International Journal of Systematic and Evolutionary Microbiology (2006), **56**, 1439–1447 DOI 10.1099/ijs.0.64061-0

Phylogenetic history of plastid-targeted proteins in the peridinin-containing dinoflagellate *Heterocapsa triquetra*

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

Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, 3529-6270 University Boulevard, Vancouver, BC, V6T 1Z4, Canada

Correspondence
Patrick J. Keeling
pkeeling@interchange.ubc.ca

The evolutionary history and relationship between plastids of dinoflagellate algae and apicomplexan parasites have been controversial both because the organelles are unusual and because their genomes contain few comparable genes. However, most plastid proteins are encoded in the host nucleus and targeted to the organelle, and several of these genes have proved to have interesting and informative evolutionary histories. We have used expressed sequence tag (EST) sequencing to generate gene sequence data from the nuclear genome of the dinoflagellate *Heterocapsa triquetra* and inferred phylogenies for the complete set of identified plastid-targeted proteins. Overall, dinoflagellate plastid proteins are most consistently related to homologues from the red algal plastid lineage (not green) and, in many of the most robust cases, they branch with other chromalveolate algae. In resolved phylogenies where apicomplexan data are available, dinoflagellates and apicomplexans are related. We also identified two cases of apparent lateral, or horizontal, gene transfer, one from the green plastid lineage and one from a bacterial lineage unrelated to plastids or cyanobacteria.

INTRODUCTION

Plastids originated by the endosymbiotic uptake of a cyanobacterium and the subsequent conversion of this endosymbiont into the highly reduced and specialized double-membrane-bound primary plastid found today in land plants and some algae. Most algal groups, however, acquired their plastids by an additional step called secondary endosymbiosis. Here, an alga containing a primary plastid is itself taken up and converted into an organelle within its new eukaryotic host (Archibald & Keeling, 2002). These secondary plastids are bounded by either three or four membranes and are found in chlorarachniophytes, euglenids, cryptomonads, heterokonts, haptophytes, dinoflagellates and apicomplexans. In all of these groups, the plastid retains only a small genome: most proteins are nuclear-encoded and are targeted post-translationally to the plastid using specific N-terminal peptides that are characteristic for either primary or secondary plastids (McFadden, 2001).

Dinoflagellates are closely related to apicomplexans and, together with ciliates and a handful of other protists, make up the alveolates (Fast *et al.*, 2002; Gajadhar *et al.*, 1991). Since many members of both dinoflagellates and apicomplexans contain secondary plastids, the most parsimonious explanation is that they share a common origin, but the evolutionary history of both plastids has proved contentious, in part because they are divergent, making them difficult to compare.

While some dinoflagellates have undergone plastid replacements through further endosymbiotic events (Delwiche, 1999; Keeling, 2004), the plastid found in the majority of photosynthetic dinoflagellates is a secondary plastid containing the distinctive pigment peridinin. This plastid is further distinguished in many ways, not least in that it has three membranes rather than four, which appears to have a significant affect on how proteins are targeted to the organelle (Nassoury *et al.*, 2003; Patron *et al.*, 2005). The genomes of dinoflagellate plastids are also the most reduced known. All but a few of their genes have moved to the nucleus (Bachvaroff *et al.*, 2004; Hackett *et al.*, 2004a), and the remaining genome (16 genes have been found so far, nearly all related to photosynthesis) has been broken up into mini-circles each generally encoding a single gene (Zhang *et al.*, 1999).

†Present address: Botany School, University of Melbourne, Parkville, VIC 3010, Australia.

Abbreviations: EST, expressed sequence tag; FBA, fructose-1,6-bisphosphate aldolase; GAPDH, glyceraldehyde-3-phosphate dehydrogenase.

The GenBank/EMBL/DDBJ accession numbers for the completed cDNA sequences reported in this study are AY826826–AY826947 and AY884246–AY884255. The dbEST accession numbers for the EST sequences reported in this study are DT379484–DT386290.

Protein maximum-likelihood trees for a further 23 plastid-targeted proteins are available as supplementary material in IJSEM Online.

Apicomplexans, on the other hand, are obligate parasites, so their plastid (the apicoplast) is non-photosynthetic and generally reduced. Its relict genome contains few genes, typically highly divergent at the sequence level, and mostly encoding housekeeping functions (Wilson *et al.*, 1996). The evolutionary origin of this plastid has been debated since its discovery, with some data interpreted as showing a green algal origin (Cai *et al.*, 2003; Funes *et al.*, 2002; Köhler *et al.*, 1997) and other data interpreted as showing a red algal origin (Blanchard & Hicks, 1999; Fast *et al.*, 2001; McFadden & Waller, 1997; Patron *et al.*, 2004; Waller *et al.*, 2003). Since nearly all of the genes remaining in the dinoflagellate plastid are related to photosynthesis, there are few plastid-encoded genes that can be compared directly between apicomplexans and dinoflagellates. Those that have been (primarily encoding small- and large-subunit rRNA) are highly divergent in both groups, making phylogenies difficult to interpret (Zhang *et al.*, 2000).

The disputes over the origin of dinoflagellate and apicomplexan plastids widened with the suggestion that both plastids originated in the ancestor of all chromalveolates. The chromalveolates are a hypothetical grouping of alveolates and chromists (cryptomonads, haptophytes and heterokonts): all eukaryotes hypothesized to have red algal-derived plastids (Cavalier-Smith, 1998). Plastid-encoded gene trees have given variable results, but multigene analyses weakly unite chromist plastids (Yoon *et al.*, 2002, 2004). Nuclear gene trees provide strong support for the alveolates and their relationship to heterokonts (Baldauf *et al.*, 2000; Harper *et al.*, 2005). Taken together, plastid and cytosolic data are therefore consistent with the chromalveolate hypothesis, but neither support the whole group: the strongest support for this comes from two nucleus-encoded genes for plastid-targeted proteins: glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and fructose-1,6-bisphosphate aldolase (FBA) (Fast *et al.*, 2001; Harper & Keeling, 2003; Patron *et al.*, 2004). In both cases, the chromalveolate plastid-targeted protein appears to have evolved in a unique way relative to other plastids, and these deviations suggest that plastids of chromalveolates share a common origin.

Phylogenies of GAPDH and FBA demonstrate the usefulness of nucleus-encoded plastid-targeted proteins for studying plastid evolution, but they are poorly sampled because the nuclear genome is practically less accessible than that of the plastid. In dinoflagellates, this problem is aggravated by their unusually large genome size, so expressed sequence tag (EST) surveys have been used to generate genomic data from several species of dinoflagellate (Bachvaroff *et al.*, 2004; Hackett *et al.*, 2004a; Patron *et al.*, 2005). Phylogenetic analyses for some of these proteins have been carried out, some showing a substantial conflict in gene trees indicating either a red or green algal origin, or high levels of lateral gene transfer (Hackett *et al.*, 2004a). Here, we have phylogenetically analysed all the plastid-targeted proteins identified from an EST survey of the dinoflagellate *Heterocapsa triquetra* (Patron *et al.*, 2005). We have inferred phylogenies for 52 proteins, eight of which have apicomplexan homologues. Overall, the dinoflagellate plastid proteins tend to branch with red algae and chromists (with red algal secondary plastids) with stronger consistency than previously observed. Of the phylogenies containing apicomplexan homologues that are resolved with reasonable support, the apicomplexans group within the red algal plastid clade and a specific relationship with the dinoflagellate homologues was evident in some of these. Many dinoflagellate plastid-targeted proteins are relatively divergent and contain unique oddities [such as a tandem fusion of translation elongation factor Ts (EF-Ts)], and a few well-resolved phylogenies appear to support lateral gene transfer, including genes derived from prokaryotes.

METHODS

EST sequencing and protein identification. *H. triquetra* CCMP 449, cultivated in Guillard's f2-Si medium at 16 °C with a 12 h : 12 h light/dark cycle, was harvested in batches and total RNA was used to construct a cDNA library as described by Patron *et al.* (2005). ESTs were 5′-sequenced and gene identification was performed at the PEPdb (http://amoebidia.bcm.umontreal.ca/pepdb/searches/welcome.php). Plastid-targeted proteins were identified as part of an analysis of the nature of plastid-targeting leader sequences in dinoflagellates reported in Patron *et al.* (2005). Briefly, EST annotation was searched for genes with known function in the plastid and the sequence database was searched using known plastid-targeted proteins from other organisms. In cases where candidate genes were determined to be full-length, they were analysed for the presence of an N-terminal leader with characteristics expected of a dinoflagellate plastid-targeting peptide (in particular the presence of a predicted signal peptide). In addition, phylogenetic analysis was carried out for all candidate genes (see below), revealing some to be related to cytosolic or mitochondrial homologues. In these cases, unless the gene encoded a leader warranting further investigation, they were no longer considered. In cases where cDNAs were truncated and the presence of a leader could not be verified, genes were considered to encode plastid-targeted proteins if they were phylogenetically related to other plastid-targeted homologues and to cyanobacterial homologues. Identification of putatively plastid-targeted proteins in apicomplexans followed previous annotation, which is based on the well-characterized leaders of annotated proteins (Ralph *et al.*, 2004) and direct localization (e.g. Jomaa *et al.*, 1999; Waller *et al.*, 1998). Completed cDNA sequences have been deposited in GenBank (accession numbers AY826826–AY826947, AY884246–AY884255) and EST sequences have been deposited in dbEST (DT379484–DT386290).

Phylogenetic analysis. Protein alignments were constructed using CLUSTAL X (Thompson *et al.*, 1997) and edited manually. All ambiguous sites of the alignments were removed from the dataset for phylogenetic analyses. The alignment data are available on request. Protein maximum-likelihood analyses used PhyML (Guindon & Gascuel, 2003) with input trees generated by BIONJ, the JTT model of amino acids substitution, proportion of variable rates estimated from the data and nine categories of substitution rates (eight variable and one invariable). One hundred bootstrap trees were calculated with PhyML initially without gamma correction categories; if the resulting trees showed resolution, the analysis was repeated with four rate categories. For distance analyses, gamma-corrected distances were calculated by TREE-PUZZLE 5.2 (Schmidt *et al.*, 2002) using the WAG substitution matrix with eight variable rate categories and invariable sites. Trees were inferred by weighted

neighbour-joining using WEIGHBOR 1.0.1a (Bruno *et al.*, 2000). Bootstrap resampling was performed using PUZZLEBOOT (shell script by A. Roger and M. Holder; http://www.tree-puzzle.de) with rates and frequencies estimated using TREE-PUZZLE 5.2.

RESULTS AND DISCUSSION

Red algal origin of *Heterocapsa* plastid-targeted proteins

A previous analysis of plastid-targeting leaders in *H. triquetra* (Patron *et al.*, 2005) identified a total of 63 distinct genes from 2022 EST clusters as likely being plastid-targeted protein-coding genes. Of these, 11 represented multiple copies of certain genes; hence there were 52 distinct plastid proteins identified in total (including several distinct lineages within the light-harvesting complex superfamily). The majority of these proteins are involved in the light or dark reactions of photosynthesis, but other activities such as transcription, translation or synthesis of fatty acids and isoprenoids were also represented, and eight proteins representing such functions had identifiable homologues in apicomplexans. These plastid proteins were each subjected to phylogenetic analyses, with the exceptions of form II ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO), which has already been analysed in detail and is known to be unique to dinoflagellates and photosynthetic proteobacteria (Whitney *et al.*, 1995), and the light-harvesting complex proteins, which we found to form a poorly resolved protein family in dinoflagellates. Furthermore, dinoflagellate GAPDH, FBA and PsbO have been analysed previously (Fast *et al.*, 2001; Harper & Keeling, 2003; Ishida & Green, 2002; Patron *et al.*, 2004).

Of the 34 discrete proteins analysed (Table 1), many of the *H. triquetra* genes were divergent compared with other plastid homologues, and the position of *H. triquetra* was completely unresolved in two cases. The other 32 proteins supported *H. triquetra* branching with other plastid homologues, as expected. Of these, *H. triquetra* branched with neither the red nor green algal lineages (i.e. the position was unresolved within plastids) or, in 18 cases, plastids were polyphyletic. *H. triquetra* branched with the red plastid lineage (red and chromistan algae) in another 13 cases, eight with moderate to strong support, as shown in three examples in Fig. 1. In the phosphoglycerate kinase phylogeny (Fig. 1a), the *H. triquetra* genes for the plastid and cytosolic enzymes are both shown, and the plastid gene branches within a relatively well supported (89 %) clade of red algal and chromists genes, as well as *Bigelowiella natans*, which has a green algal secondary plastid and whose phosphoglycerate kinase has been proposed to be derived from a red alga by lateral gene transfer (Archibald *et al.*, 2003). Similarly, *H. triquetra* phosphoribulokinase (Fig. 1b) branches specifically with chromist homologues, and most closely with the haptophyte *Isochrysis* (94–96 %). Lastly, *H. triquetra* and *Alexandrium tamarense* (Hackett *et al.*, 2004a) possess a gene for photosystem II extrinsic protein (Fig. 1c), a protein apparently lost from green algae altogether and otherwise only known in cyanobacteria, red algal plastids and their chromist derivatives.

As a whole, the phylogenies of *H. triquetra* plastid-targeted proteins are most consistent with the peridinin-containing plastid being derived from a red algal plastid. This is in line with results from plastid-encoded genes (Zhang *et al.*, 2000) and a few plastid-targeted genes (Bachvaroff *et al.*, 2004; Hackett *et al.*, 2004a). However, in some analyses, several proteins have shown a green algal origin (Hackett *et al.*, 2004a) and, in one multigene analysis, the red origin was not significantly better supported than a green algal origin (Yoon *et al.*, 2005). We see no strong evidence for a green algal origin and only a few cases that might suggest lateral gene transfer (see below). In nearly all resolved cases, *H. triquetra* branches specifically with heterokonts, haptophytes or cryptomonads, while only three proteins show *H. triquetra* branching with a red alga to the exclusion of these taxa; none of these are statistically supported. Among the more strongly supported phylogenies, the data are more consistent with the chromalveolate hypothesis than with independent plastid origins; however, broader representation of red algal taxa is required in order to test this hypothesis more thoroughly.

Relationship between dinoflagellate and apicomplexan plastid-targeted proteins

Most of the *H. triquetra* plastid-targeted proteins are involved in photosynthesis and so are not present in apicomplexans. However, eight proteins were identified from both groups (see asterisks in Table 1), and the phylogenies of four offer some resolution (Fig. 2). The phylogenies of both dimethyladenosine synthase (Fig. 2a) and queuine tRNA ribosyltransferase (Fig. 2b) support dinoflagellates and apicomplexans as sister taxa within the red plastid clade. The latter is also of interest because it is not known from green plastids. Interestingly, however, the dinoflagellate–apicomplexan clade of queuine tRNA ribosyltransferase also includes two alphaproteobacterial sequences, the chromist plastid sequences and a *Dictyostelium* homologue, and there is no clear relationship between plastid genes in general and cyanobacterial homologues. Ultimately, the source of this protein in plastids is unclear: it is possible the proteobacteria and *Dictyostelium* each acquired this gene from plastids, but it is also possible the plastid genes are not ancestrally cyanobacterial or that there are many paralogues. In any case, the *H. triquetra* gene forms a strongly supported group with *Toxoplasma* (91–99 %), suggesting that they are mostly likely closely related, and the presence of this protein in apicomplexans is not consistent with their having a plastid of green algal ancestry. 1-Deoxy-D-xylulose-5-phosphate synthase (Fig. 2c) also places apicomplexans firmly within the red plastid clade, although the position of *H. triquetra* is unresolved within this red group. Apicomplexans are, nevertheless, moderately strongly allied to the chromist *Thalassiosira* (74–82 %) and not the green plastid lineage.

Table 1. *H. triquetra* plastid-targeted proteins predicted from cDNAs sorted by inferred evolutionary origin

The phylogenetic position of *H. triquetra* protein sequences is indicated where *H. triquetra* groups within either a clade of red algal and red algal-derived plastids (Red) or green algal and green algal-derived plastids (Green). If the *H. triquetra* position is not resolved with either of these groups but still groups within the plastid clade, this is indicated by 'Plastid'. The level of support from maximum-likelihood and weighted neighbour-joining bootstrap for these groups is indicated as follows: +++, greater than 80; ++, 70–80; +, 60–70; no symbol, below 60. Proteins for which homologues are represented in apicomplexans are indicated by asterisks.

Protein (gene)	GenBank accession no.	Phylogeny	Figure/reference
Red algal/chromalveolate			
Phosphoglycerate kinase (*pgk*)	AY826862	Red +++	Fig. 1a
Phosphoribulokinase (*prk*)	AY826860	Red +++	Fig. 1b
Photosystem II extrinsic protein (*psbU*)	AY826889	Red +++ (no green)	Fig. 1c
Dimethyladenosine synthase* (*ksgA*)	AY826874	Red	Fig. 2a
Queuine tRNA ribosyltransferase* (*tgt*)	AY826892	Red +++	Fig. 2b
1-Deoxy-D-xylulose-5-phosphate synthase* (*dxs*)	AY826876	Red	Fig. 2c
Ribose-5-phosphate isomerase (*rpiA*)	AY826893	Red +++	Supplementary Fig. S1
Cytochrome *f* (*petA*)	AY826881	Red	Supplementary Fig. S2
Cytochrome *b559* (*psbF*)	AY826887	Red (plus *B. natans*)	Supplementary Fig. S3
Transketolase (*tktA*)	AY826896	Red	Supplementary Fig. S4
GAPDH*	AY884246, AY884247	Red +++	Takishita *et al.* (2005)
Fructose-1,6-bisphosphate aldolase (*fbaA*)	AAV71135	Red +++	Patron *et al.* (2004)
Oxygen-evolving enhancer 1 (*psbO*)	AAM77465	Red +++	Ishida & Green (2002)
Green algal/plant			
Oxoglutarate/malate translocator	AY826859	Green +++ (no red)	Fig. 3a
Protochlorophyllide reductase subunit (*chlL*)	AY826880	Green	Supplementary Fig. S5
Photosystem I subunit III (*psaF*)	AY826884	Green	Supplementary Fig. S6
Non-plastid			
Acetolactate synthase (*als*)	AY826826	Bacteria +++	Fig. 3b
RuBisCO form II	AY826897	Bacteria +++	Whitney *et al.* (1995)
Red/green unresolved			
Photosystem II protein L (*psbL*)	AY826888	Plastid	Supplementary Fig. S7
Thylakoid 11 kDa protein	AY826895	Plastid	Supplementary Fig. S8
Translation elongation factor Ts (*tsf*)	AY826878	Plastid	Fig. 3c
Ferredoxin* (*petF*)	AY826847, AY826848	Plastid	Supplementary Fig. S9
Ferredoxin-$NADP^+$ reductase* (*petH*)	AY826853	Plastid	Supplementary Fig. S10
Geranylgeranyl reductase/hydrogenase	AY826855	Plastid	Supplementary Fig. S11
Beta-keto-acyl reductase	AY826869	Plastid	Supplementary Fig. S12
Carbonic anhydrase (*yadF*)	AY826838–AY826840	Plastid	Supplementary Fig. S13
ATP synthase subunit gamma (*atpC*)	AY826835	Plastid	Supplementary Fig. S14
ATP synthase subunit C (*atpH*)	AY826871, AY884249–AY884255	Plastid	Supplementary Fig. S15
Cytochrome b6 (*petC*)	AY826843	Plastid	Supplementary Fig. S16
Cytochrome c6 (*petJ*)	AY826872, AY884248	Plastid	Supplementary Fig. S17
Photosystem I protein E (*psaE*)	AY826882	Plastid	Supplementary Fig. S18
Ascorbate peroxidase	AY826833	Plastid	Supplementary Fig. S19
Adenylate kinase (*adk*)	AY826832	Plastid	Supplementary Fig. S20
Photosystem I subunit XI (*psaL*)	AY826885	Plastid	Supplementary Fig. S21
Acyl carrier protein* (*acp*)	AY826829	Plastid and mitochondrion	Supplementary Fig. S22
Lipoate protein ligase*	AY826879	Plastid and mitochondrion	Supplementary Fig. S23

GAPDH provides further support for dinoflagellates and apicomplexans grouping in the chromalveolates, although not as sisters. This gene has been analysed in detail previously (Fast *et al.*, 2001; Harper & Keeling, 2003; Takishita *et al.*, 2005) and will therefore not be described here except to state that the *H. triquetra* data are consistent with previous observations that GAPDH supports the origin of both dinoflagellate and apicomplexan plastids from the red plastid clade and that the apicomplexan homologues are specifically related to those of haptophytes, not

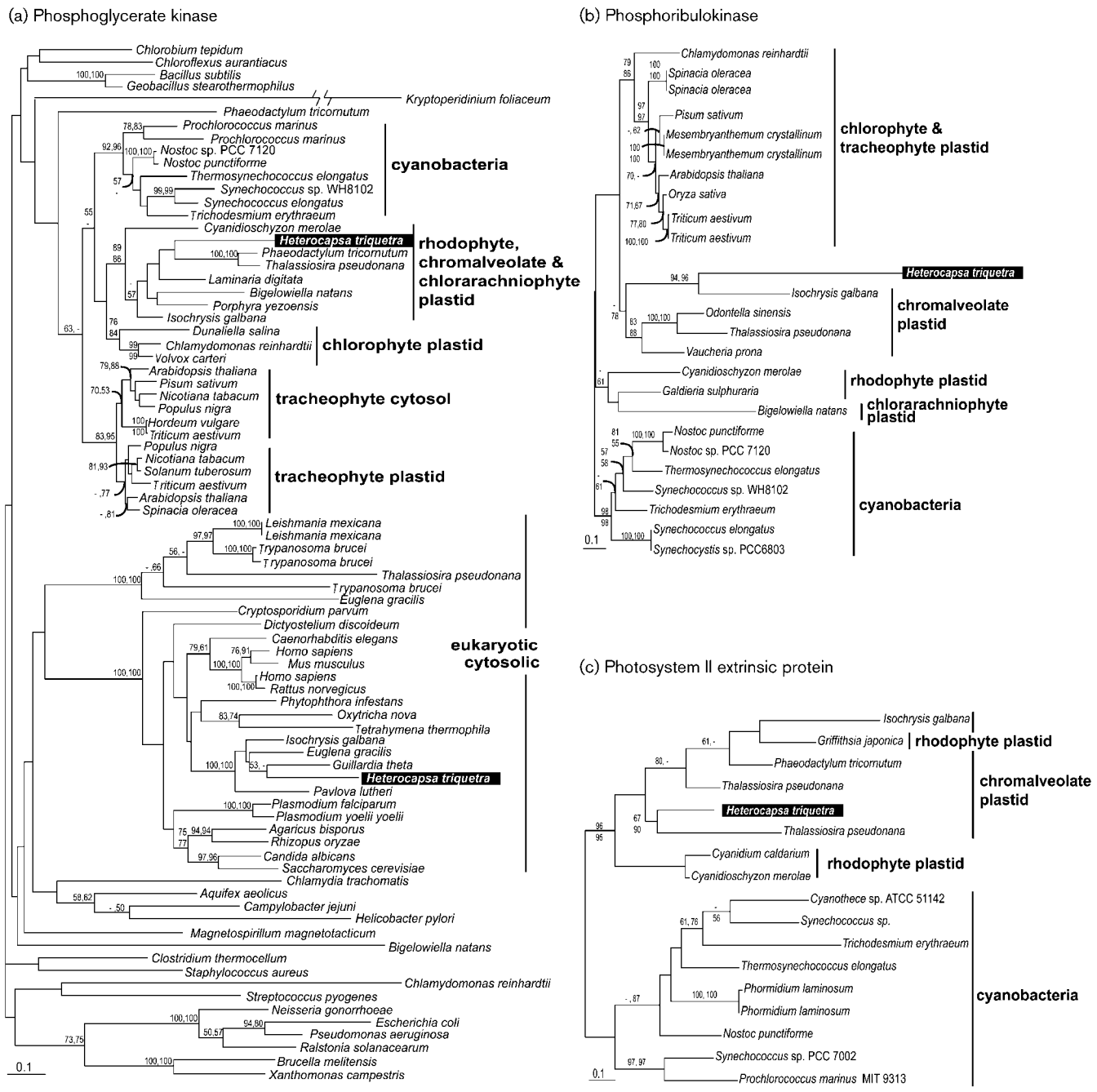

Fig. 1. Relationship of dinoflagellate plastid-targeted proteins phosphoglycerate kinase (a), queuine tRNA ribosyltransferase (b) and photosystem II extrinsic protein (c) to homologues from red algal and chromist plastids. Numbers at nodes indicate bootstrap support (ML top/left; NJ bottom/right) for major nodes over 50 % by at least one method. Major eukaryotic groups and plastids are indicated to the right.

dinoflagellates. Plastid-targeted TufA was originally used to argue for a green algal origin of apicomplexan plastids (in the absence of data from dinoflagellates) (Köhler *et al.*, 1997), but it has now been shown that the apicomplexan and dinoflagellate homologues are closely related (Hackett *et al.*, 2004a), in our view ruling this gene out as support for a green plastid in apicomplexans.

Evidence for lateral gene transfer and gene fusions in dinoflagellate plastid-targeted proteins

Previously it has been shown that *B. natans*, a chlorarachniophyte alga with a green algal secondary plastid, acquired several plastid-targeted protein genes from other

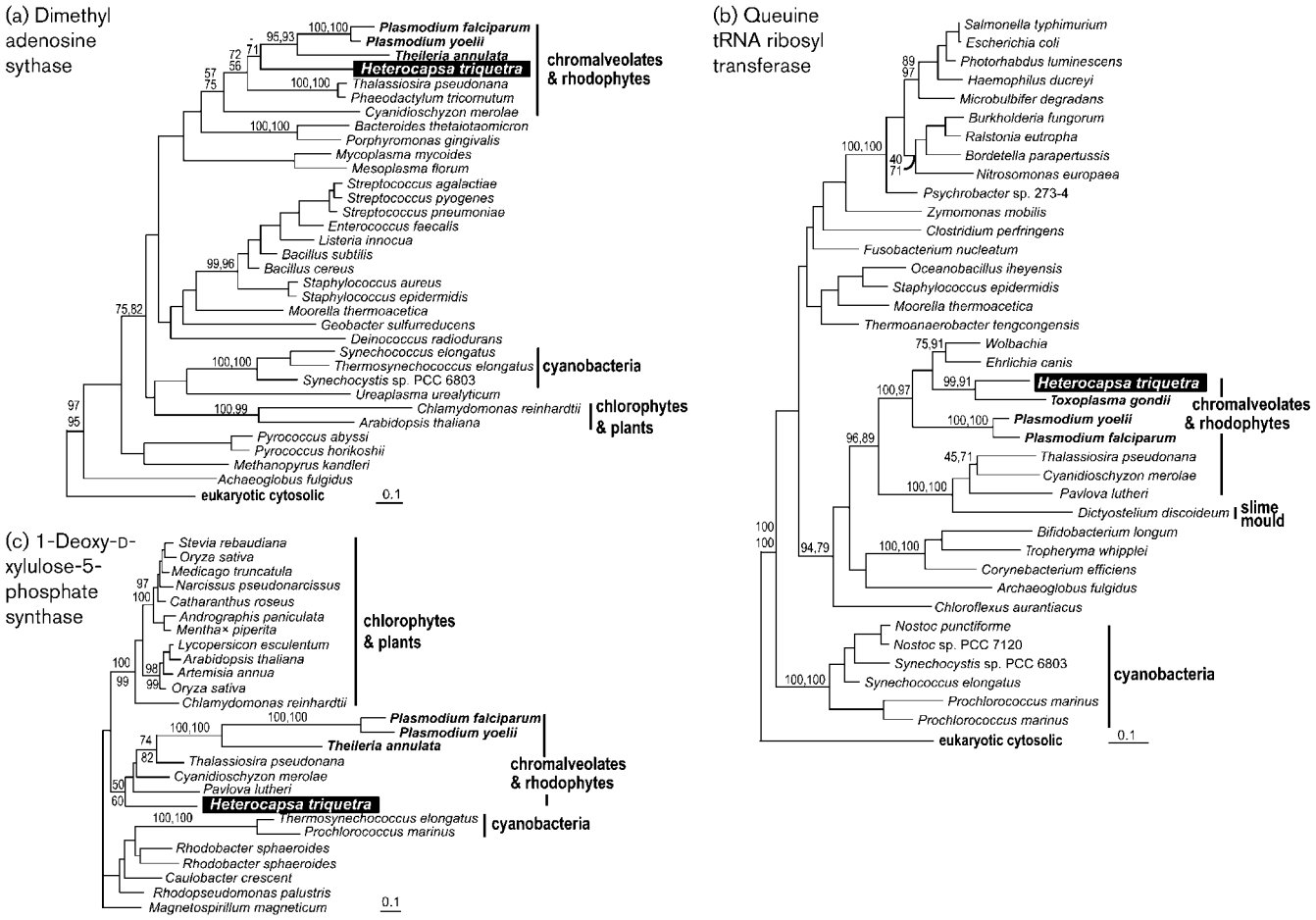

Fig. 2. Relationship between dinoflagellate and apicomplexan plastid-targeted proteins dimethyladenosine synthase (a), queuine tRNA ribosyltransferase (b) and 1-deoxy-D-xylulose-5-phosphate synthase (c). See the legend to Fig. 1 for other details.

algae, including red algae, or bacteria (Archibald *et al.*, 2003). In dinoflagellates, RuBisCO has long been known to be of a bacterial type and considered to have originated by lateral gene transfer (Rowan *et al.*, 1996; Whitney *et al.*, 1995). One additional protein, δ-aminolaevulinic acid dehydratase, has also been suggested to be derived from a green alga along with other more complex cases (Hackett *et al.*, 2004a), but the extent to which dinoflagellate plastid proteins may not originate from a single source is not well known. Other than the well-studied RuBisCO, it is similarly unclear whether any bacterial proteins have been harnessed in the plastid.

In our survey, a handful of proteins branched with the green algae or plants rather than the red plastid clade, but in all cases except one this was without support (Table 1). The one exception, the oxoglutarate/malate translocator (Fig. 3a), is interesting because, to date, the only eukaryotic sources from which it has been reported are plastids of plants and green algae. Since the preponderance of genes support the red origin of the dinoflagellate plastid, the presence of this protein in *H. triquetra* but not in finished genomes of red algae or diatoms (Armbrust *et al.*, 2004; Barbier *et al.*, 2005; Matsuzaki *et al.*, 2004) or in any other known red or red-derived plastids suggests that this dinoflagellate protein is derived by lateral gene transfer from a green source.

Even more interesting, the *H. triquetra* acetolactate synthase (Fig. 3b) appears to have originated from a non-plastid source. The protein is known to exist in cyanobacteria and other plastids, but the *H. triquetra* gene is related to neither and instead forms a highly supported group within alphaproteobacteria (100 % support), specifically related to a *Paracoccus*/*Rhodobacter* subgroup (96–98 % support). Nested well within a group of related bacteria as this is, it suggests that this gene is derived relatively recently from an alphaproteobacterial genome. The gene was represented by 13 ESTs, which indicates that it is highly expressed, and we also discovered a related homologue in a subsequent EST

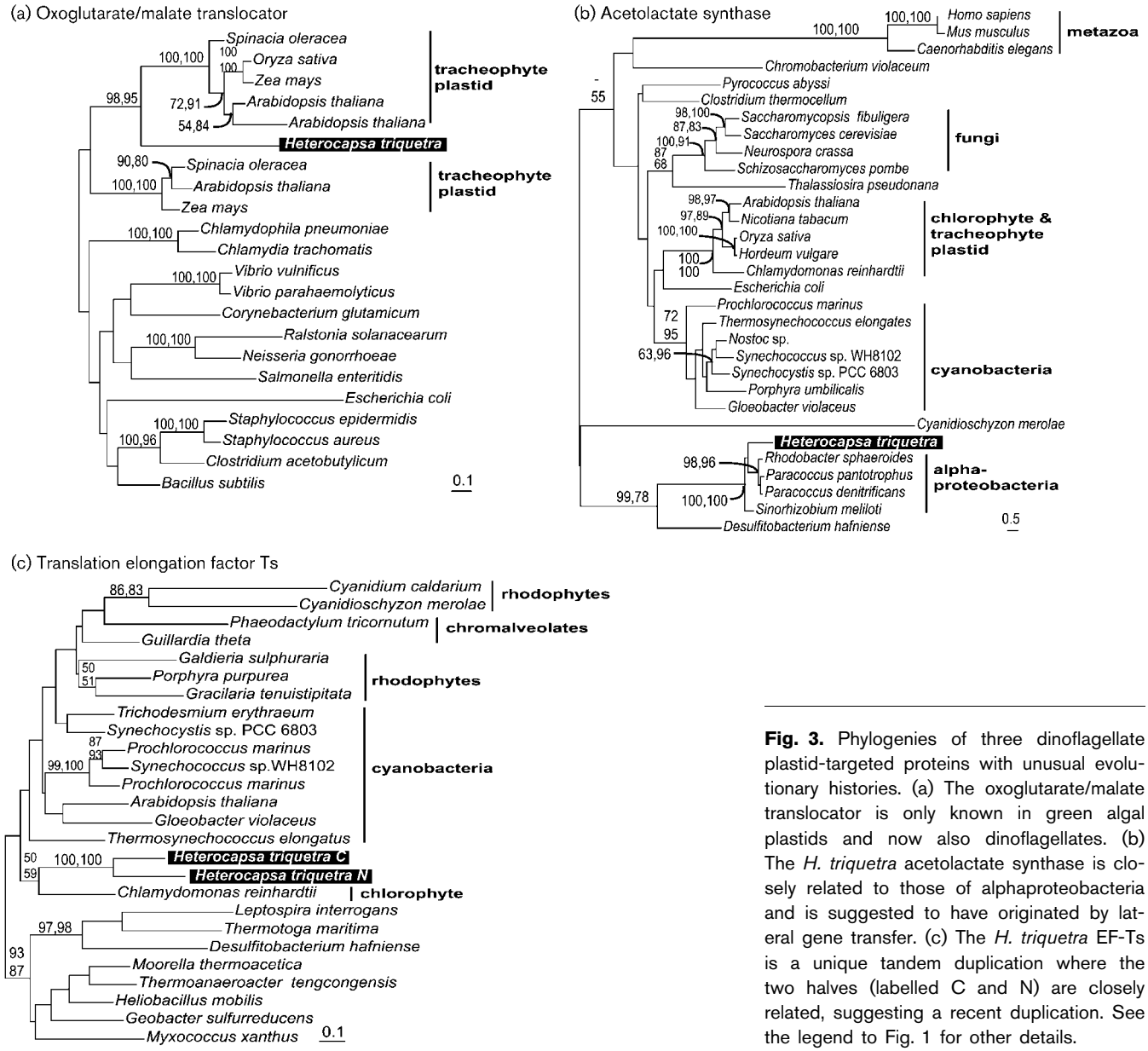

Fig. 3. Phylogenies of three dinoflagellate plastid-targeted proteins with unusual evolutionary histories. (a) The oxoglutarate/malate translocator is only known in green algal plastids and now also dinoflagellates. (b) The *H. triquetra* acetolactate synthase is closely related to those of alphaproteobacteria and is suggested to have originated by lateral gene transfer. (c) The *H. triquetra* EF-Ts is a unique tandem duplication where the two halves (labelled C and N) are closely related, suggesting a recent duplication. See the legend to Fig. 1 for other details.

survey of another species of dinoflagellate, *Karlodinium micrum* (Patron *et al.*, 2006), both strongly supporting the conclusion that this gene is encoded in the dinoflagellate genome (as opposed to being a bacterial contaminant). It is not clear whether this gene is common to other chromalveolate plastids or indeed even whether all other dinoflagellates encode this gene, but its apparent recent origin suggests that its distribution may be relatively restricted in these plastids.

Lastly, one of the *H. triquetra* genes that branches weakly with the green algae is EF-Ts. The phylogeny of this gene (Fig. 3c) is too weak to conclude much about its origin; however, it is noteworthy because of its unique structure. The *H. triquetra* gene encodes a tandem duplication of EF-Ts and the phylogeny shows that the two halves of the duplication branch together with 100 % support. This close relationship between the two halves shows that the duplication took place relatively recently, which has interesting implications for the function of this protein, which recycles GDP from elongation factor-Tu during translation elongation. Several other dinoflagellate proteins are expressed as polymers, some of which are processed and some are functional fusion proteins (Hiller *et al.*, 1995; Liu *et al.*, 2004; Rowan *et al.*, 1996). It is not clear whether the EF-Ts is processed or not. Despite this protein having a housekeeping role in translation, we could not identify a plastid version in apicomplexans, but we did find non-duplicated plastid homologues of EF-Ts in a diatom (*Phaeodactylum*) and a cryptomonad (*Guillardia*). The distribution of this

character is relatively restricted, therefore, but seeking this protein in other apicomplexans might be of interest.

Concluding remarks

Overall, the phylogeny of dinoflagellate plastid-targeted proteins supports the origin of this organelle from the red plastid lineage. This is consistent with previous suggestions based on pigmentation (chlorophyll *c* is found in dinoflagellates and chromists) and some molecular data, but the analyses described here provide more consistency than previously observed with molecular phylogenetic surveys. In general, dinoflagellate proteins also tend to branch with homologues from other chromalveolates, although no single protein shows this relationship unambiguously and the best evidence for this remains the unusual evolutionary histories of GAPDH and FBA. The few plastid proteins available from both dinoflagellates and apicomplexans tend to support a common origin of these two plastids and add further evidence for the red ancestry of apicomplexan plastids. Altogether there is very little evidence to support a green origin for either dinoflagellate or apicomplexan plastids. Confirming that the dinoflagellate plastid is related to those of apicomplexans and the chromalveolates as a whole is significant for several reasons, in particular because of the many differences between dinoflagellate and apicomplexan plastids. Distinctions in membrane number and protein targeting between dinoflagellates and other chromalveolates mean that dinoflagellates must have undergone a dramatic transformation necessitating changes to the targeting system and also to the transit peptides of perhaps hundreds of proteins (Nassoury *et al.*, 2003; Patron *et al.*, 2005). In addition, a plastid-containing ancestor of apicomplexans and dinoflagellates is interesting because deep-branching members of these two lineages, *Colpodella* and *Perkinsus*, respectively (Goggin & Barker, 1993; Kuvardina *et al.*, 2002), potentially enable us to reconstruct many characteristics of this ancestor and the subsequent evolution of both groups in detail (Leander & Keeling, 2003).

Analyses of plastid-targeted protein genes also continue to show the evolutionary complexity of dinoflagellate plastids, and indeed plastids as a whole. There are now a handful of dinoflagellate plastid proteins that do not seem to fit with the overall picture of the plastid's evolution from the red plastid lineage: this work and other studies (Hackett *et al.*, 2003) have shown some proteins being more akin to green homologues and others being only distantly related to plastid homologues in general. This is similar to observations from another myxotrophic alga, *B. natans* (Archibald *et al.*, 2003), raising interesting questions about the genetic content of such genomes in general. In addition, dinoflagellates seem to be prone to developing novelty at the molecular level (Hackett *et al.*, 2004b), as evident from their plastid proteins. Complexes appear to form between proteins from distantly related sources and new structures are found, making dinoflagellates an interesting case study of molecular diversity in microbial eukaryotes.

ACKNOWLEDGEMENTS

This work was supported by the Protist EST Program of Genome Canada/Genome Atlantic and by a grant (MOP-42517) from the Canadian Institutes for Health Research (CIHR). R. F. W. is supported by Fellowships from CIHR and the Michael Smith Foundation for Heath Research (MSFHR) and P. J. K. is a Fellow of the CIAR and a new investigator of the CIHR and MSFHR.

REFERENCES

Archibald, J. M. & Keeling, P. J. (2002). Recycled plastids: a 'green movement' in eukaryotic evolution. *Trends Genet* **18**, 577–584.

Archibald, J. M., Rogers, M. B., Toop, M., Ishida, K. & Keeling, P. J. (2003). Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. *Proc Natl Acad Sci U S A* **100**, 7678–7683.

Armbrust, E. V., Berges, J. A., Bowler, C. & 42 other authors (2004). The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**, 79–86.

Bachvaroff, T. R., Concepcion, G. T., Rogers, C. R., Herman, E. M. & Delwiche, C. F. (2004). Dinoflagellate expressed sequence tags data indicate massive transfer of chloroplast genes to the nuclear genome. *Protist* **155**, 65–78.

Baldauf, S. L., Roger, A. J., Wenk-Siefert, I. & Doolittle, W. F. (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**, 972–977.

Barbier, G., Oesterhelt, C., Larson, M. D., Halgren, R. G., Wilkerson, C., Garavito, R. M., Benning, C. & Weber, A. P. (2005). Comparative genomics of two closely related unicellular thermo-acidophilic red algae, *Galdieria sulphuraria* and *Cyanidioschyzon merolae*, reveals the molecular basis of the metabolic flexibility of *Galdieria sulphuraria* and significant differences in carbohydrate metabolism of both algae. *Plant Physiol* **137**, 460–474.

Blanchard, J. L. & Hicks, J. S. (1999). The non-photosynthetic plastid in malarial parasites and other apicomplexans is derived from outside the green plastid lineage. *J Eukaryot Microbiol* **46**, 367–375.

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.

Cai, X., Fuller, A. L., McDougald, L. R. & Zhu, G. (2003). Apicoplast genome of the coccidian *Eimeria tenella*. *Gene* **321**, 39–46.

Cavalier-Smith, T. (1998). A revised six-kingdom system of life. *Biol Rev Camb Philos Soc* **73**, 203–266.

Delwiche, C. F. (1999). Tracing the thread of plastid diversity through the tapestry of life. *Am Nat* **154** (Suppl. 4), S164–S177.

Fast, N. M., Kissinger, J. C., Roos, D. S. & Keeling, P. J. (2001). Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol Biol Evol* **18**, 418–426.

Fast, N. M., Xue, L., Bingham, S. & Keeling, P. J. (2002). Re-examining alveolate evolution using multiple protein molecular phylogenies. *J Eukaryot Microbiol* **49**, 30–37.

Funes, S., Davidson, E., Reyes-Prieto, A., Magallón, S., Herion, P., King, M. P. & Gonzalez-Halphen, D. (2002). A green algal apicoplast ancestor. *Science* **298**, 2155.

Gajadhar, A. A., Marquardt, W. C., Hall, R., Gunderson, J., Ariztia-Carmona, E. V. & Sogin, M. L. (1991). Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Crypthecodinium cohnii* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. *Mol Biochem Parasitol* **45**, 147–154.

Goggin, C. L. & Barker, S. C. (1993). Phylogenetic position of the genus *Perkinsus* (Protista, Apicomplexa) based on small subunit ribosomal RNA. *Mol Biochem Parasitol* **60**, 65–70.

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

Hackett, J. D., Maranda, L., Yoon, H. S. & Bhattacharya, D. (2003). Phylogenetic evidence for the cryptophyte origin of the plastid of *Dinophysis* (Dinophysiales, Dinophyceae). *J Phycol* **39**, 440–448.

Hackett, J. D., Yoon, H. S., Soares, M. B., Bonaldo, M. F., Casavant, T. L., Scheetz, T. E., Nosenko, T. & Bhattacharya, D. (2004a). Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. *Curr Biol* **14**, 213–218.

Hackett, J. D., Anderson, D. M., Erdner, D. L. & Bhattacharya, D. (2004b). Dinoflagellates: a remarkable evolutionary experiment. *Am J Bot* **91**, 1523–1534.

Harper, J. T. & Keeling, P. J. (2003). Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. *Mol Biol Evol* **20**, 1730–1735.

Harper, J. T., Waanders, E. & Keeling, P. J. (2005). On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. *Int J Syst Evol Microbiol* **55**, 487–496.

Hiller, R. G., Wrench, P. M. & Sharples, F. P. (1995). The light-harvesting chlorophyll a-c-binding protein of dinoflagellates: a putative polyprotein. *FEBS Lett* **363**, 175–178.

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 U S A* **99**, 9294–9299.

Jomaa, H., Wiesner, J., Sanderbrand, S. & 9 other authors (1999). Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* **285**, 1573–1576.

Keeling, P. J. (2004). Diversity and evolutionary history of plastids and their hosts. *Am J Bot* **91**, 1481–1493.

Köhler, S., Delwiche, C. F., Denny, P. W., Tilney, L. G., Webster, P., Wilson, R. J. M., Palmer, J. D. & Roos, D. S. (1997). A plastid of probable green algal origin in apicomplexan parasites. *Science* **275**, 1485–1489.

Kuvardina, O. N., Leander, B. S., Aleshin, V. V., Myl'nikov, A. P., Keeling, P. J. & Simdyanov, T. G. (2002). The phylogeny of colpodellids (Alveolata) using small subunit rRNA gene sequences suggests they are the free-living sister group to apicomplexans. *J Eukaryot Microbiol* **49**, 498–504.

Leander, B. S. & Keeling, P. J. (2003). Morphostasis in alveolate evolution. *Trends Ecol Evol* **18**, 395–402.

Liu, L., Wilson, T. & Hastings, J. W. (2004). Molecular evolution of dinoflagellate luciferases, enzymes with three catalytic domains in a single polypeptide. *Proc Natl Acad Sci U S A* **101**, 16555–16560.

Matsuzaki, M., Misumi, O., Shin-i, T. & 38 other authors (2004). Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428**, 653–657.

McFadden, G. I. (2001). Primary and secondary endosymbiosis and the origin of plastids. *J Phycol* **37**, 951–959.

McFadden, G. I. & Waller, R. F. (1997). Plastids in parasites of humans. *Bioessays* **19**, 1033–1040.

Nassoury, N., Cappadocia, M. & Morse, D. (2003). Plastid ultrastructure defines the protein import pathway in dinoflagellates. *J Cell Sci* **116**, 2867–2874.

Patron, N. J., Rogers, M. B. & Keeling, P. J. (2004). Gene replacement of fructose-1,6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. *Eukaryot Cell* **3**, 1169–1175.

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.

Patron, N. J., Waller, R. F. & Keeling, P. J. (2006). A tertiary plastid uses genes from two endosymbionts. *J Mol Biol* **357**, 1373–1382.

Ralph, S. A., van Dooren, G. G., Waller, R. F., Crawford, M. J., Fraunholz, M. J., Foth, B. J., Tonkin, C. J., Roos, D. S. & McFadden, G. I. (2004). Tropical infectious diseases: metabolic maps and functions of the *Plasmodium falciparum* apicoplast. *Nat Rev Microbiol* **2**, 203–216.

Rowan, R., Whitney, S. M., Fowler, A. & Yellowlees, D. (1996). Rubisco in marine symbiotic dinoflagellates: form II enzymes in eukaryotic oxygenic phototrophs encoded by a nuclear multigene family. *Plant Cell* **8**, 539–553.

Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504.

Takishita, K., Patron, N. J., Ishida, K., Maruyama, T. & Keeling, P. J. (2005). A transcriptional fusion of genes encoding glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and enolase in dinoflagellates. *J Eukaryot Microbiol* **52**, 343–348.

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. *Nucleic Acids Res* **25**, 4876–4882.

Waller, R. F., Keeling, P. J., Donald, R. G. & 7 other authors (1998). Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* **95**, 12352–12357.

Waller, R. F., Keeling, P. J., van Dooren, G. G. & McFadden, G. I. (2003). Comment on "A green algal apicoplast ancestor". *Science* **301**, 49.

Whitney, S. M., Shaw, D. C. & Yellowlees, D. (1995). Evidence that some dinoflagellates contain a ribulose-1,5-bisphosphate carboxylase/oxygenase related to that of the alpha-proteobacteria. *Proc Biol Sci* **259**, 271–275.

Wilson, I. R. J. M., Denny, P. W., Preiser, P. R. & 8 other authors (1996). Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J Mol Biol* **261**, 155–172.

Yoon, H. S., Hackett, J. D., Pinto, G. & Bhattacharya, D. (2002). The single, ancient origin of chromist plastids. *Proc Natl Acad Sci U S A* **99**, 15507–15512.

Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G. & Bhattacharya, D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. *Mol Biol Evol* **21**, 809–818.

Yoon, H. S., Hackett, J. D., Van Dolah, F. M., Nosenko, T., Lidie, K. L. & Bhattacharya, D. (2005). Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. *Mol Biol Evol* **22**, 1299–1308.

Zhang, Z., Green, B. R. & Cavalier-Smith, T. (1999). Single gene circles in dinoflagellate chloroplast genomes. *Nature* **400**, 155–159.

Zhang, Z., Green, B. R. & Cavalier-Smith, T. (2000). Phylogeny of ultrarapidly evolving dinoflagellate chloroplast genes: a possible common origin for sporozoan and dinoflagellate plastids. *J Mol Evol* **51**, 26–40.