electron carrier proteins, copper-binding plastocyanin and iron-binding CytC6. In fucoxanthin dinoflagellates, like in green algae and cyanobacteria, plastocyanin is the main protein carrying electrons between the cytochrome b6/f complex and Photosystem I. Based on our results we suggest that the acquisition of plastocyanin is an adaptation of fucoxanthin dinoflagellates to iron-limitation in its natural habitat. Another known adaptation to iron deficiency in many algae and cyanobacteria is a physiological switch from the iron–sulfur protein ferredoxin to the non-iron protein flavodoxin in the electron transport chain (Erdner et al. 1999; Geiss et al. 2001). Analysis of the K. brevis EST data show that both proteins, ferredoxin and flavodoxin, are present in this fucoxanthin dinoflagellate. Consistent with its role in facilitating photosynthetic electron transport under low iron conditions, flavodoxin transcript levels increase, whereas those of ferredoxin decrease under iron limitation (KL Lidie and FM Van Dolah unpublished data), and flavodoxin protein expression is induced under iron limitation, whereas only ferredoxin is present in nutrient replete conditions (A Neeley, G DiTullio, and FM Van Dolah unpublished data). Thus, the two types of plastid electron carrier proteins, red algal-derived flavodoxin (Fig. 6B, Table 1) and green algal-derived plastocyanin, likely provide fucoxanthin dinoflagellates with the ability to thrive under conditions of iron deficiency in their natural habitats.
Supplementary Material

Figures S1-S3 and the list of contributors in the JGI Production Sequencing Group are available at *Molecular Biology and Evolution* online.
Acknowledgements

The *K. brevis* EST data were produced under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program and the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC03-76SF00098 and Los Alamos National Laboratory under contract No. W-7405-ENG-36. This was Community Sequencing Program project #200407. D. B. and T. N. were supported by grants from the National Science Foundation and the National Aeronautics and Space Administration awarded to D. B. (EF 04-31117, NNG04GM17G). T. N. was partially supported by an Avis E. Cone Fellowship from the University of Iowa. F. M. V. D. and K. L. L. acknowledge support from the NOAA HAB Initiative.
Literature Cited


Lidie KB, Ryan JC, Barbier M, Van Dolah FM. 2005. Gene expression in Florida red tide dinoflagellate
Karenia brevis: analysis of an expressed sequence tag library and development of DNA microarray.

Lopez CD, Yost CS, Prusiner SB, Myers RM, Lingappa VR. 1990. Unusual topogenic sequence directs

Martin W, Brinkmann H, Savonna C, R Cerff. 1993. Evidence for a chimeric nature of nuclear genomes:
eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. Proc Natl Acad
Sci USA 90:8692-8696.

plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci USA
99:12246-12251.

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


270:93-126.


Fig. 1. - A 6-protein plastid phylogeny using genes implicated in tertiary EGT in *K. brevis*. This ML tree was inferred from the combined plastid-targeted protein sequences of *atpC*, *psbO*, *petC*, *psbU*, PGAM, and *fabG*. The numbers above and below the branches are the results of ML and NJ bootstrap analyses, respectively. Only bootstrap values ≥ 60% are shown. The thick branches indicate ≥ 0.95 posterior probability from Bayesian inference. Branch lengths are proportional to the number of substitutions per site (see scale bars). Chr, chromista; Alv, alveolates; Ra, red algae; Gr, green algae/plants; CB, cyanobacteria. Numbers in bold indicate bootstrap supports for the monophyly of chromalveolates and red algae and *K. brevis* and haptophytes.

Fig. 2. - ML tree of plastid- and mitochondrial-targeted translation elongation factor G (*fusA*). The numbers above and below the branches are the results of ML and NJ bootstrap analyses, respectively. Only bootstrap values ≥ 60% are shown. The thick branches indicate ≥ 0.95 posterior probability from Bayesian inference. Branch lengths are proportional to the number of substitutions per site (see scale bars). The lineage designations are as in Fig. 1. Numbers in bold indicate bootstrap support for the monophyly of plastid and alpha-proteobacterial clades, and *K. brevis* and haptophytes.

Fig. 3. - ML tree of plastid targeted and cytosolic fructose-1, 6-bisphosphatase (FBPase) sequences. The numbers above and below the branches are the results of ML and NJ bootstrap analyses, respectively. Only bootstrap values ≥ 60% are shown. The thick branches indicate ≥ 0.95 posterior probability from Bayesian inference. Branch lengths are proportional to the number of substitutions per site (see scale bars). The lineage designations are as in Fig. 1. Numbers in bold indicate bootstrap support for the monophyly of plastid and cytosolic isoforms of FBPase, and chromalveolates and red algae.

Fig. 4. - Proteins of green algal origin in the *K. brevis* tertiary plastid proteome. (A) ML tree of GSA-AT. (B) ML tree of *petH*. (C) ML tree of *petE*. (D) ML tree of γ-Tmt. (E) ML tree of sPPase. (F) ML tree of
sppA. The numbers above and below the branches are the results of ML and NJ bootstrap analyses, respectively. Only bootstrap values ≥ 60% are shown. Dashes represent bootstrap values below 60%. The thick branches indicate ≥ 0.95 posterior probability from Bayesian inference. Branch lengths are proportional to the number of substitutions per site (see scale bars). The lineage designations are as in Fig. except that Gl is glaucophytes. Numbers in bold indicate bootstrap support for the monophyly of chromalveolates and green algae/plants and fucoxanthin dinoflagellates *K. brevis* and *K. micrum* and haptophytes.

Fig. 5. - ML tree of chromalveolate plastid targeted NAD-GAPDH. The numbers above and below the branches are the results of ML and NJ bootstrap analyses, respectively. Only bootstrap values ≥ 60% are shown. The thick branches indicate ≥ 0.95 posterior probability from Bayesian inference. Branch lengths are proportional to the number of substitutions per site (see scale bars). Chr, chromista; Alv, alveolates. Numbers in bold indicate bootstrap support for the monophyly of *gapC1-pd* isoform of *K. brevis* and peridinin dinoflagellates and *gapC1-fd* isoform of fucoxanthin dinoflagellates and haptophytes. This tree represents a portion of the complete NAD-GAPDH tree provided in supporting materials (Fig. S3).

Fig. 6. - Origin of plastid-targeted proteins in *K. brevis*. Fucoxanthin dinoflagellate plastid proteome evolution through primary, secondary, and tertiary endosymbioses is shown from top to bottom. The sources of *K. brevis* genes encoding plastid targeted proteins are denoted using different colors in three sections of the figure corresponding to primary, secondary, and tertiary endosymbioses. (a) Genes derived from the primary host through vertically (black font/arrow) and cyanobacteria through EGT (blue font/arrow). (b) Genes derived from the secondary host vertically (black font/arrow), red algal endosymbiont through EGT (red font/arrow), and green algae through HGT (green font/arrow). (c) Genes derived by fucoxanthin dinoflagellates from the peridinin dinoflagellate host vertically (black font/arrow) and haptophyte tertiary endosymbiont through EGT (purple font/arrow). The question mark indicates
uncertainty about the time for the gene transfer. * - numbers of cyanobacteria- and host-derived genes are indicated for *Arabidopsis* plastid proteome (Reumann, Inoue, and Keegstra 2005).
Table 1. Plastid targeted proteins in the fucoxanthin dinoflagellate *Karenia brevis*.

<table>
<thead>
<tr>
<th>#</th>
<th>Protein name</th>
<th>Gene/Protein abbreviation</th>
<th>Accession number</th>
<th>Leader peptide</th>
<th>Gene source in plastid proteome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1’</td>
</tr>
<tr>
<td>1</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase isoform C1-fd</td>
<td>gapC1-fd/ GAPDHp-fd</td>
<td>DQ531601*</td>
<td>st</td>
<td>Cyt</td>
</tr>
<tr>
<td>2</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase isoform C1-pd</td>
<td>gapC1-pd/ GAPDHp-pd</td>
<td>DQ531575</td>
<td>st</td>
<td>Cyt</td>
</tr>
<tr>
<td>3</td>
<td>Cytochrome b6-f complex iron-sulfur subunit</td>
<td>petC/ RISP</td>
<td>DQ531586</td>
<td>t</td>
<td>CB</td>
</tr>
<tr>
<td>4</td>
<td>Ferredoxin</td>
<td>petF/ Fd</td>
<td>DQ531588*</td>
<td>st</td>
<td>CB</td>
</tr>
<tr>
<td>5</td>
<td>Ferredoxin NADPH reductase</td>
<td>petH/ FNR</td>
<td>DQ531589</td>
<td>st</td>
<td>CB</td>
</tr>
<tr>
<td>6</td>
<td>Flavodoxin</td>
<td>isiB /Flv</td>
<td>DQ531584*</td>
<td>st</td>
<td>CB</td>
</tr>
<tr>
<td>7</td>
<td>Oxygen-enhancer 1 protein</td>
<td>psbO</td>
<td>DQ531590</td>
<td>st</td>
<td>CB</td>
</tr>
<tr>
<td>8</td>
<td>Photosystem II 12kD extrinsic protein</td>
<td>psbU</td>
<td>DQ531591*</td>
<td>st</td>
<td>CB</td>
</tr>
<tr>
<td>9</td>
<td>Fructose-1,6-bisphosphate aldolase class II</td>
<td>fbaC1/ FBA-II</td>
<td>DQ531583</td>
<td>st</td>
<td>Cyt</td>
</tr>
<tr>
<td>10</td>
<td>Adenylate kinase</td>
<td>ADK</td>
<td>DQ531597</td>
<td>-</td>
<td>CB</td>
</tr>
<tr>
<td>11</td>
<td>Soluble inorganic pyrophosphatase</td>
<td>sPPase</td>
<td>DQ531593*</td>
<td>st</td>
<td>Euk</td>
</tr>
<tr>
<td>12</td>
<td>ATP synthase gamma subunit</td>
<td>atpC</td>
<td>DQ531581</td>
<td>st</td>
<td>CB</td>
</tr>
<tr>
<td>13</td>
<td>Plastocyanin</td>
<td>petE/ Pc</td>
<td>DQ531587*</td>
<td>at</td>
<td>CB</td>
</tr>
<tr>
<td>14</td>
<td>Phosphoglycerate mutase</td>
<td>PGAM</td>
<td>DQ531576</td>
<td>st</td>
<td>CB</td>
</tr>
<tr>
<td>15</td>
<td>Fructose-1,6 bisphosphatase</td>
<td>fbp/ FBPase</td>
<td>DQ531600</td>
<td>st</td>
<td>Mt_Cyt</td>
</tr>
<tr>
<td>16</td>
<td>Putative ribulose-1,5 bisphosphate carboxylase</td>
<td>RBCM</td>
<td>DQ531595</td>
<td>st</td>
<td>Mt_Cyt</td>
</tr>
<tr>
<td>17</td>
<td>LH chlorophyll binding proteins</td>
<td>not analyzed</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Electron transport and carbohydrate metabolism*
<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Gene</th>
<th>Accession</th>
<th>Signal Peptide</th>
<th>Cyanobacteria</th>
<th>Red Algae</th>
<th>Green Algae</th>
<th>Haptophytes</th>
<th>Peridinin Dinoflagellates</th>
<th>Fucoxanthin Dinoflagellates</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>Translation elongation factor G fusA/EF-G</td>
<td></td>
<td>DQ531572</td>
<td></td>
<td>Mt</td>
<td>RA</td>
<td>H</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Translation elongation factor Ts EF-Ts</td>
<td></td>
<td>DQ531577</td>
<td>st</td>
<td>CB</td>
<td>nr</td>
<td>nr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>DNA gyrase subunit A gyrA</td>
<td></td>
<td>DQ531573</td>
<td></td>
<td>CB</td>
<td>RA</td>
<td>nr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Beta-ketoacyl-ACP reductase fabG/ BKR</td>
<td></td>
<td>DQ531582</td>
<td>st</td>
<td>CB</td>
<td>RA</td>
<td>H</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Clp protease</td>
<td></td>
<td>DQ531574</td>
<td></td>
<td></td>
<td></td>
<td>nr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Clp protease</td>
<td></td>
<td>DQ531578</td>
<td>st</td>
<td>CB</td>
<td>nr</td>
<td>nr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>DegP serine-type peptidase degP</td>
<td></td>
<td>DQ531596</td>
<td>t</td>
<td>CB</td>
<td>RA</td>
<td>nr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Periplasmic serine protease IV sppA</td>
<td></td>
<td>DQ531592</td>
<td>st</td>
<td>CB</td>
<td>GA</td>
<td>nr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>M48-like peptidase</td>
<td></td>
<td>DQ531585</td>
<td>st</td>
<td>CB</td>
<td>GA</td>
<td>nr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>RNA helicase VDL</td>
<td></td>
<td>DQ531579</td>
<td>st</td>
<td></td>
<td></td>
<td>nr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Glutamate 1-semialdehyde 2,1-aminomutase</td>
<td></td>
<td>DQ531598</td>
<td></td>
<td>CB</td>
<td>GA</td>
<td>nr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Gamma-tocopherol O-methyltransferase _-Tmt</td>
<td></td>
<td>DQ531594</td>
<td>st</td>
<td>CB</td>
<td>GA</td>
<td>nr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Conserved hypothetical protein</td>
<td></td>
<td>DQ531580</td>
<td>st</td>
<td>CB</td>
<td>RA</td>
<td>nr</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The sequence represents one member of a closely related family of proteins in *K. brevis*; s - signal peptide (> 0.9 SignalP probability); a - signal anchor peptide (> 0.9 SignalP probability); t - transit peptide; 1’ – Archaeplastida; 2’ – Chromalveolata; 3’ – fucoxanthin dinoflagellates. CB - cyanobacteria; RA - red algae; GA - green algae; H - haptophytes; PD - peridinin dinoflagellates; Mt
- gene of mitochondrial or pro-mitochondrial origin; Cyt - host gene encoding cytosolic protein; Euk – host gene of eukaryotic origin (if the original gene copy was lost from the host genome); nr- protein phylogeny was not resolved.
Origin of Plastid Targeted Proteins in Fucoxanthin Dinoflagellates

**HOST GENES**

**ENOSYMBIOTIC / HORIZONTAL GENE TRANSFER**

**A**

**PRIMARY ENDOSYMBIOSIS**

- FBPase
- fusA
- sppA
- sPPase (~1200 genes*)

**Large-scale EGT from Cyanobacterium (~900 genes*)**

**Rhodophytes**

**B**

**SECONDARY ENDOSYMBIOSIS**

- gapC1
- FBAII

**Chromalveolates**

- atpC
- degP
- fabG
- FBPase
- Flv
- fusA
- gyrA
- petC
- petE
- sPPase
- Tmt

**C**

**TERTIARY ENDOSYMBIOSIS**

- gapC1-pd

**Fucoxanthin Dinoflagellates**

- atpC
- fabG
- FBAII
- fusA
- petC
- petE
- sppA
- Tmt