Tertiary Endosymbiosis Driven Genome Evolution in Dinoflagellate Algae

Hwan Su Yoon,* Jeremiah D. Hackett,* Frances M. Van Dolah,† Tetyana Nosenko,* Kristy L. Lidie,‡ and Debashish Bhattacharya* 

*Department of Biological Sciences and Roy J. Carver Center for Comparative Genomics, University of Iowa; and †Biotoxins Program, NOAA National Ocean Service, Center for Coastal Environmental Health and Biomolecular Research, Charleston, South Carolina

Dinoflagellates are important aquatic primary producers and cause “red tides.” The most widespread plastid (photosynthetic organelle) in these algae contains the unique accessory pigment peridinin. This plastid putatively originated via a red algal secondary endosymbiosis and has some remarkable features, the most notable being a genome that is reduced to 1–3 gene minicircles with about 14 genes (out of an original 130–200) remaining in the organelle and a nuclear-encoded proteobacterial Form II Rubisco. The “missing” plastid genes are relocated to the nucleus via a massive transfer unequalled in other photosynthetic eukaryotes. The fate of these characters is unknown in a number of dinoflagellates that have replaced the peridinin plastid through tertiary endosymbiosis. We addressed this issue in the fucoxanthin dinoflagellates (e.g., Karenia brevis) that contain a captured haptophyte plastid. Our multiprotein phylogenetic analyses provide robust support for the haptophyte plastid replacement and are consistent with a red algal origin of the chromalveolate plastid. We then generated an expressed sequence tag (EST) database of 5,138 unique genes from K. brevis and searched for nuclear genes of plastid function. The EST data indicate the loss of the ancestral peridinin plastid characters in K. brevis including the transferred plastid genes and Form II Rubisco. These results underline the remarkable ability of dinoflagellates to remodel their genomes through endosymbiosis and the considerable impact of this process on cell evolution.

Introduction

It is now generally accepted that the photosynthetic organelle of eukaryotes (plastid) originated through endosymbiosis (Gray 1992; Bhattacharya and Medlin 1995; Palmer 2003; Bhattacharya, Yoon, and Hackett 2004), whereby a single-celled protist engulfed and retained a foreign photosynthetic cell. Over time, the foreign cell was reduced to a plastid and was vertically transmitted to subsequent generations. Most plastids have originated either through primary or secondary endosymbiosis. The first results from the engulfment of a photosynthetic prokaryote (cyanobacterium) and gave rise to a plastid bound by two membranes, whereas the second results from the engulfment of an eukaryotic alga and resulted in a plastid bound by three or four membranes. Only the dinoflagellates have undergone tertiary endosymbiosis, which is the engulfment of an alga containing a secondary plastid (Bhattacharya, Yoon, and Hackett 2004). The primary endosymbiosis is believed to have given rise to the plastid in the common ancestor of the red, green, and glaucophyte algae (Moreira, Le Guyader, and Philippe 2000; Stibitz, Keeling, and Bhattacharya 2000; MatsuZaki et al. 2004; Sanchez Puerta, Bachvaroff, and Delwiche 2004).

After their split from the green algae, a red algal cell was engulfed by a nonphotosynthetic protist and reduced to a plastid. This cell evolved chlorophyll c and putatively was the common ancestor of the protist superassemblage chromalveolates (Cavalier-Smith 1999), comprising the Chromista (cryptophytes, haptophytes, and stramenopiles, Cavalier-Smith 1986) and the Alveolata (dinoflagellates, ciliates, and apicomplexans, Van de Peer and De Wachter 1997). Although not yet recovered in host gene trees, studies of plastid genes and plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) are consistent with chromalveolate monophyly (e.g., Durnford et al. 1999; Fast et al. 2001; Yoon et al. 2002; Harper and Keeling 2003; Yoon et al. 2004).

Within dinoflagellates, the most common plastid type is bound by three membranes and contains the unique accessory pigment, peridinin. Other significant characteristics of peridinin plastids are a highly reduced genome encoding about 14 proteins on 1–3 gene minicircles, in addition to the large and small subunits of the plastid rRNA and minicircles encoding pseudogenes (Zhang, Green, and Cavalier-Smith 1999; Barbrook and Howe 2000; Zhang, Cavalier-Smith, and Green 2002; Howe et al. 2003). The distribution of minicircle genes in different dinoflagellates remains uncertain. Although apparently plastid encoded in taxa such as Procorcentrum spp. and Amphidinium spp. (Zhang, Green, and Cavalier-Smith 1999; Koumandou et al. 2004), these genes have been localized to the nucleus in Ceratium horridum (Laatsch et al. 2004) and provisionally also in Pyrocystis lunula (rpl28 and rpl33, unpublished data; see GenBank accession number AF490367). Plant and algal plastids generally contain a circular genome that is about 150 kb in size and encode between 130 and 200 genes. In contrast, the minicircles encode the core subunits of the photosystem (atpA, atpB, perB, perD, psaA, psaB, psbA–E, and psbI) and rRNA (16S, 23S rRNA), and the remaining genes required for plastid function have been transferred to the nucleus (e.g., in Alexandrium tamarense, Amphidinium carterae, and Lingulodinium polyedrum; Bachvaroff et al. 2004; Hackett et al. 2004) and encode a targeting sequence for plastid import (Nassoury, Cappadocia, and Morse 2003; Bachvaroff et al. 2004; Hackett et al. 2004). In addition to this remarkable development, the normal plastid-encoded Form I Rubisco was replaced in peridinin dinoflagellates with a nuclear Form II enzyme of alpha-proteobacterial origin.

As noteworthy as this set of evolutionary developments may seem, in some dinoflagellates, the peridinin plastid was replaced by one from a cryptophyte, a haptophyte, a stramenopile, or a green alga (Bhattacharya, Yoon, and Hackett...
The genomic consequences for dinoflagellates of these tertiary endosymbioses (except for the green algal replacement which was a successive secondary event) remain unknown. For example, were the nuclear-encoded plastid genes completely lost or is there now a set of genes, potentially of nonoverlapping functions, of both chromalveolate and tertiary endosymbiotic origin, and what of the Form II Rubisco? Was this gene lost or are there genes encoding both Forms I and II enzymes in these taxa? To answer these questions, we studied in detail one particular dinoflagellate tertiary endosymbiosis, the replacement of the peridinin plastid with one of haptophyte origin in the Gymnodiniales. The plastids in taxa such as *Karenia* spp., *Karlodinium micrum*, and *Takeyama* spp. contain chlorophylls c1 + c2 and 19'-hexanoyloxy-fucoxanthin and/or 19'-butanoyloxy-fucoxanthin but lack peridinin (e.g., Daugbjerg, Hansen, and Moestrup 2000; De Salas et al. 2003), similar to the haptophyte algae. Phylogenetic analyses support a haptophyte origin of fucoxanthin plastids (Tengs et al. 2000; Ishida and Green 2002), but their relationship to peridinin plastids has never been convincingly demonstrated in the context of a taxonomically broadly sampled multigene phylogeny. A previous DNA-based attempt at resolving this question using plastid- and minicircle-encoded *psbA* (Yoon, Hackett, and Bhattacharya 2002) led to the artificial clustering of fucoxanthin and peridinin dinoflagellates. This misleading result was most likely caused by codon-usage heterogeneity in the DNA sequences (Inagaki et al. 2004). Use of protein data appears, therefore, to be of critical importance in resolving dinoflagellate plastid relationships.

Given these existing data, our study had two major aims. The first was to establish, using robust multiprotein (photosystem I P700 chlorophyll a apoprotein A1 [PsaA], photosystem I P700 chlorophyll a apoprotein A2 [PsaB], photosystem II reaction center protein D1 [PsbA], photosystem II 44 kDa apoprotein [PsbC], and photosystem II D2 reaction center protein [PsbD]) analyses, the position of fucoxanthin plastids in a tree that included peridinin, red, chromist, green algal, glaucophyte, and cyanobacterial sequences. The second aim was to determine the impact of tertiary plastid endosymbiosis on nuclear genome evolution. To do this, we analyzed a unigene set of 5,138 expressed sequence tags (ESTs) that we generated from a light- and dark-harvested culture of the toxic fucoxanthin dinoflagellate *Karenia brevis*.

### Materials and Methods

#### Taxon Sampling and Sequencing

We determined the sequence of five minicircle-encoded genes (*psaA*, *psbA*, *psbC*, and *psbD*) in peridinin dinoflagellates and their putative plastid-encoded homologs from fucoxanthin-containing taxa as well as from chromist, red, and glaucophyte algae. A total of 60 new sequences were determined in this study (see table 1 in the Supplementary Material).

The algal cultures were frozen in liquid nitrogen and ground with glass beads using a glass rod and/or Mini-BeadBeater™ (Biospec Products Inc., Bartlesville, Okla.). Total genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Santa Clarita, Calif.). Polymerase chain reactions (PCR) were done using specific primers for each of the genes (Yoon, Hackett, and Bhattacharya 2002; Yoon et al. 2002, 2004). Degenerate primers were designed for the *psbC* gene (psbC60F: TGC YTG GTG GWC WGG TAA TGC; psbC100F: GGT AAR TTM YTM GGT GCT CAT; psbC1160R: TCT TGC CAS GTY TGS ATA TCA; psbC1200R: CCT ASS GTC GCA TGR TCA TA) and *psbD* (psbD50F: GAT GAT TGG YAA AAC GAG A; psbD90F: TTT TWT TTA TCG GYT GGT CYG G; psbD1010R: CAA GAT CAR CCW CAT GAA AAC; psbD1040R: GAG GAG GTW TTA CCA CGT GGA AA). PCR products were purified using the QiAquick PCR Purification kit (Qiagen) and were used for direct sequencing using the BigDye™ Terminator Cycle Sequencing Kit (PE- Applied Biosystems, Norwalk, Conn.), and an ABI-3100 at the Roy J. Carver Center for Comparative Genomics at the University of Iowa. Some PCR products were cloned into pGEM-T vector (Promega, Madison, Wis.) prior to sequencing.

#### Phylogenetic Analyses

The DNA sequence for each photosystem gene was translated, and the amino acid data for each protein were manually aligned with the available data from GenBank using SeqPup (Gilbert 1995). Ambiguous positions were excluded from the phylogenetic analysis (the alignment is available upon request from D.B.). We analyzed each amino acid data set individually as well as a concatenated data set of PsA (264 aa), PsB (277 aa), PsB (319 aa), PsB (334 aa), PsB (296 aa) (total of 1,490 aa). Because we were unable, in spite of repeated attempts, to PCR amplify the *psbA* gene from *K. brevis*, *Karenia mikimotoi,* and *Heterocapsa triquetra* and the *psbC* gene from *K. mikimotoi,* these sequences were coded as missing data in the alignment. The cyanobacterium *Nostoc* sp. PCC7120 was used to root the tree.

The maximum likelihood (ML) method was used to infer the phylogenetic relationships of the different plastids using the concatenated data. We did separate analyses for data sets that included all dinoflagellates or excluded the peridinin-containing taxa. The ML analyses were done with proml in PHYLIP V3.6b (Felsenstein 2004) using the JTT + Γ evolutionary model (Jones, Taylor, and Thornton 1992). The alpha values for the gamma distribution were calculated using Bayesian inference with MrBayes V3.0b4 (Huelsenbeck and Ronquist 2001). The mean value for the gamma parameter was calculated from the post “burn-in” trees (see below) in the posterior distribution. Global rearrangements and randomized sequence input order with five replicates were used to find the best ML tree. To test the stability of monophyletic groups in the ML trees, 100 bootstrap replicates were analyzed with proml as described above (Felsenstein 1985), except that one round of random taxon addition was used in the analysis. In the Bayesian inference of the amino acid data, we used the WAG + Γ model (Whelan and Goldman 2001). Metropolis-coupled Markov chain Monte Carlo from a random starting tree was run for 1,000,000 generations with
trees sampled each 500 cycles. Four chains were run simultaneously, of which three were heated and one was cold, with the initial 250,000 cycles (500 trees) discarded as the burn-in. A consensus tree was made with the remaining 1,500 trees to determine the posterior probabilities at the different nodes. Minimum evolution (ME) bootstrap (500 replicates) analyses were done using PUZZLEBOOT V1.03 (http://hades.biochem.dal.ca/Rogerlab/Software/software.html) and Tree-Puzzle V5.1 (Schmidt et al. 2002) to generate the WAG + Γ distance matrices. For the individual plastid proteins, an ML tree was inferred using proml and the JTT + Γ model as described above. One hundred bootstrap replicates were analyzed with phyML V2.4.3 (Guidon and Gascuel 2003) and the JTT + Γ model to infer support for nodes in these trees. Bayesian posterior probabilities for nodes were calculated with MrBayes using the WAG + Γ model. We also analyzed a data set of publicly available chromalveolate cytosolic and plastid-targeted GAPDH sequences that included the novel K. brevis data using ML, ME, and Bayesian methods as described above for the five-protein alignment.

In addition to the protein data analyses, we also inferred a tree from the concatenated DNA sequences of the five minicircle genes. This ML tree was inferred using PAUP* (Swoford 2002) and the site-specific general time-reversible (GTR) model (Rodriguez et al. 1990) with different evolutionary rates for each codon position. Bootstrap analyses were done using phyML (100 replicates) and the ME method (GTR + Γ + I model, using PAUP*). The Bayesian posterior probability was calculated using MrBayes and the site-specific GTR model.

Tree Topology Tests

To assess the positions of the dinoflagellate plastids in our tree, we used MacClade V4.05 (D. R. Maddison and W. P. Maddison 2002) to generate 38 alternative topologies from that shown in the best ML tree (fig. 1). In these trees, the fucoxanthin clade was moved to 16 alternative positions, and Peridinium foliaceum and peridinin clade to 8 and 14 alternative positions, respectively. The one-sided Kishino-Hasegawa (KH) test (Goldman, Anderson, and Rodrigo 2000) was implemented using Tree-Puzzle V5.1 to assign probabilities to the best ML and the alternative trees.

We also implemented the approximately unbiased (AU) test. Because of the high computing time required to calculate the site-by-site likelihoods, we generated the WAG + Γ distance matrices. For the individual plastid proteins, an ML tree was inferred using proml and the JTT + Γ model as described above. One hundred bootstrap replicates were analyzed with phyML V2.4.3 (Guidon and Gascuel 2003) and the JTT + Γ model to infer support for nodes in these trees. Bayesian posterior probabilities for nodes were calculated with MrBayes using the WAG + Γ model. We also analyzed a data set of publicly available chromalveolate cytosolic and plastid-targeted GAPDH sequences that included the novel K. brevis data using ML, ME, and Bayesian methods as described above for the five-protein alignment.

In addition to the protein data analyses, we also inferred a tree from the concatenated DNA sequences of the five minicircle genes. This ML tree was inferred using PAUP* (Swoford 2002) and the site-specific general time-reversible (GTR) model (Rodriguez et al. 1990) with different evolutionary rates for each codon position. Bootstrap analyses were done using phyML (100 replicates) and the ME method (GTR + Γ + I model, using PAUP*). The Bayesian posterior probability was calculated using MrBayes and the site-specific GTR model.

Results and Discussion

Chromalveolate Phylogeny and Plastid Evolution

Under the chromalveolate hypothesis, all taxa containing chlorophyll c (i.e., a chromophytic plastid) share a single common ancestor and comprise two monophyletic clades, the chromists and the alveolates (Cavalier-Smith 2000). Phylogenetic analyses provide mixed support for this plastid-based view of eukaryotic relationships. Analysis of nuclear genes indicates a single origin of alveolates (e.g., Gajadhar et al. 1991; Baldauf et al. 2000; Stechmann and Cavalier-Smith 2003), but chromist monophyly is only indirectly supported by plastid trees (Dumford et al. 1999; Yoon et al. 2002, 2004; Hagopian et al. 2004), and the haptophytes and cryptophytes have unresolved positions in nuclear and mitochondrial gene analyses (e.g., Bhattacharya and Weber 1997; Sanchez Puerta, Bachvaroff, and Delwiche 2004). Nuclear gene trees do, however, suggest a specific relationship between the stramenopiles and alveolates (Van de Peer and De Wachter 1997; Baldauf et al. 2000; Nozaki et al. 2003). Despite this uncertainty regarding both the monophyly and the internal
branching order, the union of chromalveolate taxa is poten-
tially confirmed by the existence of a gene replacement in
which the cytosolic GAPDH gene was duplicated and re-
targeted to the plastid, uniquely in these taxa (Fast et al. 2001;

Furthermore, in contrast to the convincing evidence for a red algal secondary endosymbiotic origin of the chromist plastid (most parsimoniously, via a single event; Yoon et al. 2002, 2004), verifying the source of the alveolate plastid has proven more challenging. The ciliates have apparently lost this organelle, whereas the parasitic apicomplexans retain a remnant plastid (apicoplast) that contains a reduced genome of about 35 kb in size (Williamson et al. 1994). The apicoplast genes are highly enriched in A’s and T’s (e.g., 86.9% AT-content in Plasmodium falciparum, Wilson et al. 1996) and therefore have extreme divergence rates, rendering them of limited utility in phylogenetic anal-
yses (e.g., Kohler et al. 1997; Hackett et al. 2004). The api-
coplast is hypothesized to be of chromalveolate descent (e.g., McFadden and Waller 1999) or to have originated through a replacement of the ancestral plastid with another of eukaryotic provenance, resulting in the four-membrane–
bound organelle in these taxa (Bodyl 1999). The dinoflagel-
late minicircle genes are also highly divergent and pose problems for tree reconstruction, in particular, when using DNA sequences (e.g., due to codon-usage heterogeneity) and single proteins or with limited taxon sampling (e.g., Zhang, Green, and Cavalier-Smith 1999; Yoon, Hackett, and Bhattacharya 2002; Inagaki et al. 2004). For these rea-
sons, we chose to address dinoflagellate plastid origin using a taxonomically broadly sampled data set of five proteins that are encoded on minicircles in peridinin dinoflagellates and are putatively borne on the haptophyte-derived plastid genome in the fucoxanthin-containing taxa K. brevis and

Fig. 1.—Phylogenetic analysis of algal and plant plastids. This ML tree was inferred from the combined plastid protein sequences of PsaA, PsaB, PsbA, PsbC, and PsbD. The results of an ML bootstrap analysis are shown above the branches, whereas the values below the branches result from an ME bootstrap values

The branch lengths are proportional to the number of substitutions per site (see scale in figure). The probabilities (using the one-sided KH test) for placing the peridinin plastid clade in three alternate positions are shown in the smaller gray boxes with the arrows.
K. mikimotoi. We assume here that the minicircle genes are ultimately of plastid origin whether they are now localized in this organelle or in the nucleus.

Dinoflagellate Plastid Origin

The ML plastid tree inferred from the concatenated protein (1,490 aa) data set is shown in figure 1. This tree has significant Bayesian support for all the nodes, and the chromalveolate plastids (excluding apicoplasts) form a monophyletic group within the red algae as sister to the non-Cyanidiales clade. The fucoxanthin dinoflagellate plastids are positioned within the haptophytes as sister to Emiliania huxleyi and Isochrysis sp. (Isochrysidales, Prymnesiophyceae; Edvardsen et al. 2000), whereas the monophyletic peridinin plastids diverge within the stramenopiles as sister to the diatoms Skeletonema costatum and Odontella sinensis. Peridinium foliaceum is robustly positioned within this clade, confirming the origin of this dinoflagellate plastid through tertiary endosymbiosis (Chesnix, Morden, and Schmieg 1996; Chesnix et al. 1997; Schnepf and Elbrachter 1999). The bootstrap analyses provide, however, only weak to moderate support for many of the nodes in the tree. To determine whether the highly divergent dinoflagellate clades are the cause of this phenomenon (i.e., due to their long branches), we removed in separate analyses either the peridinin or the fucoxanthin plastid sequences from the alignment and recalculated the bootstrap values. In these trees, the peridinin and fucoxanthin clades were positioned in the same place as shown in figure 1 (i.e., within the stramenopile plastids for the peridinin clade and within the haptophyte plastids for the fucoxanthin clade [e.g., see fig. S1 in the Supplementary Material]). These data suggest that the divergent dinoflagellate sequences do not exhibit significant long-branch attraction in our protein trees. However, removal of the peridinin plastid sequences results in a marked increase in bootstrap support for most of the nodes of interest in the tree, now with moderate support for chromalveolate monophyly (ML: 49%–62%, ME: 86%–86%) and strong support for haptophyte + fucoxanthin dinoflagellate (ML: 68%–93%, ME: <50%–73%) and stramenopile plastid monophyly (ML: 100%, ME: 96%) as shown in the gray boxes in figure 1 (see fig. S1). Removal of the fucoxanthin clade did not, however, change significantly the bootstrap values in the resulting tree (unpublished data). Our multiprotein data, therefore, provide convincing evidence for the independent evolutionary origins of fucoxanthin (putatively plastid encoded) and peridinin (minicircle encoded) genes in dinoflagellates, consistent with previous studies (Tengs et al. 2000; Ishida and Green 2002; Takahata et al. 2005). Analysis of the individual protein data sets (see fig. S2 in the Supplementary Material) shows significant topological instability, as would be expected for trees that are derived from single proteins (often) with variable divergence rates among taxa.

To test the positions of the fucoxanthin and peridinin dinoflagellate plastids in figure 1, we generated 38 alternative topologies that tested the divergence point of these tax (including Peridinium) and assessed their probabilities using the one-sided KH test. All 16 trees in which the fucoxanthin dinoflagellate clade was moved to alternative positions in figure 1 (e.g., to the base of the chromalveolates \( P < 0.001 \)) or to the base of the peridinin clade \( P = 0.01 \)) were significantly rejected \( P < 0.05 \). Seven alternative positions for P. foliaceum were also significantly rejected \( P < 0.000 \) except for the monophyly of this taxon with the Odontella + Skeletonema clade \( P = 0.122 \). In contrast, of 14 alternative divergence points for the peridinin dinoflagellate clade, the nonrejected positions included at the bases of the stramenoples, stramenoples + (haptophytes), and chromalveolates \( P = 0.33, 0.07, 0.05 \), respectively [see fig. 1]).

We also generated two sets of 25 topologies of a 14-taxon backbone ML tree that excluded the dinoflagellates, in which the fucoxanthin and peridinin clades were added individually to each branch. These pools of topologies were analyzed with the AU test. The results of this analysis are shown in figure 2 and lead to some clear conclusions about dinoflagellate plastid phylogeny. First, the fucoxanthin dinoflagellate plastids receive overwhelming support for their origin from a prymnesiophyte tertiary endosymbiont \( P = 0.995 \), and all alternative positions have significantly lower probabilities (fig. 2A). Second, the peridinin dinoflagellate plastids clearly are not related to this organelle in haptophytes (as suggested in Yoon, Hackett, and Bhattacharya 2002), but rather, their most likely position is sister to the diatom O. sinensis \( P = 0.799 \). However, as is apparent in figure 2B, many alternative positions are also included in the confidence set (i.e., \( P > 0.05 \)) of trees, in particular, in the region defining the radiation of the Cyanidiales and non-Cyanidiales red algae and the cryptists. The divergence point predicted from the chromalveolate hypothesis (as sister to the cryptists, see filled circle in fig. 2B) has a probability of 0.153.

Analysis of the 16-taxon ML backbone tree with the AU test provides further support for the conclusions described above (fig. 2C). Here, the Karenia sp. plastids are positioned as sister to the prymnesiophyte alga as in figure 1 and as strongly suggested in figure 2A. The tree of highest probability in this pool specified a sister relationship between the peridinin plastids and O. sinensis \( P = 0.805 \); the position as sister to the cryptists had \( P = 0.145 \). There were again many alternative positions possible for the peridinin plastids within the confidence set. Their divergence within the haptophytes was rejected except as sister to the Karenia spp. sequences \( P = 0.059 \). However, this likely results from long-branch attraction between the relatively highly divergent dinoflagellate sequences because, in the absence of the fucoxanthin plastids, there is no support for a specific relationship between Akashiwo + Heterocapsa and the haptophytes (see fig. 2B). Our results are therefore consistent with (but do not prove) the chromalveolate hypothesis that posits a red algal origin of the plastid in this group. Intriguingly, the weak support that we find for a specific relationship between peridinin and stramenopile plastids, which is also supported by the analyses of nuclear genes (e.g., Van de Peer and De Wachter 1997; Baldauf et al. 2000; Nozaki et al. 2003; Hackett et al. 2004), suggests a paraphyletic cryptista. Under this (speculative) scenario, the stramenoples and alveolates share a specific relationship independent of the cryptista and haptophytes.
We also inferred a tree using the DNA sequences of the plastid genes (see fig. S3 in the Supplementary Material) and found that these more extensive data still recover the artifactual clustering of peridinin and fucoxanthin plastids within the haptophytes that was reported in Yoon, Hackett, and Bhattacharya (2002). A recent paper by Inagaki et al. (2004) suggests that this misplacement of the peridinin clade (using the DNA data) is explained by similar codon-usage patterns for constant leucine, serine, and arginine residues in PsbA among fucoxanthin dinoflagellates and some haptophytes. Usage of the protein data corrects for this codon-usage heterogeneity and clearly supports the independent origins of these two types of dinoflagellate plastids.

**Endosymbiotic Gene Replacement**

Additional support for the haptophyte plastid replacement in *K. brevis* comes from analysis of GAPDH sequences. A previous study (Ishida and Green 2002) provided phylogenetic evidence for the haptophyte origin of the nuclear-encoded photosystem gene, oxygen-evolving enhancer protein 1 (*psbO*), in *K. brevis* putatively through tertiary endosymbiotic gene replacement. Inspection of our EST data set turned up both cytosolic and plastid-targeted GAPDH genes in *K. brevis*. Phylogenetic analysis of these data shows that the *K. brevis* plastid-targeted GAPDH gene is distantly related to this sequence in peridinin dinoflagellates but rather is sister to the prymnesiophyte, *Isochrysis galbana*, within the haptophyte clade (fig. 3; see also Takishita, Ishida, and Maruyama 2004). This indicates that the ancestral plastid-targeted gene of chromalveolate origin in *K. brevis* was presumably replaced by the haptophyte homolog (fig. 3). In contrast, the cytosolic GAPDH in *K. brevis* is nested within a monophyletic clade of dinoflagellate plastids that were likely vertically inherited in these taxa. These results are consistent with the multiprotein plastid tree shown in figure 1 and support a prymnesiophyte source for the fucoxanthin dinoflagellate plastid.

**Nuclear Genome Transformation in Fucoxanthin Dinoflagellates**

Given the relatively robust view of dinoflagellate plastid evolution that has resulted from our study, we then asked what happened to the nuclear genome of fucoxanthin dinoflagellates after the haptophyte plastid replacement? In particular, what was the fate of the previously transferred nuclear genes of plastid function demonstrated in peridinin dinoflagellates and of the proteobacterial Rubisco? To address these issues, we generated a genomic data set of 5,138 unique ESTs from *K. brevis*. We first searched the *K. brevis* data set for homologs of each of the 48 nuclear-encoded plastid-targeted genes that were uncovered in the EST unigene set from *A. tamarense* (Hackett et al. 2004) and from other dinoflagellate EST projects (e.g., Bachvaroff et al. 2004). None of these genes were present in the *K. brevis* EST set. In particular, peridinin-containing *A. tamarense* encodes 15 photosynthetic genes in its nucleus (atpE, atpF, atpH, petA, psaC, psaI, psbF, psbH, psbI, psbK, psbL, psbN, psbT, rpl2, and rps19) that are restricted, in all known cases, to the plastid genome of other algae and plants (Hackett et al. 2004). These landmark nuclear markers of peridinin dinoflagellates were absent from our *K. brevis* EST set. In addition, there was no evidence of a nuclear-encoded Form II Rubisco gene, although...
The typical Form I sequence of haptophyte origin has previously been isolated (Yoon, Hackett, and Bhattacharya 2002). These results imply that the haptophyte tertiary endosymbiosis resulted in a genome transformation in K. brevis, whereby it lost the unique plastid characters typical of most peridinin dinoflagellates.

Although we have no direct evidence for plastid localization of the K. brevis photosysytem genes used in this study because we did not determine the 5'-terminus of the genes (i.e., to detect a potential plastid-targeting sequence), analysis of the G + C-content of these sequences shows them to have a nucleotide content that is typical of plastid genes (see table 2 in the Supplementary Material). In a comparison of genes known to be encoded in the plastid genome of the red alga Porphyra purpurea (39.8% G + C) and in the red algal–derived plastid in E. huxleyi (41.6% G + C), the Karenia spp. photosystem genes had a similar G + C-content of 41.3% (K. brevis) and 41.8% (K. mikimotoi). In contrast, the nuclear-encoded genes in K. brevis had a markedly higher G + C-content of 52.8%, a trait shared with these genes in A. tamarense (i.e., 60.0%). These data are consistent with the idea that the K. brevis genes are located in the plastid but do not allow us to determine whether they are encoded on a typical genome or on minicircles. The most parsimonious explanation is that the haptophyte plastid replacement resulted in a typical genome in this species, but this hypothesis awaits verification through the direct sequence analysis of this organelle-encoded DNA (T. Nosenko and D. Bhattacharya, unpublished data).

It should be also noted that although the majority of the ESTs were derived from the dark-grown K. brevis library, we were able to identify many plastid-targeted genes in these libraries. Comparison of our ESTs to sequence databases using a Blast e-value cutoff of 1 x 10^-10 resulted in the identification of a number of transcripts that encode plastid-targeted proteins usually found in the nucleus, including numerous light-harvesting proteins, flavodoxin, ferredoxin, plastocyanin, GAPDH, and fructose-1,6-bisphosphate aldolase class I and II (J. D. Hackett, T. Nosenko, and D. Bhattacharya, unpublished data). We also recognize that our EST data potentially represent a fraction of the K. brevis nuclear gene complement; therefore, additional cDNA sequencing (D. Bhattacharya and F. M. Van Dolah with the Joint Genome Institute, unpublished data) may yet result in the identification of some coding regions of plastid function in this species that are typical of peridinin dinoflagellates. Normally, however, photosynthetic genes are found in the initial 200–300 EST sequences determined from a light-harvested dinoflagellate cDNA library (e.g., in A. tamarense, Hackett et al. 2004).

Conclusions

A model of dinoflagellate plastid evolution that summarizes the current state of knowledge is presented in figure 4. Presently, the most parsimonious explanation is that the alveolate ancestor contained a plastid of red algal secondary endosymbiotic origin (Fast et al. 2001; Bhattacharya, Yoon, and Hackett 2004). This organelle was lost in the ciliates, and its genome was significantly reduced in the parasitic apicomplexans and underwent a major transformation.
(PT1) at the base of the dinoflagellates (Bachvaroff et al. 2004; Hackett et al. 2004). Major evolutionary innovations in the dinoflagellates were the origin of minicircle genes and large-scale plastid gene transfer to the nucleus. Our EST data show that a second plastid genome transformation (PT2) occurred following the haptophyte tertiary endosymbiosis that resulted in the loss of PT1 characters and reversion to a state typical of “normal” algae (e.g., origin of Form I Rubisco, absence of peridinin). The presence of an intact, typical plastid genome in fucoxanthin dinoflagellates remains to be verified. The well-supported sister group relationship of these taxa with peridinin-containing species in the Gymnodiniales (e.g., *Akashiwo* spp.; Daugbjerg et al. 2000; Zhang, Bhattacharya, and Lin 2005) strongly suggests, however, that the fucoxanthin plastid is derived from one which contained minicircles (e.g., as in *A. sanguinea*; see fig. 1). Loss of the existing nuclear-encoded plastid genes in the *K. brevis* ancestor, which presumably were shielded from Muller’s ratchet in the nucleus, may indicate that selection favors coevolved proteins (i.e., in the captured haptophyte plastid genome) rather than a mixture of haptophyte and ancient chromalveolate origin. Alternatively, the regulation of plastid function may be more efficient when the 15 landmark nuclear-encoded plastid genes of peridinin dinoflagellates are transcribed and translated in the organelle (see Hackett et al. 2004; Koumandou et al. 2004).

We also uncovered a second case of endosymbiotic gene transfer in which the existing plastid-targeted GAPDH in *K. brevis*, presumably of chromalveolate origin, was replaced by the homolog from the haptophyte tertiary endosymbiont (as for psbO, see Ishida and Green 2002). An intriguing question that remains is whether PT2-type transformations have also occurred in other dinoflagellates that have undergone plastid replacement such as *P. foliaceum* (diatom plastid, see fig. 1) and *Lepidodinium viride* (green algal plastid; e.g., Watanabe et al. 1991) or whether these taxa have adopted different strategies for incorporating the genomic information encoded in the captured organelle. In conclusion, our data underline the remarkable ability of dinoflagellates to transform their genomes through endosymbiosis and identify these protists as ideal models for understanding this critical process in eukaryotic evolution.

**Supplementary Material**

Tables 1 and 2 and figures S1–S3 are available at *Molecular Biology and Evolution* online (www.mbe.oupjournals.org).

**Acknowledgments**

This work was supported by grants from the National Science Foundation awarded to D.B. (DEB 01-07754, MCB 02-36631) and from NOAA/ECOHAB to F.M.V.D.

**Literature Cited**


Kamykowski, D., E. J. Milligan, and R. E. Reed. 1998. Biochem-
eries in the orientation of the autotrophic dinofla-
Goldman, N., J. P. Anderson, and A. G. Rodrigo. 2000. Likeli-
———. 2004. PHYLIP (phylogenetic inference package). Ver-
———. 2003. A simple, fast, and accurate
———. 2004. Dinoflagellate tertiary endosymbiosis 1307
Dinoflagellate Tertiary Endosymbiosis 1307
Downloaded from https://academic.oup.com/mbe/article-abstract/22/5/1299/1066953 by guest on 12 January 2019

sequences of Sarccosystis maris, Thelthria annulata and Cryptothecodium cohnii reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. Mol. Biochem.
Parasitol. 45:147–154.
Gilbert, D. G. 1995. SeqPup, a biological sequence editor and analysis program for Macintosh computer. Indiana University, Bloomington, Ind.
Goldman, N., J. P. Anderson, and A. G. Rodrigo. 2000. Likeli-
Guidon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum like-
relationships in the orientation of the autotrophic dinofla-
Nassoury, N., M. Cappodocia, and D. Morse. 2003. Plastid ultra-
Swofford, D. L. 2002. PAUP*: phylogenetic analysis using parsi-


Geoffrey McFadden, Associate Editor

Accepted February 21, 2005