JMB

A Tertiary Plastid Uses Genes from Two Endosymbionts

Nicola J. Patron, Ross F. Waller and Patrick J. Keeling*

Canadian Institute for Advanced Research, Department of Botany, University of British Columbia Vancouver, Canada V6T 1Z4

The origin and subsequent spread of plastids by endosymbiosis had a major environmental impact and altered the course of a great proportion of eukaryotic biodiversity. The ancestor of dinoflagellates contained a secondary plastid that was acquired in an ancient endosymbiotic event, where a eukaryotic cell engulfed a red alga. This is known as secondary endosymbiosis and has happened several times in eukaryotic evolution. Certain dinoflagellates, however, are unique in having replaced this secondary plastid in an additional (tertiary) round of endosymbiosis. Most plastid proteins are encoded in the nucleus of the host and are targeted to the organelle. When secondary or tertiary endosymbiosis takes place, it is thought that these genes move from nucleus to nucleus, so the plastid retains the same proteome. We have conducted large-scale expressed sequence tag (EST) surveys from *Karlodinium micrum*, a dinoflagellate with a tertiary haptophyte-derived plastid, and two haptophytes, *Isochrysis galbana* and *Pavlova lutheri*. We have identified all plastid-targeted proteins, analysed the phylogenetic origin of each protein, and compared their plastid-targeting transit peptides. Many plastid-targeted genes in the *Karlodinium* nucleus are indeed of haptophyte origin, but some genes were also retained from the original plastid (showing the two plastids likely co-existed in the same cell), in other cases multiple isoforms of different origins exist. We analysed plastid-targeting sequences and found the transit peptides in *K. micrum* are different from those found in either dinoflagellates or haptophytes, pointing to a plastid with an evolutionarily chimeric proteome, and a massive remodelling of protein trafficking during plastid replacement.

*Corresponding author

Introduction

Endosymbiosis has played a critical role at several turning points in eukaryotic evolution, not least in the origin and subsequent spread of plastids, or chloroplasts. These few events are responsible for significant changes to marine and terrestrial environments, the composition of the oceans and atmosphere, and the course of evolution of a large proportion of eukaryotic diversity. And yet these events took place so long ago, and the organelles are now so highly integrated, that many of the effects of this process at the genomic level are only poorly understood. Plastids ultimately most likely originated from a single primary endosymbiosis between a eukaryotic host and a cyanobacterium, giving rise to plants, green and red algae and glaucophytes.[1,2] Many endosymbiont genes were lost, but many others were transferred to the host nucleus. In the best-studied cases, the protein products of these genes are targeted to plastids where they function in their original environment. This targeting is mediated by an N-terminal extension called a transit peptide (TP), which is recognized by protein complexes in the two plastid membranes.[3,4] It has also been suggested that many genes moved from the endosymbiont to the nucleus and their products took on roles in the cytoplasm, but only a few such cases have been examined carefully, so the impact of such a contribution to the nuclear genome is not certain.[5]

Present address: R. F. Waller, School of Botany, University of Melbourne, Melbourne, Australia.

Abbreviations used: TP, transit peptide; SP, signal peptide; EST, expressed sequence tag.

E-mail address of the corresponding author: pkeeling@interchange.ubc.ca

On several occasions these primary algae were themselves taken up by another eukaryote and transformed into an organelle, a process called secondary endosymbiosis. Most or all of the genes for plastid-targeted proteins in the primary algal nucleus moved once again, this time to the new secondary host nucleus and, except in cryptophytes and chlorarachniophytes, the nucleus of the primary host was lost. Secondary plastids are situated within the host endomembrane system, so plastid-targeting leaders consist of a signal peptide (SP) followed by a TP. The SP directs the proteins to the endomembrane system *via* the signal recognition particle, and from there the TP directs them to the plastid.[6]

One algal group that contains a secondary plastid derived from the red algal plastid lineage is the dinoflagellates. Dinoflagellates are alveolates and sisters to apicomplexan parasites (e.g. the malaria parasite, *Plasmodium*). Their secondary plastid, called the peridinin plastid, is distinguished by the pigment peridinin and by having three bounding membranes rather than the four common to most secondary plastids. This difference is associated with specific changes to the way proteins are targeted to the plastid.[7] Dinoflagellates are also unique among eukaryotic algae in that they have taken the process of endosymbiosis one step further: in several independent lineages, the peridinin plastid has been replaced, either with a successive secondary plastid[8] or with a plastid from another secondary alga, resulting in tertiary plastids.[9–13] These organisms are not only fascinatingly complex cells, but they also have characteristics that make them useful models to study the effects on the genome of endosymbiosis: the endosymbiotic events took place recently compared to primary or secondary endosymbioses and, more importantly, the extant relatives of the host and endosymbiont are well defined and genomic information is available. Taken together, these characteristics allow us to track the fate of genes from various genomes during endosymbiosis and genomic integration with much more confidence than is possible with older events where the participants are not as well defined. These are important considerations, because all three processes, primary, secondary, and tertiary endosymbiosis, involved massive levels of gene transfer and major renovations to protein trafficking systems that are difficult to untangle without a clear idea of the ancestral conditions of the participant cells.

Karlodinium and *Karenia* are two related genera of toxic bloom forming dinoflagellates that have tertiary plastids of haptophyte origin.[13,14] In such organisms, it is thought that the nuclear genes for proteins targeted to the original peridinin plastid would be lost and replaced by homologues from their new endosymbiont, a process dubbed genome transformation.[15] Indeed, the first investigations of the genome of *Karenia brevis* confirmed that several such gene replacements took place: plastid-targeted isoforms of GAPDH, PsbO and Rubisco were all found to be haptophyte-like,[15–17] but a lack of data from haptophytes prevented any genome-wide assessment.[15] To investigate the impact of endosymbiosis on the host genome, plastid proteome, and the plastid-trafficking system in greater depth, we have carried out expressed sequence tag (EST) surveys from *Karlodinium micrum*, and two haptophytes, *Isochrysis galbana* and *Pavlova lutheri*. We conducted phylogenetic analyses on all genes for plastid-targeted proteins in all three surveys, and found that some *K. micrum* plastid-targeted proteins are from the haptophtye endosymbiont, some are ancestrally dinoflagellate, and some have multiple isoforms. We also analysed *K. micrum* TP characteristics and found that they resemble those of neither haptophytes nor dinoflagellate. Overall, some features of the *K. micrum* plastid are replacements from its tertiary endosymbiont, others have been uniquely and extensively remodelled from the secondary plastid, which suggests that the two plastids co-existed simultaneously.

Results

Identification of plastid-targeted proteins and targeting sequences

We 5′ end-sequenced 16,544 *K. micrum* ESTs, which assembled into 11,903 unique clusters. Of these, 38 clusters were identified as encoding 17 distinct plastid proteins (21 were isoforms, 14 of which were light harvesting proteins) by similarity to known plastid proteins, or known plastid isoforms of proteins. We also sequenced the 5′ end of cDNA sequences from two haptophyte species, *I. galbana* and *P. lutheri*, generating 12,216 and 7588 ESTs, respectively. We assembled 6106 unique clusters from *I. galbana* and identified 30 clusters encoding plastid-proteins. Of these, 15 encoded light-harvesting protein isoforms and five encoded ribosomal-protein subunits. From *P. lutheri* we assembled 3383 unique clusters from which we identified 20 plastid-targeted proteins, of which eight encoded ribosomal-protein subunits. All unique clones for plastid protein genes were fully sequenced.

Of the 38 plastid proteins identified from *K. micrum* 26 included full-length leaders. From the haptophytes, 23 *I. galbana* and 12 *P. lutheri* clusters encoded full-length proteins, in addition five that were found to contain complete TPs and part of the signal peptide. The accession numbers and completeness of all sequences used in these analyses are given in Supplementary Data, Table 1.

Genes originating from the tertiary endosymbiont are present in the nucleus

Of the 17 distinct plastid proteins identified in *K. micrum*, homologues from both peridinin-containing dinoflagellates and haptophytes were identified in 14 cases. None of the 14 genes are

plastid-encoded in either haptophytes or dino-flagellates,[18] so all are most likely derived either from pre-existing dinoflagellate nucleus-encoded genes, or nucleus–nucleus transfers from the haptophyte endosymbiont. For each protein, the origin of the *K. micrum* gene was inferred using phylogenetic methods. Since a specific question about a relatively recent endosymbiotic event was being asked, the overall robustness of the phylogeny is not as important as whether the positions of the relevant taxa are well supported; whether there is robust pairing of *K. micrum* with either hapto-phytes or dinoflagellates, to the exclusion of the other. In six phylogenies the *K. micrum* homologue was related to haptophytes (Figure 1). In three cases, acyl carrier protein (ACP), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), and csp41, *K. micrum* formed a strongly supported clade with haptophytes to the exclusion of dinoflagellates (88–100% support), and modest support was also seen in the ferredoxin:NADPH oxidoreductase phylogeny (65–77% support). PsaE also supported this relationship, but in this case the branch uniting *K. micrum* and haptophytes did not receive boot-strap support, although the distantly related dinoflagellate clade was strongly supported (100%) to the exclusion of the *K. micrum* sequence. Class II fructose bisphosphate aldolase (FBA2) is an interesting example, because the plastid isoform placed *K. micrum* and *K. brevis* within a haptophyte/heterokont/cryptomonad clade to the exclusion of the dinoflagellate with weak support, while the cytosolic isoform branched strongly with the dinoflagellate to the exclusion of haptophtyes, exactly as one would expect from a tertiary endosymbiotic plastid replacement. In general, phylogenies such as these are the expected result for *K. micrum*: they show that a broad range of nucleus-to-nucleus gene transfers took place when the haptophyte integrated with its dinoflagellate host.

Genes originating from the secondary endo-symbiont are retained and used by the tertiary endosymbiont
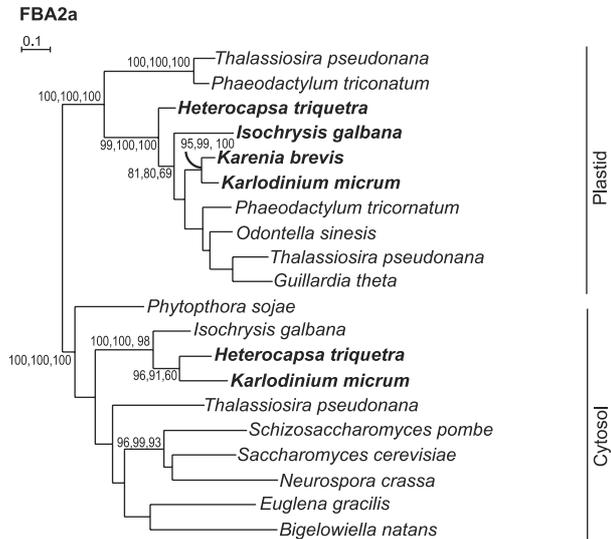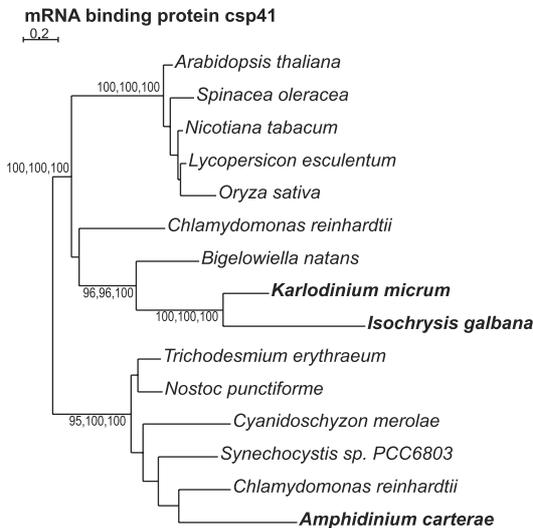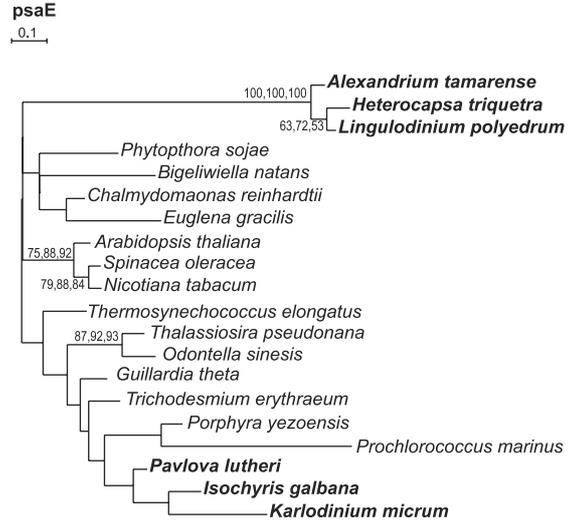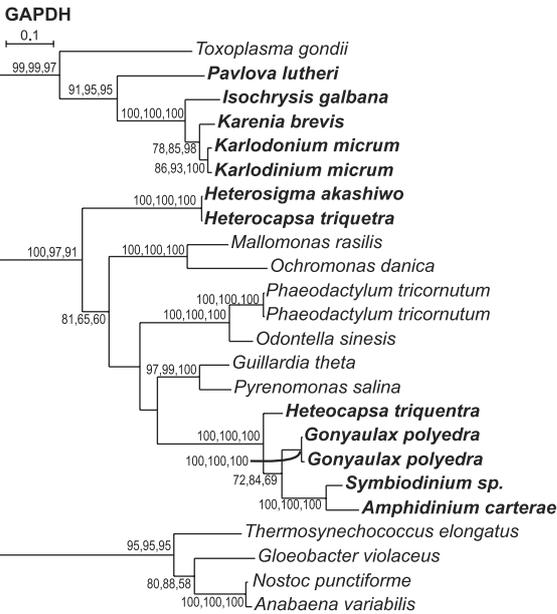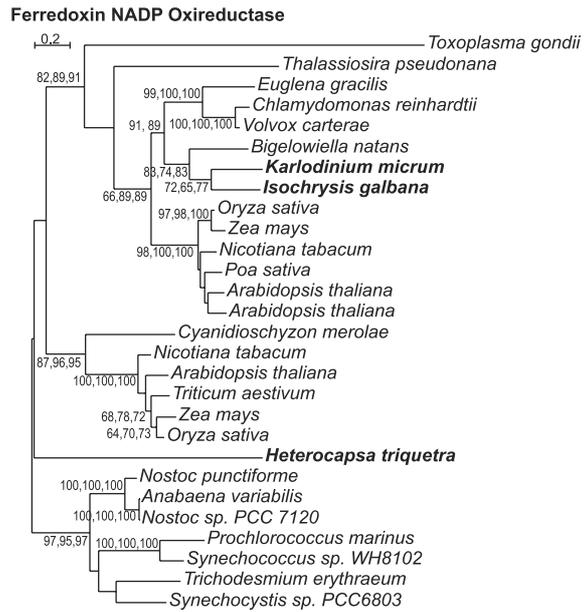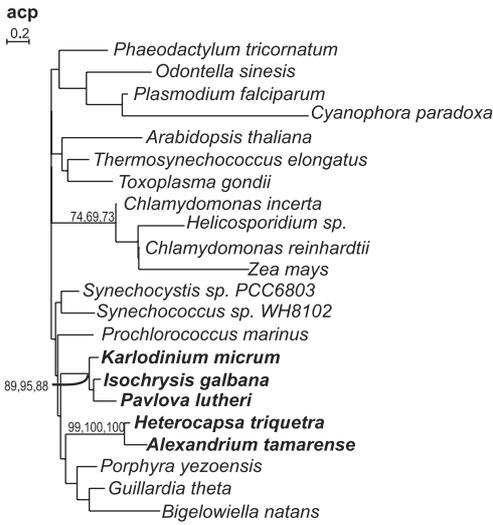
The presence of haptophyte-derived genes in the *K. micrum* nucleus should not come as a surprise because these proteins are simply maintaining their functional compartmentalization. However, phylo-genies inferred for three proteins suggest that *K. micrum* has also retained genes derived from the ancestral secondary plastid, and that their products are now targeted to the tertiary plastid. In two cases, thylakoid-bound ascorbate peroxidase and phosphoribulokinase, *K. micrum* branches with the dinoflagellate *H. triquetra* to the exclusion of the haptophyte *I. galbana* with, or near, 100% support (Figure 2). In phosphoribulokinase phylogeny, *I. galbana* is the closest relative to this pair. A third phylogeny of ATP synthase gamma subunit also shows this relationship, but only with modest support (61–78%) and the *K. micrum* and *H. triquetra*

sequences are both similarly divergent. Approxi-mately unbiased tests on all three phylogenies were consistent with the levels of support from boot-strapping. Eight alternative positions in each of the thylakoid-bound ascorbate peroxidase and phos-phoribulokinase phylogenies were all rejected at the 1% level. These tests also examined the 100 maximum likelihood bootstrap trees, and inspect-ing these showed 14 and 19 trees were accepted at the 5% level in thylakoid-bound acsorbate peroxi-dase and phosphoribulokinase, respectively, and all of these trees included the *K. micrum*/*H. triquetra* grouping (Figure 2). In ATP synthase gamma, nine alternate positions were tested and all were rejected at the 1% level, except those placing *K. micrum* as sister to *I. galbana* or *I. galbana* and *H. triquetra*, which were not rejected at the 5% level. Overall, the *K. micrum* thylakoid-bound ascorbate peroxi-dase and phosphoribulokinase are clearly derived from a peridinin plastid, whereas ATP synthase gamma also may be, but the data are equivocal and long branch attraction cannot be ruled out in this case. The relationship between *K. micrum* and dinoflagellates for this gene may be real or may be an artifact due to their shared divergence.

Three phylogenies, ferredoxin, glutamate semi-aldehyde aminomutase, and ribulose phosphate isomerase did not resolve the position of *K. micrum* with any support. Interestingly, however, the maximum likelihood tree of all three showed it branching with the dinoflagellates rather than the haptophytes (Supplementary Data, Figure 1) but, as this is not supported, these proteins are not discussed further.

Isoforms of mutiple origins

In addition to genes for plastid-targeted proteins of either haptophyte or dinoflagellate origin, in two instances we found mutiple isoforms in *K. micrum* with different evolutionary origins (Figure 3). Perhaps the most interesting of these genes is the *O*-acetyl serine (thiol) lyase (OASTL) component of cysteine synthase. Cysteine biosynthesis in bacteria and plastids takes place *via* acetylserine, whereas in the eukaryotic cytosol it proceeds *via* cystathionine using distinct or distantly related enzymes.[19] Homologues of cystathionine-beta-synthase were found in the ESTs of *H. triquetra*, *K. brevis* and *A. carterae*, suggesting they have the cytosolic path-way. OASTL is a key enzyme in the bacterial/plastid pathway, and this was also found in *A. tamarense* and *K. micrum*, suggesting that dino-flagellates may have both pathways. *K. micrum* has two isoforms of OASTL, but in this case they both fall into a single clade containing haptophytes, dinoflagellates, and a diatom. Interestingly one isoform branches very strongly (99–100% support) with the dinoflagellate, *A. tameranse*, suggesting this plastid gene has been retained from the peridinin plastid. The other isoform is not closely related to haptophytes or dinoflagellates, but has a very interesting structure. The C terminus of this protein

acp

0.2

- *Phaeodactylum tricornatum*
- *Odontella sinesis*
- *Plasmodium falciparum*
- *Cyanophora paradoxa*
- *Arabidopsis thaliana*
- *Thermosynechococcus elongatus*
- *Toxoplasma gondii*
- *Chlamydomonas incerta*
74,69,73 *Helicosporidium sp.*
- *Chlamydomonas reinhardtii*
- *Zea mays*
- *Synechocystis sp. PCC6803*
- *Synechococcus sp. WH8102*
- *Prochlorococcus marinus*
- **Karlodinium micrum**
89,95,88 **Isochrysis galbana**
- **Pavlova lutheri**
99,100,100 **Heterocapsa triquetra**
- **Alexandrium tamarense**
- *Porphyra yezoensis*
- *Guillardia theta*
- *Bigelowiella natans*

Ferredoxin NADP Oxireductase

0.2

- *Toxoplasma gondii*
82,89,91 *Thalassiosira pseudonana*
99,100,100 *Euglena gracilis*
91,89 *Chlamydomonas reinhardtii*
100,100,100 *Volvox carterae*
66,89,89 83,74,83 *Bigelowiella natans*
Karlodinium micrum
72,65,77 **Isochrysis galbana**
97,98,100 *Oryza sativa*
Zea mays
98,100,100 *Nicotiana tabacum*
Poa sativa
Arabidopsis thaliana
Arabidopsis thaliana
- *Cyanidioschyzon merolae*
87,96,95 *Nicotiana tabacum*
100,100,100 *Arabidopsis thaliana*
Triticum aestivum
68,78,72 *Zea mays*
64,70,73 *Oryza sativa*
Heterocapsa triquetra
100,100,100 *Nostoc punctiforme*
Anabaena variabilis
100,100,100 *Nostoc sp. PCC 7120*
97,95,97 *Prochlorococcus marinus*
100,100,100 *Synechococcus sp. WH8102*
Trichodesmium erythraeum
Synechocystis sp. PCC6803

GAPDH

0.1

- *Toxoplasma gondii*
99,99,97 **Pavlova lutheri**
91,95,95 **Isochrysis galbana**
100,100,100 **Karenia brevis**
78,85,98 **Karlodonium micrum**
86,93,100 **Karlodinium micrum**
100,100,100 **Heterosigma akashiwo**
Heterocapsa triquetra
100,97,91 100,100,100 *Mallomonas rasilis*
Ochromonas danica
Phaeodactylum tricornutum
100,100,100 *Phaeodactylum tricornutum*
100,100,100 *Odontella sinesis*
81,65,60 97,99,100 *Guillardia theta*
Pyrenomonas salina
Heteocapsa triquentra
100,100,100 **Gonyaulax polyedra**
100,100,100 **Gonyaulax polyedra**
72,84,69 **Symbiodinium sp.**
100,100,100 **Amphidinium carterae**
Thermosynechococcus elongatus
95,95,95 *Gloeobacter violaceus*
80,88,58 *Nostoc punctiforme*
100,100,100 *Anabaena variabilis*

psaE

0.1

- 100,100,100 **Alexandrium tamarense**
Heterocapsa triquetra
63,72,53 **Lingulodinium polyedrum**
- *Phytopthora sojae*
- *Bigeliwiella natans*
- *Chalmydomaonas reinhardtii*
- *Euglena gracilis*
75,88,92 *Arabidopsis thaliana*
Spinacea oleracea
79,88,84 *Nicotiana tabacum*
- *Thermosynechococcus elongatus*
87,92,93 *Thalassiosira pseudonana*
Odontella sinesis
- *Guillardia theta*
- *Trichodesmium erythraeum*
- *Porphyra yezoensis*
- *Prochlorococcus marinus*
Pavlova lutheri
Isochyris galbana
Karlodinium micrum

mRNA binding protein csp41

0.2

- *Arabidopsis thaliana*
100,100,100 *Spinacea oleracea*
- *Nicotiana tabacum*
- *Lycopersicon esculentum*
- *Oryza sativa*
- *Chlamydomonas reinhardtii*
- *Bigelowiella natans*
100,100,100 96,96,100 **Karlodinium micrum**
100,100,100 **Isochrysis galbana**
- *Trichodesmium erythraeum*
- *Nostoc punctiforme*
95,100,100 *Cyanidoschyzon merolae*
- *Synechocystis sp. PCC6803*
- *Chlamydomonas reinhardtii*
- **Amphidinium carterae**

FBA2a

0.1

100,100,100 *Thalassiosira pseudonana*
Phaeodactylum triconatum
100,100,100 **Heterocapsa triquetra**
99,100,100 **Isochrysis galbana**
95,99, 100 **Karenia brevis**
81,80,69 **Karlodinium micrum**
Phaeodactylum tricornatum
Odontella sinesis
Thalassiosira pseudonana
Guillardia theta

Plastid

- *Phytopthora sojae*
100,100, 98 *Isochrysis galbana*
Heterocapsa triquetra
96,91,60 **Karlodinium micrum**
100,100,100 *Thalassiosira pseudonana*
Schizosaccharomyces pombe
96,99,93 *Saccharomyces cerevisiae*
Neurospora crassa
Euglena gracilis
Bigelowiella natans

Cytosol

is fused to glutaredoxin, suggesting that this protein may be controlled by redox in a novel fashion.

Three distinct psbU homologues were found in *K. micrum*, two of which were found to branch with other haptophytes as expected, but the third formed a strongly supported grouping with the diatom *Thalassiosira pseudonana*, which in turn branched weakly with dinoflagellates. There has clearly been some paralogy in the history of this protein and so not much can be concluded without further sampling. The close relationship between *K. micrum* and a diatom may be a sampling error, or may be a result of lateral gene transfer, which has been observed between diatoms and dinoflagellate enolases.[20]

Transit peptides of *K. micrum* are unique

Phylogenetically, the plastid protein complement of *K. micrum* is a composite of proteins from the tertiary haptophyte plastid and some residual proteins from the original dinoflagellate peridinin-containing plastid, and perhaps a few proteins from other sources altogether. All these proteins now appear to function in the same compartment, and therefore one would expect them to use a single plastid-targeting system, but would this resemble a dinoflagellate or haptophyte system? In order to infer this, the characteristics of targeting need to be known from both dinoflagellates and haptophytes, but only a few targeted proteins have been previously available from haptophytes. Accordingly, we analysed the chemical characteristics of 51 haptophyte pre-sequences from our EST data and a small number of published sequences. Overall, these pre-sequences have features typical of other secondary plastid-targeting sequences, namely, they consist of a signal peptide followed by a transit peptide. Signal peptides were typical in composition with the majority of residues being hydrophobic and a signal peptidase cleavage site detectable using Signal-P.[21] The TPs were also representative of secondary TPs with a reduction in acidic residues (3.1% compared to 12% in the mature protein) and an increase in both alanine (17.5% compared to 10.4%) and arginine residues (10.4% compared to 5%). Phenylalanine is the most common residue at the first position of the haptophyte TPs. The motif described as FVAP,[7] consisting of a phenylalanine residue followed by three bases, none of which are either acidic or basic and no more than one of which is polar-uncharged, is a conspicuous feature of the plot generated using Sequence Logo Generation,[22] where the frequency of residues at each of the first 12 positions of the 51 haptophyte TPs are displayed (Figure 4(a)).

However, such a FVAP motif occurs in only 11 of the 51 haptophyte TPs in total. This is lower than that of dinoflagellate transit peptides analysed concurrently, in which case 41 of the 51 transit peptides (80%) contained a FVAP-like motif. In haptophytes the presence of a phenylalanine at or near the first position of the TP is also lower (31% of haptophyte sequences) than has been described in diatoms, or red algae, where 80% and 91% of transit peptides began with phenylalanine in *Thalassiosira pseudonana*[23] and *C. merolae*,[7] respectively.

By comparison, the TPs of peridinin-containing dinoflagellates are found to contain novel features compared to other secondary plastids.[7] Despite these TPs having characteristics generally associated with all TPs, including the FVAP motif that is widespread among non-green plastids,[7] dinoflagellate TPs are unusual occurring in multiple classes. Class I TPs contain an additional hydrophobic region, functioning as a transmembrane helix, that is involved in transport *via* Golgi-vesicles analogous to plastid targeting in Euglenids[24,25] and thought to be related to plastids having three membranes. Class II dinoflagellate TPs do not contain a transmembrane domain, and the method by which these proteins are sorted remains unknown. In addition, peridinin-containing dinoflagellate TPs show a unique bias for glutamine residues.[7]

The 26 full-length plastid-targeting proteins identified in *K. micrum* were all found to encode a typical signal peptide followed by a peptide extension consistent with a transit peptide. However, when the composition of these TP sequences was compared to 51 peridinin-containing dinoflagellate TPs and 51 haptophyte TPs, the *K. micrum* TPs were found to share certain features in common with all TPs, some features unique to dinoflagellates, and some features never observed in any TP. The percentage of serine residues (10.6% compared to 5.7% in the mature protein) and arginine residues (8.4% compared to 3.6% in the mature protein) are increased, as is common in most described transit peptides. Also the percentage of glutamine residues is elevated (6.8% compared to 2.6% in the mature protein) in common with dinoflagellates.[7] However, two features common to all other TPs known are absent in *K. micrum*. Firstly, the percentage of acid residues is not decreased; 14.3% of residues in transit peptides compared to 12.9% of residues in mature proteins, and given no elevation of basic residues, *K. micrum* transit peptides lack the net positive charge otherwise common to TPs (Figure 5). Secondly, there was no enrichment of alanine residues. In addition only one of the 26 TPs contained a FVAP-like motif and the overall

Figure 1. Maximum likelihood trees inferring a haptophyte origin in *K. micrum* for RNA-binding protein csp41, GAPDH, ferredoxin-dependent NADP oxireductase, FBA-2a, acyl carrier protein and psaE. Trees were inferred using ProML. Numbers at nodes indicate bootstrap values over 60 from (left to right: ProML, phyML and Weighbor). PD, peridinin dinoflagellate; H, haptophyte.
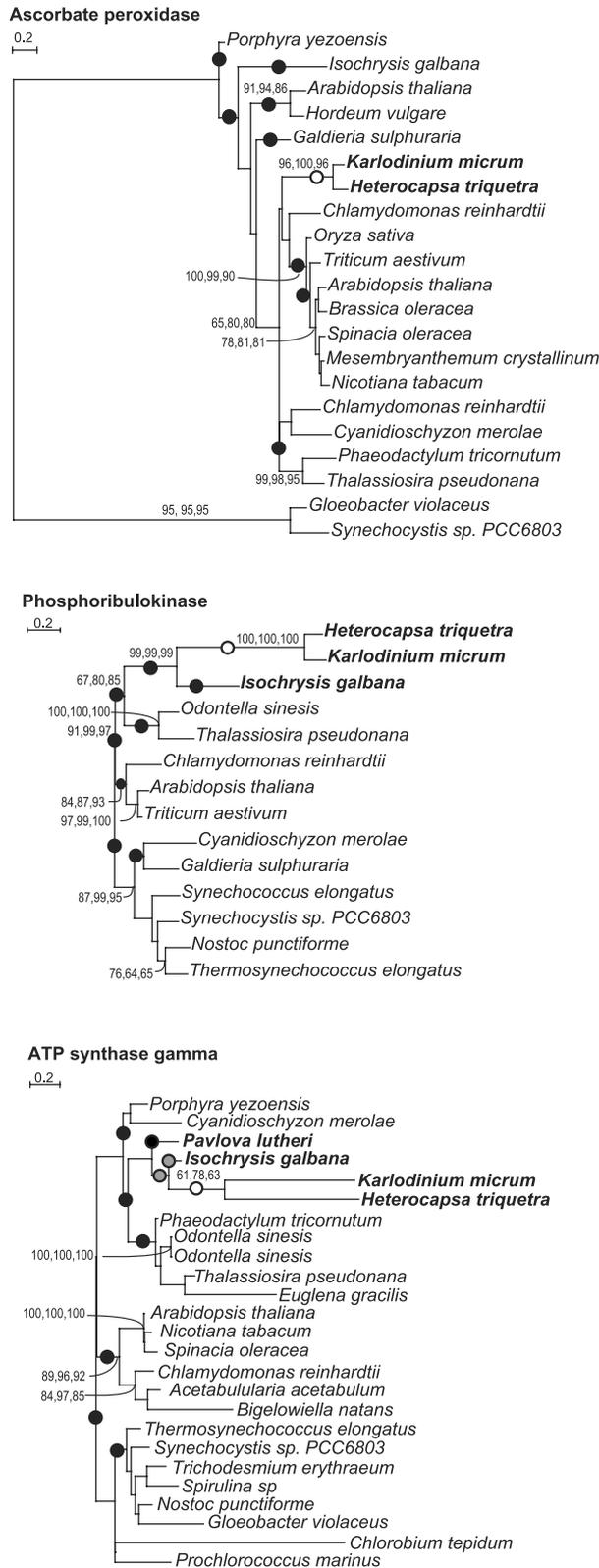
Ascorbate peroxidase


Phosphoribulokinase

ATP synthase gamma

Figure 2. Maximum likelihood trees inferring a peridinin dinoflagellate origin in *K. micrum* for thylakoid-bound ascorbate peroxidase, phosphoribulokinase and ATP-synthase gamma. Trees were inferred using ProML. Numbers at nodes indicate bootstrap values over 60 from (left to right: ProML, phyML and Weighbor). PD, peridinin dinoflagellate; H, haptophyte (AU test nodes).

Cysteine synthase
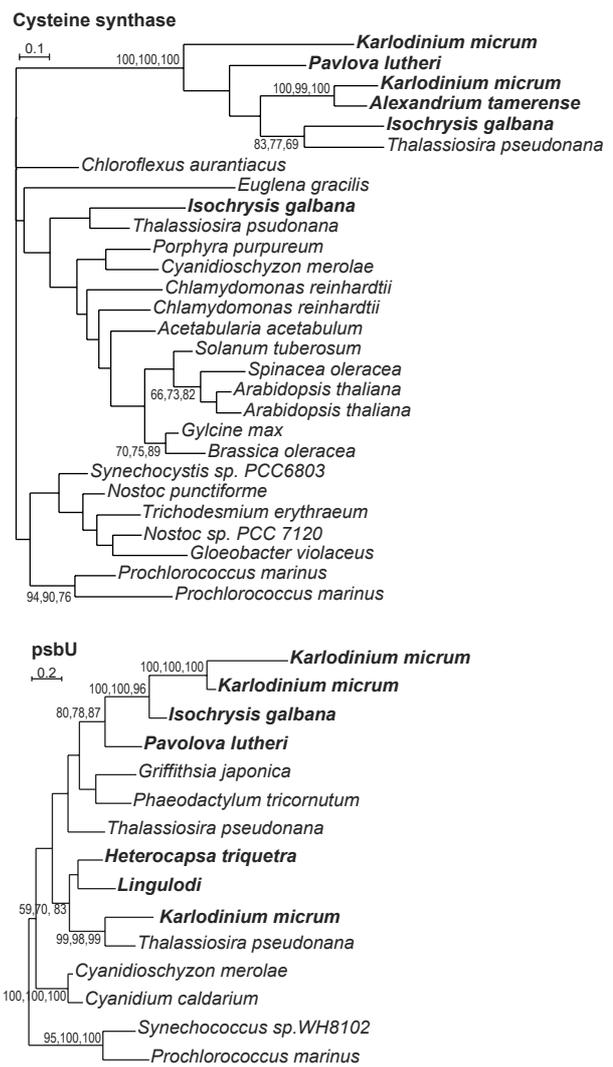
psbU

Figure 3. Maximum likelihood trees inferred for proteins (cysteine synthase, and psbU) with isoforms of different origins in *K. micrum*. Trees were inferred using ProML. Numbers at nodes indicate bootstrap values over 60 from (left to right: ProML, phyML and Weighbor). PD, peridinin dinoflagellate; H, haptophyte.

composition of the first 12 residues after the signal-peptide cleavage site is very different from either haptophytes or dinoflagellates, as shown in plots of amino acid composition (Figure 4).

While these compositional peculiarities were seen in all *K. micrum* transit peptides, two transit peptides (GAPDH and ACP) contain a hydrophobic region predicted to be a transmembrane helix.[26] This is not found in haptophyte transit peptides, but is common to class I dinoflagellate transit peptides.[7] In other dinoflagellates, GAPDH has a class I transit peptide, but ACP has a class II transit peptide, which lacks a transmembrane domain. The phylogenies inferred for both of these proteins suggest that the *K. micrum* genes originated in the haptophyte endosymbiont, and so these leaders have been acquired or refashioned after replacement of the secondary plastid.
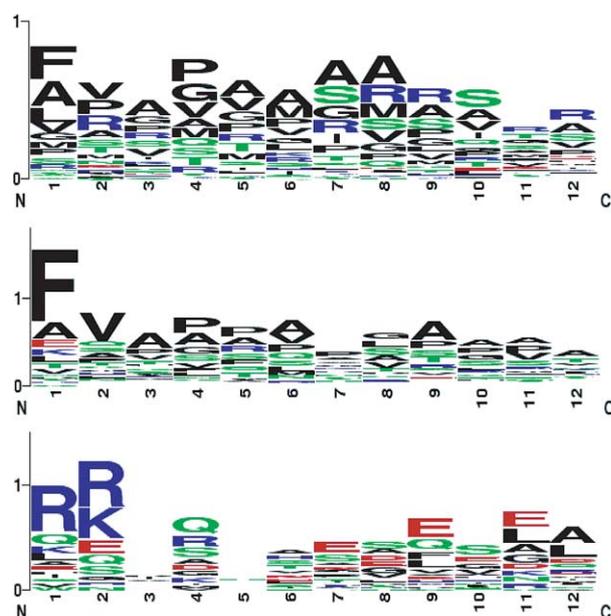
Figure 4. Sequence Logo Generation plots of the first 12 residues of the transit peptides from (a) 51 proteins from haptophytes, (b) 51 proteins from peridinin-containing dinoflagellates and (c) 26 proteins from *K. micrum*. The height of an amino acid letter is proportional to its frequency at a given position; the height of a stack of letters is proportional to information content at a given position. Red letters indicate acidic residues, blue letters indicate basic residues, green letters indicate polar-uncharged residues and black letters indicate hydrophobic non-polar residues.

Figure 5. Amino acid composition of targeting peptides. Percentage bar charts of the first 25 residues of the transit peptides of 51 haptophyte, 51 peridinin-containing dinoflagellate, 26 *K. micrum* proteins and the entire mature regions of 26 *K. micrum* proteins. Red, polar acidic residues (D, E); blue, polar basic residues (H, K, R); green, polar uncharged residues (C, N, Q, S, T,W, Y); yellow, hydrophobic non-polar residues (A, F, G, I, L, M, P, V).

Discussion

The integration of the tertiary plastids in *Karlodinium* and *Karenia* offers several unique advantages to studying the process of endosymbiosis: the event was relatively recent, both host and endosymbiont are well characterized phylogenetically, and the host is unusual in that it, at least ancestrally, already had a plastid so genes from the original plastid could play a role in the process of integration. Previous analyses suggested this was not the case, since the only (three) plastid-targeted genes obtained from *K. micrum* and *K. brevis* are all of haptophyte origin, and several photosystem genes usually found in the nucleus of dinoflagellates are probably encoded in the plastid genome of *K. brevis*.[13,15–17] Those data confirmed that the plastid originated by tertiary endosymbiosis with a haptophyte and were taken to suggest that all plastid proteins originated with the new tertiary plastid.[15] Indeed, this is expected because the process of moving a gene from genome to genome within a cell is probably relatively trivial, whereas expecting a protein from one plastid to function in the heterologous environment of another plastid is likely more difficult.

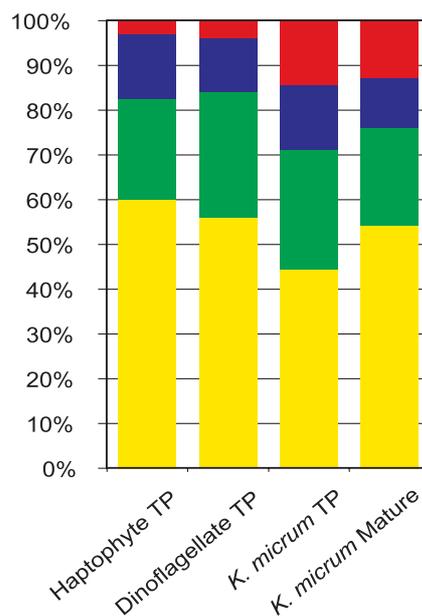Here, however, we have shown that the expected model is not true in *K. micrum*, which has retained

genes from both the original secondary plastid, as well as from the new tertiary plastid, in the nucleus. The phylogenies of two proteins (ascorbate peroxidase and phosphoribulokinase) show strong support for a dinoflagellate origin and four more show this with weaker support (ATP synthase gamma, ferredoxin, glutamate semialdehyde aminomutase, and ribulose phosphate isomerase). It is possible that the secondary plastid persists undetected in a relic form, but given the unique transit peptide characteristics seen in both sets of plastid proteins it is more likely that only one plastid type is present and uses proteins from various sources. Nevertheless, the presence of these proteins, although few, is highly significant as it implies that the dinoflagellate and the haptophyte plastids probably co-existed for some time, otherwise the dinoflagellate genes for plastid proteins would have likely been corrupted or lost before having the chance to function in the new endosymbiont. Interestingly, so far none of the dinoflagellate genes identified are involved in photosynthesis, raising the possibility that the dinoflagellate that engulfed the haptophyte was non-photosynthetic at the time, utilizing its plastid for anabolic pathways, such as fatty acid and isprenoid synthesis, as occurs for many current non-photosynthetic dinoflagellates and their apicomplexan relatives.[27] If this were the case it would also provide an explanation for why a plastid was re-acquired, as it would have been the only way to revert to autotrophism. If the host still utilized its

original plastid for anabolic pathways, and the haptophyte-derived one principally for photosynthesis, the maintenance of two plastids would have been selectively advantageous for a period, until the equivalent anabolic pathways in the haptophyte symbiont could be exploited. So why were some of the original dinoflagellate genes retained and put to use in the tertiary plastid if the proteins that evolved under selection in this compartment likely have a functional advantage? Despite the dinoflagellate genes having evolved for a different plastid environment, they would nonetheless already be located, and perhaps appropriately regulated, in the host nucleus. Any such proteins that could function in the tertiary endosymbiont would only need to adopt the necessary changes to the targeting TP for import into the new plastid. Thus this could represent a significant head-start in gene replacement.

It is possible that the differences that we now see between the plastid-targeting TPs for the tertiary plastid in *K. micrum* and either of the progenitor cells was also driven by the co-existence of two divergent plastids in one cell. In *K. micrum* the TPs lack of positive charge is very unusual for a transit peptide, and they also lost the ancient FVAP motif that both host and endosymbiont plastids would have used. With gene migration to the nucleus, discrimination of plastid genes would be essential to ensure that each plastid type imported the correct protein complements in sufficient quantities. This change is likely to have occurred at the level of TP recognition. It is also possible that modification of plastid membrane topology may have driven targeting changes. This tertiary plastid is apparently bounded by three membranes,[28] implying the loss of at least three endosymbiont membranes. It is difficult to model how such a re-organization event may have occurred, however it would certainly have impacted on protein targeting to the plastid. Whatever the case, the makeup of the *K. micrum* plastid proteome serves to show that the lines between host and symbiont can become muddied in endosymbiotic organellogenesis: in the case of *K. micrum* this is especially so since the tertiary plastid replaced a similar organelle and they appear to have co-existed in the cell, ideal conditions to create an organelle with a chimeric proteome.

Methods

Library construction and EST data generation

P. lutheri (CCMP 1325), *I. galbana* (CCMP 1323) and *K. micrum* (CCMP 415) were cultivated in Guillard's f2-Si medium at 16 °C with a 12 h light/12 h dark cycle. Cells were harvested during the light cycle and total RNA was purified using TRIzol (Invitrogen). *P. lutheri* and *K. micrum* RNA was used to construct cDNA libraries in λ-zap and mass-excised to pBluescript (Amplicon Express), and *I. galbana* RNA was used to construct a cDNA library in the pcDNA3.1 vector (DNA technologies). ESTs were sequenced from the 5′ end, and quality and vector-trimming, and cluster assembly was carried out at PEPdb†. Proteins were identified based on similarity searches to public databases. All EST sequences are available on PEPdbPUB†.

Identification of plastid proteins and phylogenetic analyses

Clusters in EST databases putatively identified as encoding proteins of plastid function were identified by browsing the database manually. Proteins representing known plastid functions were also used to directly search the three databases. All clusters thus identified were sequenced on both strands and protein alignments to homologues from other taxa were constructed using Clustal X[29] and refined in MacClade (Sinauer Associates, MA, USA). Alignments are available on request. Maximum likelihood phylogenies were inferred using ProML 3.6[30] with the James-Taylor Thornton probability model,[31] global rearrangements and randomized input orders and nine rates categories with rates and the proportions of invariable sites calculated by Tree-Puzzle version5.2.[32] Bootstrap resampling was carried out using PUZZLE-BOOT.[32] Bootstraps with ProML were obtained in the same way, but with a single rates category. Bootstraps were also obtained using PhyML with nine rates categories[33] and Weighbor.[34]

Proteins were predicted to be full length if they contained an N-terminal extension in comparison to homologues from cyanobacteria and if that extension was consistent with features of described targeting sequences, that is a signal peptide with a signal-peptidase cleavage site, as predicted by SignalP,[21] followed by a TP. TP analyses included only presequences with a predicted signal-peptidase cleavage site.

Tree-topology tests

Alternative tree topologies of were tested using the Approximately Unbiased (AU) test.[35] Hypothetical alternative topologies were constructed by moving *K. micrum* to be sisters to haptophyte and to other major branches on each tree. Test trees were "padded" with 100 maximum likelihood bootstrap trees and likelihood values for each character position were calculated for all topologies using PAML version 3.12.[36] Likelihoods were calculated using the WAG and gamma distributed rates, with alpha parameters estimated as described above. Likelihoods were used to calculate *p*-values in AU tests using CONSEL version 0.1d.[37]

Data bank accession codes

Fully sequenced clones have been deposited in GenBank (DQ11859–DQ11864 and DQ191742–DQ191744). Accession numbers are listed in full in Tables S1–3.

---

† http://amoebidia.bcm.umontreal.ca/public/pepdb/ agrm.php

Acknowledgements

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j. jmb.2006.01.084

References

1. Gray, M. W. & Spencer, D. F. (1996). Organellar evolution. In *Evolution of Microbial Life* (Roberts, D. M., Sharp, P., Alderson, G. & Collins, M. A., eds), vol. 54, pp. 109–126, Cambridge University Press, Cambridge, UK.
2. Rodriguez-Ezpeleta, N., Brinkmann, H., Burey, S. C., Roure, B., Burger, G., Loffelhardt, W. *et al.* (2005). Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr. Biol.* **15**, 1325–1330.
3. Bruce, B. D. (2001). The paradox of plastid transit peptides: conservation of function despite divergence in primary structure. *BBA-Mol. Cell Res.* **1541**, 2–21.
4. Jarvis, P. & Soll, J. (2002). Toc, Tic, and chloroplast protein import (vol 1541, pg 64, 2001). *BBA-Mol. Cell Res.* **1590**, 177–189.
5. Martin, W. & Herrmann, R. G. (1998). Gene transfer from organelles to the nucleus: How much, what happens, and why? *Plant Physiol.* **118**, 9–17.
6. van Dooren, G. G., Schwartzbach, S. D., Osafune, T. & McFadden, G. I. (2001). Translocation of proteins across the multiple membranes of complex plastids. *BBA-Mol. Cell Res.* **1541**, 34–53.
7. Patron, N. J., Waller, R. F., Archibald, J. M. & Keeling, P. J. (2005). Complex protein targeting to dinoflagellate plastids. *J. Mol. Biol.* **348**, 1015–1024.
8. Bhattacharya, D., Yoon, H. S. & Hackett, J. D. (2004). Photosynthetic eukaryotes unite: endosymbiosis connects the dots. *Bioessays*, **26**, 50–60.
9. Chesnick, J. M., Kooistra, W., Wellbrock, U. & Medlin, L. K. (1997). Ribosomal RNA analysis indicates a benthic pennate diatom ancestry for the endosymbionts of the dinoflagellates *Peridinium foliaceum* and *Peridinium balticum* (Pyrrhophyta). *J. Euk. Microbiol.* **44**, 314–320.
10. Watanabe, M. M., Suda, S., Inouye, I., Sawaguchi, T. & Chihara, M. (1990). Lepidodinium-Viride Gen Et Sp-Nov (Gymnodiniales, Dinophyta), a green dinoflagellate with a chlorophyll a-containing and B-containing endosymbiont. *J. Phycol.* **26**, 741–751.
11. Chesnick, J. M., Morden, C. W. & Schmieg, A. M. (1996). Identity of the endosymbiont of *Peridinium foliaceum* (Pyrrophyta): analysis of the rbcLS operon. *J. Phycol.* **32**, 850–857.
12. Schnepf, E. & Elbrachter, M. (1988). Cryptophycean-like double membrane-bound chloroplast in the dinoflagellate, *Dinophysis ehrenb*-evolutionary, phylogenetic and toxicological implications. *Botanica Acta*, **101**, 196–203.
13. Yoon, H. S., Hackett, J. D. & Bhattacharya, D. (2002). A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. *Proc. Natl Acad. Sci. USA*, **99**, 11724–11729.
14. Tengs, T., Dahlberg, O. J., Shalchian-Tabrizi, K., Klaveness, D., Rudi, K., Delwiche, C. F. & Jakobsen, K. S. (2000). Phylogenetic analyses indicate that the 19′ hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Mol. Biol. Evol.* **17**, 718–729.
15. Yoon, H. S., Hackett, J. D., Van Dolah, F. M., Nosenko, T., Lidie, L. & Bhattacharya, D. (2005). Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. *Mol. Biol. Evol.* **22**, 1299–1308.
16. Takishita, K., Ishida, K. I. & Maruyama, T. (2004). Phylogeny of nuclear-encoded plastid-targeted GAPDH gene supports separate origins for the peridinin- and the fucoxanthin derivative-containing plastids of dinoflagellates. *Protist*, **155**, 447–458.
17. Ishida, K. & Green, B. R. (2002). Second- and third-hand chloroplasts in dinoflagellates: phylogeny of oxygen-evolving enhancer 1 (PsbO) protein reveals replacement of a nuclear-encoded plastid gene by that of a haptophyte tertiary endosymbiont. *Proc. Natl Acad. Sci. USA*, **99**, 9294–9299.
18. Puerta, M. V., Bachvaroff, T. R. & Delwiche, C. F. (2005). The complete plastid genome sequence of the haptophyte *Emiliania huxleyi*: a comparison to other plastid genomes. *DNA Res.* **12**, 151–156.
19. Mendoza-Cozatl, D., Loza-Tavera, H., Hernandez-Navarro, A. & Moreno-Sanchez, R. (2005). Sulfur assimilation and glutathione metabolism under cadmium stress in yeast, protists and plants. *FEMS Microbiol. Rev.* **29**, 653–671.
20. Harper, J. T. & Keeling, P. J. (2004). Lateral gene transfer and the complex distribution of insertions in eukaryotic enolase. *Gene*, **340**, 227–235.
21. Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795.
22. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190.
23. Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H. *et al* (2004). The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science*, **306**, 79–86.
24. Sulli, C. & Schwartzbach, S. D. (1996). Precursor protein topology in Golgi vesicles during transport to Euglena chloroplasts. *Mol. Biol. Cell*, **7**, 457–463.
25. Sulli, C., Fang, Z. W., Muchhal, U. & Schwartzbach, S. D. (1999). Topology of Euglena chloroplast protein precursors within endoplasmic reticulum to Golgi to chloroplast transport vesicles. *J. Biol. Chem.* **274**, 457–463.
26. Hoffman, K. & Stoffel, W. (1993). TMbase—a database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler*, **374**, 166.

27. Ralph, S. A., van Dooren, G. G., Waller, R. F., Crawford, M. J., Fraunholz, M. J., Foth, B. J. *et al.* (2004). Tropical infectious diseases: metabolic maps and functions of the *Plasmodium falciparum* apicoplast. *Nature Rev. Microbiol.* **2**, 203–216.

28. Dodge, J. D. (1989). Phylogenetic relationships of dinoflagellates and their plastids. In *The Chromophyte Algae: Problems and Perspectives* (Green, J. C., Leadbeater, B. S. C. & Diver, W. L., eds), pp. 207–227, Clarenden Press, Oxford.

29. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.* **25**, 4876–4882.

30. Felsenstein, J. (1993). *Phylip—Phylogenetic Inferance Package*, Department of Genetics, University of Washington, Seattle, WA, USA.

31. Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comp. Appl. Biosci.* **8**, 275–282.

32. Schmidt, H. & Hasseler, A. v. (2003). Maximum-likelihood analysis using TREE_PUZZLE. In *Current Protocols in Bioinformatics* (Baxevanis, D., Davison, R. D. M., Page, G., Stromo, G. B. & Stein, L., eds), vol. unit 6.6, Wiley, New York.

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

34. Bruno, W. J., Socci, N. D. & Halpern, A. L. (2000). Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Mol. Biol. Evol.* **17**, 189–197.

35. Shimodaira, H. (2004). Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. *Ann. Stat.* **32**, 2616–2641.

36. Yang, Z. H. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comp. Appl. Biosci.* **13**, 555–556.

37. Shimodaira, H. & Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics*, **17**, 1246–1247.

Edited by J. Karn