

Phylogeny of Dinoflagellate Plastid Genes Recently Transferred to the Nucleus Supports a Common Ancestry with Red Algal Plastid Genes

Yunling Wang · Simon Joly · David Morse

Received: 20 October 2006 / Accepted: 9 January 2008 / Published online: 6 February 2008
© Springer Science+Business Media, LLC 2008

Abstract It is generally accepted that peridinin-containing dinoflagellate plastids are derived from red alga, but whether they are secondary plastids equivalent to plastids of stramenopiles, haptophytes, or cryptophytes, or are tertiary plastids derived from one of the other secondary plastids, has not yet been completely resolved. As secondary plastids, plastid gene phylogeny should mirror that of nuclear genes, while incongruence in the two phylogenies should be anticipated if their origin was as tertiary plastids. We have analyzed the phylogeny of plastid-encoded genes from *Lingulodinium* as well as that of nuclear-encoded dinoflagellate homologues of plastid-encoded genes conserved in all other plastid genome sequences. Our analyses place the dinoflagellate, stramenopile, haptophyte, and cryptophyte plastids firmly in the red algal lineage, and in particular, the close relationship between stramenopile plastid genes and their dinoflagellate nuclear-encoded homologues is consistent with the hypothesis that red algal-type plastids have arisen only once in evolution.

Keywords Chloroplast · Dinoflagellate · Secondary endosymbiosis · Molecular phylogeny

Introduction

Photosynthetic organisms contain a bewildering array of different chloroplast types whose evolutionary origins are best described by the endosymbiotic hypothesis (Delwiche 1999; Martin et al. 1998; Palmer 2003). For example, rhodophytes, chlorophytes, and glaucocystophytes contain plastids surrounded by two bounding membranes. This membrane structure suggests that these organelles were derived from an ancestral endosymbiotic event between a free-living photosynthetic organism and a nonphotosynthetic host. Separate phylogenetic reconstructions using proteins known to be encoded either in the plastid or in the nuclear genome both suggest that these three groups are monophyletic (Rodriguez-Ezpeleta et al. 2005), and these so-called primary plastids would thus share common ancestry. However, many other algae have complex plastids, surrounded by either three (euglenids and the peridinin-containing dinoflagellates) or four (the stramenopiles, cryptophytes, haptophytes, apicomplexans, and chloroarchniophytes) membranes. These membrane structures are consistent with secondary endosymbiotic events, where an organism containing a primary plastid established permanent residence in a new host cell, and these organelles are termed secondary plastids to reflect this. These secondary plastids are generally divided into two general classes, those clearly derived from red algae (stramenopiles, cryptophytes, haptophytes, and dinoflagellates) and those derived from green algae (euglenids and chloroarchniophytes).

Interestingly, the dinoflagellates appear particularly adept at recruiting plastids from other organisms. Several species have been found to contain what are now believed to be tertiary plastids, obtained through endosymbiosis with an organism containing a secondary plastid (Patron

Y. Wang · D. Morse (✉)
Institut de Recherche en Biologie Végétale, Département de
Sciences Biologiques, Université de Montréal,
4101 Sherbrooke est, Montréal, Québec, Canada H1X 2B2
e-mail: david.morse@umontreal.ca

S. Joly
Allan Wilson Centre, Massey University, Palmerston North,
New Zealand

et al. 2006; Tengs et al. 2000). These unusual plastid types differ in pigment content from the most common type of dinoflagellate plastid, which contains the carotenoid peridinin as an accessory pigment in light harvesting. This peridinin plastid is exemplified by the well-studied species *Alexandrium tamarense*, *Amphidinium carterae*, *Heterocapsa niei*, and *Lingulodinium polyedrum* (formerly *Gonyaulax polyedra*). These plastids have an unusual genome architecture, with only 12–16 genes encoded as a series of small generally unigenic minicircles (Koumandou et al. 2004; Wang and Morse 2006; Zhang et al. 1999) instead of the typically 130–200 genes found on the single circular plastid chromosome. The remaining genes normally encoded by the plastid genome have all been transferred to the nucleus in the peridinin dinoflagellates (Bachvaroff et al. 2004; Hackett et al. 2004). Other dinoflagellates, such as *Karenia brevis* and *Karlodinium micrum*, contain plastids that are surrounded by four membranes and, instead of peridinin, contain 19'-hexanoyloxyfucoxanthin, a carotenoid characteristic of haptophytes. A haptophyte origin for these plastids origin has been confirmed by phylogenetic analyses (Patron et al. 2006; Yoon et al. 2005). In addition, *Lepidodinium viride* contains a green algal plastid (Watanabe et al. 1987), while *Dinophysis* contains a cryptophyte plastid (Hackett et al. 2003). Finally, some species contain a membrane-bound eukaryotic endosymbiont, thought to be a stramenopile in the dinoflagellates *Peridinium balticum* (Chesnick et al. 1997) and *Kryptoperidinium foliaceum* (McEwan and Keeling 2004), and this has been interpreted as an intermediate stage in endosymbiotic plastid acquisition. The species containing unusual plastids, as well as nonphotosynthetic species, are dispersed among the dinoflagellate rRNA trees, suggesting that the peridinin-containing plastid is the ancestral state and that there have been many losses and replacements (Saldarriaga et al. 2001).

The propensity of the dinoflagellates to take in plastids by tertiary endosymbiosis leads to the question of whether the ancestor of the peridinin-type plastid is itself a tertiary plastid. Previous studies have shown a plastid gene phylogeny consistent with a close relationship either to the stramenopiles (Yoon et al. 2005) or to the haptophytes (Bachvaroff et al. 2005). While these may appear on the surface to be relatively minor differences, the evolutionary consequences are remarkably profound; grouping of stramenopile and dinoflagellate plastids to the exclusion of those of haptophytes agrees with nuclear gene phylogenies and is thus consistent with a single secondary plastid origin, while grouping of the haptophyte and dinoflagellate plastids to the exclusion of stramenopiles is incongruent with nuclear gene phylogenies and would support instead either multiple secondary plastid origins or a tertiary plastid origin. Unfortunately, the phylogenetic

reconstructions of plastid-encoded genes, which would be expected to most faithfully recapitulate plastid ancestry, are hampered by a relatively high rate of sequence evolution that could result in long-branch attraction artifacts in phylogenetic reconstruction (Felsenstein 1978), the paucity of different sequences available in the peridinin-type plastid, and the unusual minicircular chromosomal architecture that has removed all traces of gene order within the genome. The limited gene content found in dinoflagellate plastids is exacerbated in apicoplasts, the relic plastids found in apicomplexans (Waller and McFadden 2005). Nuclear gene phylogenies clearly place apicomplexans as a sister group to the dinoflagellates (Wolters 1991) as part of the Alveolates (Fast et al. 2002), a larger grouping that also includes the nonphotosynthetic ciliates. However, the highly reduced gene content of the apicoplasts has led to some controversy about their possible origin from red (Fast et al. 2001; Waller et al. 2003) or green (Funes et al. 2002; Kohler et al. 1997) plastids.

To resolve the phylogeny of the dinoflagellate plastids despite their reduced gene content, various studies have pursued phylogenetic analyses of nuclear-encoded plastid-directed proteins. Unfortunately, many are uninformative, such as the Rubisco (Delwiche and Palmer 1996) and GAPDH (Fagan et al. 1998) sequences commonly employed to reconstruct plastid phylogeny in other organisms; these genes have revealed an astonishing degree of lateral gene transfer that confounds interpretation of the phylogenetic reconstructions. Some genes appear to defy accepted phylogenies, such as the phosphoribulokinase in secondary red algal-type plastids, which supports an assemblage with green primary plastids (Petersen et al. 2006). The variety of different evolutionary patterns demonstrated by individual gene phylogenies (Hackett et al. 2004; Patron et al. 2006) may reflect the fact that they may have been derived from many different sources. Alternatively, the incongruences could reflect the small amount of phylogenetic information contained in any single gene.

We therefore considered the possibility that genes recently transferred to the nucleus from plastids might provide an acceptable alternative. An examination of the rate of sequence evolution among different classes of genes in the peridinin dinoflagellates shows higher rates among minicircle genes (remaining in the plastid) and generally lower rates among nuclear genes thought to have been transferred from the plastid genome (Bachvaroff et al. 2006). This suggested that the phylogeny of genes found in the nucleus of peridinin dinoflagellates but still maintained in the plastids of other organisms might allow the plastid phylogeny to be inferred more precisely. We find that phylogenetic reconstructions based on this class of sequence do indeed provide a more robust support for a

relationship to stramenopile plastids than previously observed and thus support a red algal ancestry for the peridinin-containing dinoflagellate plastids.

Methods

Sequence Analyses

Three approaches were undertaken to address the phylogenetic origin of peridinin-containing plastids. In the first we took advantage of our recent determination of the *Lingulodinium* plastid genome by saturating coverage of plastid gene expression (Wang and Morse 2006). The phylogeny of these genes was of interest because it was possible that they could have revealed a similarity to other red algal sequences that was hidden by sequence divergences as seen with other dinoflagellates. In all, 10 sequences were used for the reconstructions (*atpA* and *-B*; *petB* and *-D*; *psaA* and *-B*; *psbA*, *-B*, *-C*, and *-D*). These sequences have been deposited in GenBank under accession numbers DQ264844 through DQ264857. The concatenated sequences with gaps removed contained 4088 characters.

For the second analyses we sought nuclear-encoded plastid-directed genes. Two approaches were tested, the first using screens of dinoflagellate EST banks using Phootprinter (<http://theileria.ccb.sickkids.ca/phylo/cgi-bin/phootprinter.cgi>) to identify genes found in *Alexandrium tamarense* and at least one other peridinin-containing dinoflagellate (*Lingulodinium polyedrum* or *Amphidinium carterae*) that were absent from the dinoflagellate *Karenia brevis*, whose plastid is derived from a haptophyte. This approach did not take into account recent ESTs from the dinoflagellate *Karlodinium micrum*, whose plastids are also haptophyte-derived (Patron et al. 2006), and may have identified genes unlikely to be in the plastid at the time when the peridinin plastid became established. The phylogeny of many of these sequences has been reported previously (Patron et al. 2006), and as they appear to have widely divergent relationships suggestive of lateral gene transfer, they were not pursued further.

As an alternative approach to identifying nuclear-encoded plastid-directed genes, we tested all genes common to all completely sequenced primary and secondary plastid genomes for homologues to sequences in dinoflagellate EST banks (using tBLASTn searches and restricting the search to EST banks with dinoflagellate as the organism). The dinoflagellate sequences found in these EST banks are derived using primers exploiting their poly(A) tail and are thus not plastid-encoded, as plastid-encoded genes are polyuridylylated in dinoflagellates (Wang and Morse 2006). All sequences recovered were examined

individually using MacVector software for the presence of an N-terminal extension consistent with known plastid targeting leader sequences (Nassoury et al. 2003; Patron et al. 2005) and the presence of a stop codon at the expected position. Twenty sequences were identified as nuclear-encoded plastid-directed proteins using these criteria (*atpE*, *-F*, and *-H*; *petA*; *psaC* and *-J*; *psbF*, *-H*, *-K*, *-L*, *-N*, and *-T*; *rpl15* and *-16*; *rps2*, *-7*, *-11*, *-12*, and *-19*; *ycf3*). All the individual sequences were aligned using CLUSTAL W (MacVector), and a phylogenetic grouping among the red algal sequences was confirmed by maximum likelihood. Gaps were then removed from the individual gene alignments before concatenation, to produce a final alignment containing 2160 characters.

Interestingly, a subset of these nuclear-encoded plastid-directed genes (*rps7*, *-11*, *-12*, and *-19*; *rpl14*) was found to have potential apicomplexan homologues. Plasmodium homologues were identified by BLAST searches of the EST databank (limited to Alveolata), and some (but not all) could be characterized as potentially plastid directed proteins by virtue of N-terminal extensions enrichment in lysine (K) and asparagine (N) residues (not all were long enough to show this feature) (Foth et al. 2003). Of these five genes, *rps7* clearly showed a different evolutionary history from the others, placing *Porphyra* and *Synechocystis* within Streptophyta, sister to *Zea mays* (data not shown). Consequently, this gene was removed from the concatenated analysis, resulting in a dataset of four genes totaling 467 characters that was analyzed separately in order to test for a predicted relationship to Apicomplexans. To address this same issue differently, the same five ribosomal sequences were recovered from the sequenced plastid genomes from the apicomplexans *Eimeria tenella*, *Theileria parva* and *Toxoplasma gondii*, resulting in a concatenated dataset of 625 amino acids. Preliminary maximum likelihood (ML) analyses did not show evidence of phylogenetic incongruence among genes in this dataset (data not shown), which justifies the concatenated analysis. Despite a lower degree of sequence similarity, and correspondingly longer branch lengths, the presence in the organelle genome makes the origin of these sequences unambiguous.

Phylogenetic Analyses

For each concatenated dataset, phylogenies were reconstructed using ML and Bayesian approaches. For each dataset, the best evolutionary model was selected according to the Akaike Information Criterion in ProtTest (version 1.3; Abascal et al. 2005), fitting the model on a BIONJ tree and excluding matrices developed for mitochondrial and reverse transcriptase genes; the CpREV (Adachi et al.

2000) + I + Γ model was selected for all datasets except for the nuclear-encoded plastid genes, for which CpREV + Γ was selected. ML phylogenies were reconstructed using the PHYML program (Guindon and Gascuel 2003). The stability of the nodes for the ML analysis was assessed by nonparametric bootstrap (Felsenstein 1985) in PHYML using 100 replicates. The Bayesian phylogenies were obtained using the MrBayes program (version 3.1.2; Ronquist and Huelsenbeck 2003), where the rate heterogeneity (four discrete categories) and the proportion of invariable sites were optimized independently for the different genes included in the analyses. Five million generations were run for two independent analyses of four chains each, sampling every thousand generations. The convergence of the two MCMC runs to the stationary distribution was determined by looking at the standard deviation of splits frequencies (always <1%) and by the convergence of the parameter values in the two independent runs. Proper mixing of the runs for all parameters was confirmed in Tracer (version 1.3; Rambaut and Drummond 2005) to ensure convergence had been reached (standard deviation of clade posterior probabilities <1%, convergence of lnL values). Convergence was always reached before 1 million generations in all analyses and the trees sampled during the last 4 million generations from the two separate runs were used for computing posterior probabilities of trees and clades. All trees were rooted using cyanobacterial sequences. In order to investigate whether the phylogenies obtained were dependent on the evolutionary model, we also analyzed all datasets with the WAG (Whelan and Goldman 2001) + Γ model because WAG was the second-best substitution matrix selected by Prot-Test for all datasets.

To compare the topology of the trees obtained using ML and Bayesian approaches and also with different datasets, Kishino–Hasegawa (KH; only two trees compared) and Shimodaira–Hasegawa (SH; more than two trees compared) comparisons were performed to test whether the topological differences observed were statistically significant (Kishino and Hasegawa 1989; Shimodaira and Hasegawa 1999). To compare the trees obtained for the plastid-encoded and the nuclear-encoded datasets, the dinoflagellates—which were represented by different species—were assumed to be monophyletic and were represented by a single species (the choice of the species did not affect the results), and the SH test was performed using both nuclear and plastid alignments. The trees obtained with datasets including *Apicomplexans* could not be compared to that of the first two datasets because they included a different set of species. KH and SH analyses were performed using the Proml program from the Phylip package (version 3.65; Felsenstein 2004) using the JTT substitution matrix (Jones et al. 1992).

Results

The phylogenetic analyses performed using plastid-encoded genes might be expected to most faithfully reconstruct the evolutionary relationships between the different organelles, as there is little likelihood that lateral gene transfer has occurred between organelles to obscure their origins. We therefore analyzed the phylogeny of all sequences obtained from an EST project that identified all plastid-encoded genes in the dinoflagellate *Lingulodinium*. These sequences were combined with cyanobacterial genes and plastid-encoded genes representative of the primary and secondary plastids whose organellar genome sequences are available. As shown (Fig. 1), both ML and Bayesian analyses produced reconstructions with similar topology and the two trees were not statistically different according to a KH test ($p > 0.05$). The three classes of primary plastids found in green algae and higher plants (chlorophytes and streptophytes), red algae, and glaucocystophytes are all clustered separately from one another, as are the cyanobacterial individuals. Furthermore, the secondary plastids derived from red algae are found together with the red algal primary plastids. However, while the dinoflagellate sequences are solidly within a red algal clade using both analyses, the closest relative to the dinoflagellates differs between the two. A close relationship of the dinoflagellate plastid genes to those of the stramenopile *Odontella* is strongly supported in the Bayesian analysis, while the ML tree places the dinoflagellates sister to the main red algal lineage. The long branch lengths found for the different dinoflagellates are typical of this class of organisms and suggestive of a high evolutionary rate within the organelles (Bachvaroff et al. 2006; Zhang et al. 1999). In fact, the basal position of the Dinophyceae in the ML tree is unexpected for a secondary plastid derived from red algae and may be the result of long branch attraction.

To obtain a different perspective on the evolutionary position of the dinoflagellate plastids, a gene set found in all other sequenced plastid genomes and containing homologues in dinoflagellate EST banks was prepared. Only potential dinoflagellate homologues that were plastid-directed as defined by the presence of N-terminal extensions consistent with a plastid-directed protein (Nassoury et al. 2003; Patron et al. 2005) were retained for analysis. All sequences individually produced consistent phylogenetic inferences (data not shown), so they were concatenated for a more robust analysis. ML and the Bayesian analyses resulted in identical trees in which the dinoflagellates grouped with the stramenopiles with moderate to high support (Fig. 2). Once again, we note the relatively long branch lengths of the dinoflagellate sequences, which have been interpreted as a result of high mutation rates occurring in the plastids prior to transfer of

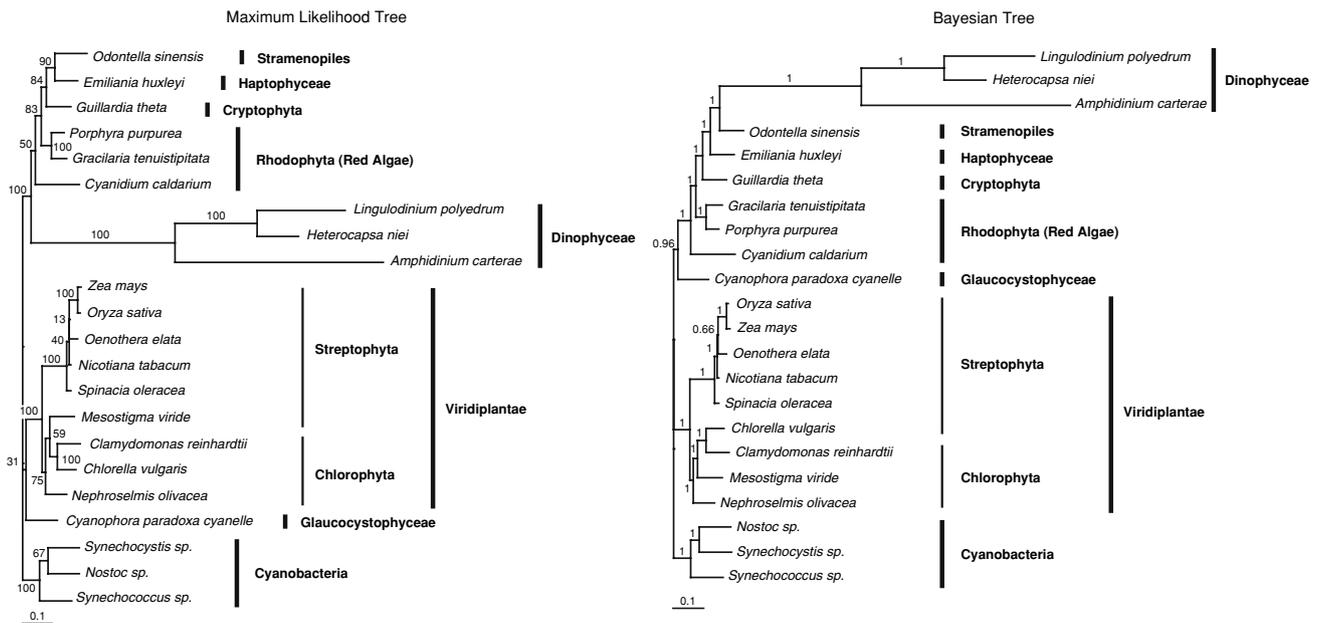
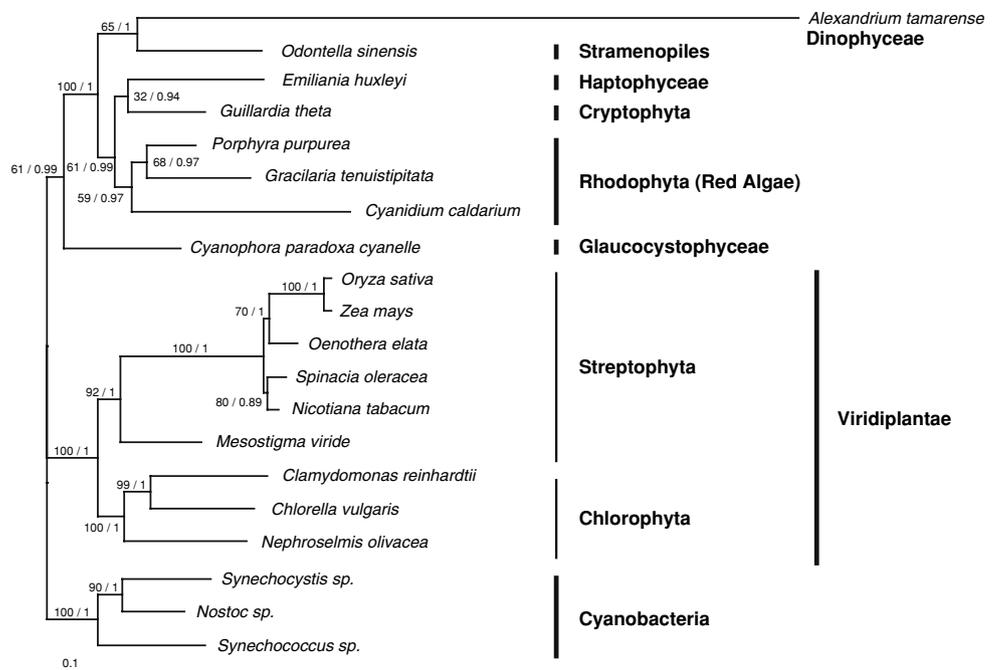


Fig. 1 Phylogenetic trees constructed using maximum likelihood (left) and Bayesian (right) methods (CpREV + I + Γ) from 10 plastid-encoded protein sequences (4088 amino acids). Support values (bootstrap for the maximum likelihood tree and clade posterior

probabilities for the Bayesian tree) are shown at internodes, and branch lengths are proportional to the number of substitutions per site (scale bars at bottom)

Fig. 2 Maximum likelihood tree (CpREV + Γ) of 20 nuclear-encoded plastid-directed dinoflagellate protein sequences (2160 amino acids); these sequences are encoded in the plastid for all other species. Bootstrap values (left) and posterior probabilities (right) for clades are shown at internodes, and branch lengths are proportional to the number of substitutions per site (scale bars at bottom). The topology of the Bayesian consensus tree was identical to that of the maximum likelihood tree



the plastid genes to the nucleus (Bachvaroff et al. 2006). Despite this, the divergence of the dinoflagellate sequences is proportionally smaller relative to the other species in the nuclear-encoded dataset than is found in the plastid-encoded dataset, suggesting that the position of the dinoflagellate in the phylogenetic analyses is less likely to be influenced by convergent evolution.

In general, the sequences encoded in the nucleus gave phylogenetic relationships similar to those when genes encoded in the dinoflagellate plastids were analyzed, yet their topologies are significantly different according to the SH test ($p < 0.001$). Notwithstanding the ML plastid-encoded tree, all other analyses placed the dinoflagellate *Alexandrium* close to the stramenopile sequence with

moderate to strong support. Thus, taken together, these analyses suggest that among the species sampled the plastids in the dinoflagellates may be most closely related to those in the stramenopiles.

While these analyses can be taken as support of an ancestral secondary endosymbiotic event from which the secondary plastids belonging to the red lineage have all been derived, it must be noted that stramenopiles are the closest relatives to the dinoflagellates among the species shown. Thus it is still possible that the dinoflagellate plastid might have been derived from a tertiary endosymbiotic event between a stramenopile and a dinoflagellate host. To attempt to distinguish between these possibilities, we attempted to find nuclear-encoded plastid-directed dinoflagellate gene homologues in the *Plasmodium* genome. The rationale for this approach is that apicomplexans are most closely related to the dinoflagellates (Wolters 1991) and, together with the ciliates, form the Alveolata. Thus the dinoflagellates would be closer to the apicomplexans if the plastid were derived from a single ancestral secondary endosymbiotic event, but would be closer to the stramenopiles if the plastid were derived from a tertiary endosymbiosis with a stramenopile. The apicomplexans have a relic plastid (McFadden et al. 1996), although its reduced gene complement and lack of photosynthetic genes make a rigorous determination of its origin extremely

difficult. Our analysis recovered a selection of five potential ribosomal gene homologues, some of which were found to contain N-terminal extensions consistent with possible plastid targeting sequences (Foth et al. 2003). However, only four of these were included in the concatenated analysis because one of the genes (*rps7*) showed a different evolutionary history from the others, placing *Porphyra* and *Synechocystis* within Streptophyta, sister to *Zea mays* (data not shown). The small number of characters included in the analysis is reflected by the low support obtained for several nodes in the phylogenies (Fig. 3). Moreover, the differences between the two trees are not significant (KH test). The analyses grouped the dinoflagellate and the apicomplexan together with high support, but the extreme divergence of this group from the remaining sequences makes ambiguous the exact position of this branch on the phylogeny. The proximity of the apicomplexan sequences with the plastid-directed ribosomal protein genes of dinoflagellates is interesting, however, and may reflect a real grouping. Alternatively, it may reflect long branch attraction artifacts between the dinoflagellate and the apicomplexan sequences. These caveats make it difficult to reach a definite conclusion from these analyses, although it is tempting to speculate that the long branch lengths in both apicomplexan and dinoflagellate sequences might have resulted from a rapid evolution

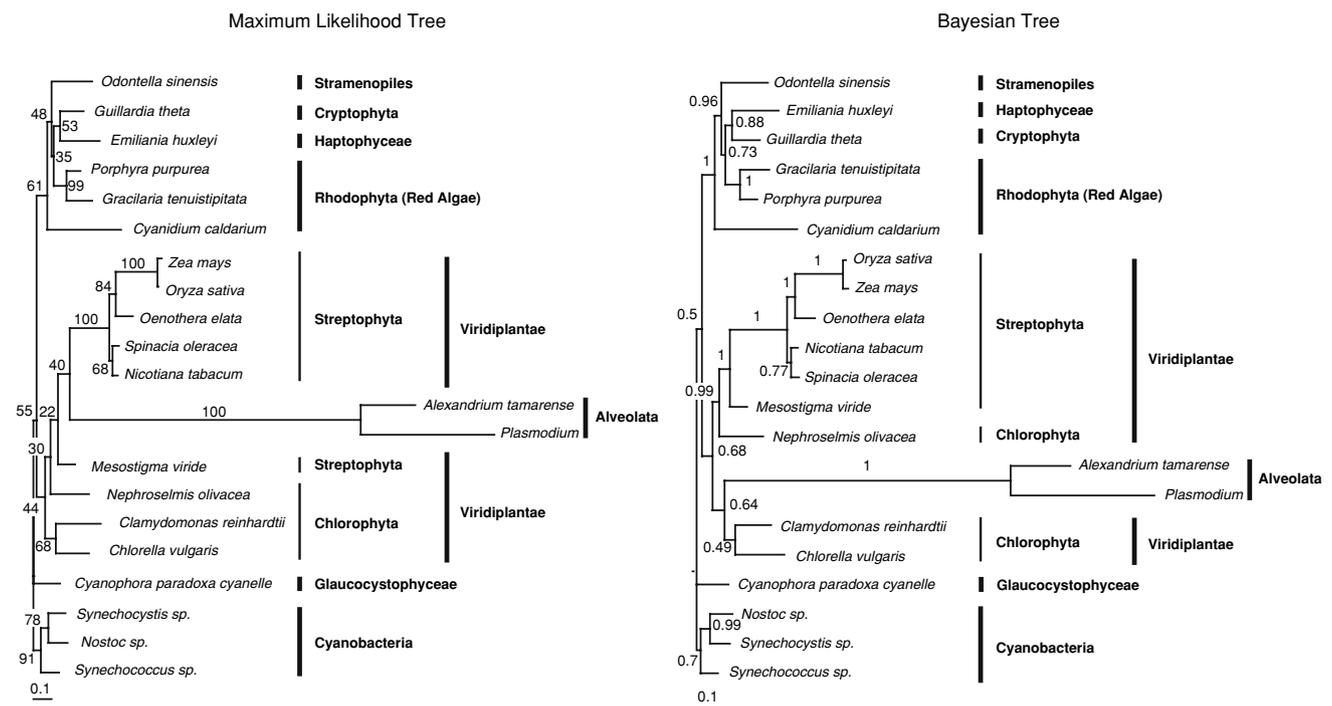


Fig. 3 Phylogenetic trees constructed using maximum likelihood (left) and Bayesian (right) methods (CpREV + I + Γ) of four nuclear-encoded plastid-directed protein sequences also found in the *Plasmodium* genome (467 amino acids). Bootstrap values or posterior

probabilities are shown at internodes, and branch lengths are proportional to the number of substitutions per site (scale bars at bottom)

of the ancestral plastid prior to divergence of the dinoflagellates and the apicomplexans. We also note a weak affinity between the alveolate and the Viridiplantae sequences in these analyses, an interesting observation given the controversy surrounding the origin of apicomplexan plastid as derived from either a red or a green algal precursor.

As a complement to these analyses using nuclear-encoded *Plasmodium* sequences, an analysis was also carried out with the same sequences obtained from the sequenced plastid genomes of three other apicomplexans (Fig. 4). Bayesian and ML analyses produce trees of identical topology except for the poorly supported relative positions of *Nephroselmis* and *Mesostigma*, and concur in the grouping of apicomplexans with dinoflagellates. Interestingly, the Alveolata group is immediately adjacent to the stramenopiles, as would be predicted by a common origin of the plastids. However, bootstrap support and clade posterior probability is only moderate for this grouping, as can also be evidenced by the inclusion of the rhodophyte *Cyanidium* sequences within the group.

All analyses, when performed using the suboptimal WAG + Γ model, resulted in trees identical to those presented here except for (i) the position of *Nicotiana* and *Oenothera*, which are interchanged in the Bayesian analysis of the chloroplast dataset (Fig. 1); (ii) a modified

position of *Guillardia* and *Cyanidium* within the Cryptophyta/Haptophyceae/Rhodophyta clade in the ML analyses of the dinoflagellate nuclear-encoded genes (Fig. 2); and (iii) the position of *Cyanophora*, which was basal to the red algae clade for datasets 3 and basal to the green algae clade for dataset 4. Overall, the few modifications observed are all in regions that received low support, indicating that the results obtained are robust with respect to the evolutionary model used.

Discussion

Relationships among the plastids containing chlorophyll c are difficult to resolve. In many of the previous analyses, there is strong support for the monophyletic origin of the chlorophyll c plastid from red algae since the plastid phylogeny (Harper and Keeling 2003; Yoon et al. 2002, 2005) and the phylogeny of the host cells (Baldauf et al. 2000; Rodriguez-Ezpeleta et al. 2005) are in good agreement. Taken together, these results suggest that the different chlorophyll c-containing plastids may have arisen from a single secondary endosymbiotic event. However, it must be noted that other studies present slightly different views, such as a phylogenetic analysis with concatenated plastid proteins, which placed the haptophyte *Emiliana*

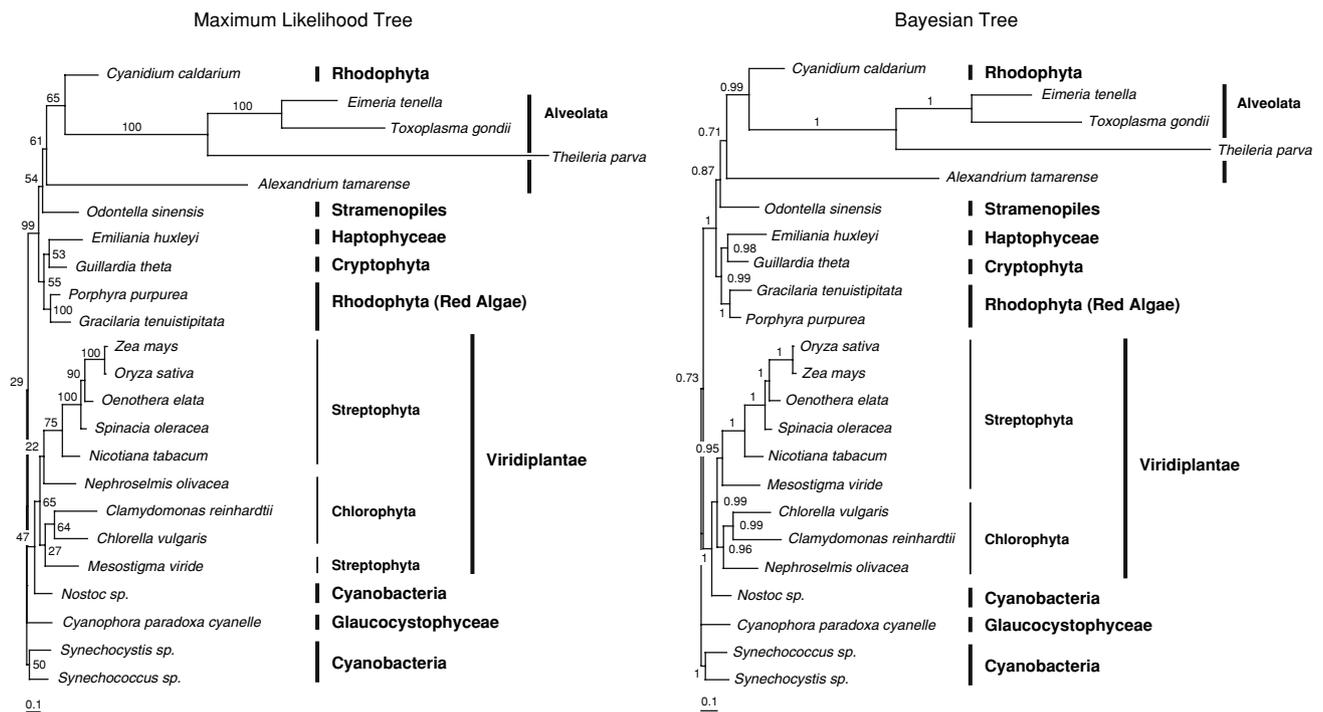


Fig. 4 Phylogenetic trees constructed using maximum likelihood (left) and Bayesian (right) methods (CpREV + I + Γ) of five nuclear-encoded plastid-directed protein sequences also found in the plastids of *Eimeria*, *Theileria*, and *Toxoplasma* (625 amino acids).

Bootstrap values or posterior probabilities are shown at internodes, and branch lengths are proportional to the number of substitutions per site (scale bars at bottom)

sister to the dinoflagellate *Amphidinium*, with the stramenopile *Odontella* sister to this previous clade (Bachvaroff et al. 2005). In part, these differences reflect an inherent difficulty in resolving the relationships between these various groups, although differences in datasets and methods of analysis may also contribute. However, should the dinoflagellate plastids turn out to be most closely related to the haptophytes, the evolutionary consequences would be considerable: instead of a single origin, the secondary plastids must have evolved several times, either as multiple independent events from closely related red algae or as a single secondary endosymbiosis followed by a series of tertiary endosymbiotic events. Clearly, this is an important issue that is strongly handicapped by the rapid evolutionary rate of the dinoflagellate plastid genes and the paucity of genetic information available in dinoflagellate plastids. Our analyses using plastid-encoded genes from *Lingulodinium* do not help in resolving the plastid origin, as our sequences are as divergent as those from other dinoflagellates. However, given the clustering of the dinoflagellates and the high rate of sequence divergence, these plastid phylogenies might be of potential use in phylogenetic reconstructions within the dinoflagellates.

A number of authors have addressed plastid phylogeny by turning to nuclear-encoded plastid-directed gene sequences. For example, the *psbO* gene phylogeny in both peridinin- and fucoxanthin-containing dinoflagellate plastids shows that the latter is related to haptophytes and distinct from that of the former (Ishida and Green 2002), although stramenopile sequences were not included in these analyses. Stramenopiles were included in an analysis of intrinsic light harvesting proteins, which placed dinoflagellates relatively basal in a red algal clade including stramenopiles, with haptophytes and cryptomonads more distantly related (Durnford et al. 1999). Recently, EST sequencing projects have provided a wealth of data that can be mined for plastid genes in both dinoflagellates and other organisms, and as an example, phylogenetic analyses of 34 discrete plastid-directed proteins from the dinoflagellate *Heterocapsa* were reported (Waller et al. 2006). While these sequences showed a general tendency to branch with red algae or with red algal secondary plastids, the precise position varied between individual phylogenies. These individual gene phylogenies thus support a monophyletic origin for the dinoflagellate plastid but cannot provide an unambiguous answer as to the plastid from which it is derived. This is because the little information contained in any single gene makes phylogenetic reconstructions from them more likely to be affected by stochastic error (i.e., sampling error). By concatenating multiple genes, stochastic error is minimized in phylogenetic reconstruction, although systematic errors such as long branch attraction may still influence the analyses (Delsuc et al. 2005).

One interesting prediction that should follow from a monophyletic origin of dinoflagellate plastids is that nuclear-encoded genes characteristic of peridinin-type plastids may still remain in the few dinoflagellates thought to have obtained their current plastids via tertiary endosymbiosis. Indeed, a bank of over 11,000 unique ESTs from *Karlodinium micrum*, whose plastids are of haptophyte origin, clearly showed genes homologous to both peridinin-containing dinoflagellates and haptophytes (Patron et al. 2006). Disappointingly, however, an analysis of over 5000 unique genes from *Karenia brevis*, whose plastids are also of haptophyte origin, shows no evidence of any characters normally associated with the peridinin-type plastids (Yoon et al. 2005). These authors' view that these results might represent extensive genome remodeling is intriguing, although the possibility that the genes are present and simply not expressed cannot be excluded.

We have identified a number of nuclear sequences in dinoflagellates that are encoded by plastids in all other organisms. These sequences are clearly nuclear-encoded because they are found in libraries representing polyadenylated transcripts rather than the polyuridylation characteristic of plastid-encoded genes (Wang and Morse 2006) and are plastid-directed because they contain N-terminal extensions characteristic of plastid targeting sequences (Nassoury et al. 2003; Patron et al. 2005). These sequences also have relatively long branch lengths in phylogenetic analyses, consistent with the idea that they may have experienced to some degree the high evolutionary rates seen for genes currently within the organelle (Bachvaroff et al. 2006). These genes are thus likely to have been transferred to the nucleus from the peridinin-containing plastid itself. It is thus significant that these analyses more strongly support a relationship to the diatom plastids than the analyses using genes currently residing in the dinoflagellate plastids. Moreover, the divergence of dinoflagellate nuclear-encoded genes is proportionally smaller compared to those of other species than is the divergence of dinoflagellate plastid-encoded genes, suggesting that the phylogeny obtained from the former is less likely to be influenced by convergent evolution that could lead to long branch attraction (Felsenstein 1978). It was somewhat surprising, given the proportionally shorter branch lengths for the nuclear-encoded dinoflagellate sequences, that the chlorophyll *c* plastids were not monophyletic in either the ML or the Bayesian trees (Fig. 2). It is still possible that long branch attraction pulled the clade of the dinoflagellates and stramenopiles to the base of the rhodophyte clade, resulting in a paraphyletic chlorophyll *c* group. However, it seems unlikely that the dinoflagellate and stramenopile clade is itself an artifact of long branch attraction, as it is also found in the Bayesian tree using plastid-encoded sequences (Fig. 1).

The hypothesis of a single evolutionary origin of the red algal-type plastids predicts a close relationship between apicomplexan and dinoflagellate plastid sequences, similar to the relationship obtained with bona fide nuclear genes (Harper et al. 2005; Van de Peer and De Wachter 1997). However, the provenance of some nuclear genes that show the relationship between apicomplexans and dinoflagellates, such as the nuclear-encoded plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (Fast et al. 2001), is not always clear, as their evolutionary origins may be obscured by lateral gene transfer (Fagan et al. 1998). Thus these analyses may simply be mirroring the phylogeny of other nuclear-encoded genes. In contrast, the ribosomal genes tested in our analyses (Fig. 3) are more likely to be derived from the plastid, as they are retained by the plastid genome in all other species examined. This is certainly the case for the three other apicomplexans whose plastid sequences contain the same ribosomal sequences (Fig. 4). Although the phylogenies need to be interpreted cautiously because of the high sequence divergence of apicomplexans, they do not show evidence of a tertiary endosymbiotic event. Therefore, it is most parsimonious to consider the dinoflagellate plastid as having the same endosymbiont origin as all other chlorophyll c-containing plastids.

The discovery of secondary plastids has opened a window into a fascinating world where transfer of organelles through endosymbiosis has been combined with movement of genetic material through lateral gene transfer to generate genetic diversity. The dissection of the different elements involved in this process requires a large number of different approaches that can be brought to bear on the problem. In the analyses performed here, we have attempted to tease out the evolutionary threads leading to the peridinin-containing dinoflagellate plastids by exploiting gene sequences still retained in other plastid genomes. It will be of interest to determine if any ciliate sequences can be obtained that will further reinforce the conclusion that red algal secondary plastids have all had a common origin.

References

- Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105
- Bachvaroff TR, Concepcion GT, Rogers CR, Herman EM, Delwiche CF (2004) Dinoflagellate expressed sequence tag data indicate massive transfer of chloroplast genes to the nuclear genome. *Protist* 155:65–78
- Bachvaroff TR, Sanchez Puerta MV, Delwiche CF (2005) Chlorophyll c-containing plastid relationships based on analyses of a multigene data set with all four chromalveolate lineages. *Mol Biol Evol* 22:1772–1782
- Bachvaroff TR, Sanchez-Puerta MV, Delwiche CF (2006) Rate variation as a function of gene origin in plastid-derived genes of peridinin-containing dinoflagellates. *J Mol Evol* 62:42–52
- Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290:972–977
- Chesnick JM, Kooistra WH, Wellbrock U, Medlin LK (1997) Ribosomal RNA analysis indicates a benthic pennate diatom ancestry for the endosymbionts of the dinoflagellates *Peridinium foliaceum* and *Peridinium balticum* (Pyrrhophyta). *J Eukaryot Microbiol* 44:314–320
- Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet* 6:361–375
- Delwiche CF (1999) Tracing the thread of plastid diversity through the tapestry of life. *Am Nat* 154:S164–S177
- Delwiche C, Palmer J (1996) Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plasmids. *Mol Biol Evol* 13:873–882
- Durnford DG, Deane JA, Tan S, McFadden GI, Gantt E, Green BR (1999) A phylogenetic assessment of the eukaryotic light-harvesting antenna proteins, with implications for plastid evolution. *J Mol Evol* 48:59–68
- Fagan T, Hastings JW, Morse D (1998) The phylogeny of glyceraldehyde-3-phosphate dehydrogenase indicates lateral gene transfer from cryptomonads to dinoflagellates. *J Mol Evol* 47:633–639
- Fast NM, Kissinger JC, Roos DS, Keeling PJ (2001) Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol Biol Evol* 18:418–426
- Fast NM, Xue L, Bingham S, Keeling PJ (2002) Re-examining alveolate evolution using multiple protein molecular phylogenies. *J Eukaryot Microbiol* 49:30–37
- Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool* 27:401–410
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791
- Felsenstein J (2004) PHYLIP: phylogeny inference program, version 3.65. University of Washington, Seattle
- Foth BJ, Ralph SA, Tonkin CJ, Struck NS, Fraunholz M, Roos DS, Cowman AF, McFadden GI (2003) Dissecting apicoplast targeting in the malaria parasite *Plasmodium falciparum*. *Science* 299:705–708
- Funes S, Davidson E, Reyes-Prieto A, Magallon S, Herion P, King MP, Gonzalez-Halphen D (2002) A green algal apicoplast ancestor. *Science* 298:2155
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704
- Hackett JD, Maranda L, Yoon HS, Bhattacharya D (2003) Phylogenetic evidence for the cryptophyte origin of the plastid of *Dinophysis*. *J Phycol* 39:440–448
- Hackett JD, Yoon HS, Soares MB, Bonaldo MF, Casavant TL, Scheetz TE, Nosenko T, Bhattacharya D (2004) Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. *Curr Biol* 14:213–218
- Harper JT, Keeling PJ (2003) Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. *Mol Biol Evol* 20:1730–1735
- Harper JT, Waanders E, Keeling PJ (2005) On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. *Int J Syst Evol Microbiol* 55:487–496
- Ishida K, Green BR (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

- Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8:275–282
- Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J Mol Evol* 29:170–179
- Kohler S, Delwiche CF, Denny PW, Tilney LG, Webster P, Wilson RJ, Palmer JD, Roos DS (1997) A plastid of probable green algal origin in Apicomplexan parasites. *Science* 275:1485–1489
- Koumandou VL, Nisbet RE, Barbrook AC, Howe CJ (2004) Dinoflagellate chloroplasts—Where have all the genes gone? *Trends Genet* 20:261–267
- Martin W, Stoebe B, Goremykin V, Hapsmann S, Hasegawa M, Kowallik KV (1998) Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393:162–165
- McEwan ML, Keeling PJ (2004) HSP90, tubulin and actin are retained in the tertiary endosymbiont genome of *Kryptoperidinium foliaceum*. *J Eukaryot Microbiol* 51:651–659
- McFadden GI, Reith ME, Munholland J, Lang-Unnasch N (1996) Plastid in human parasites. *Nature* 381:482
- Nassoury N, Cappadocia M, Morse D (2003) Plastid ultrastructure defines the protein import pathway in dinoflagellates. *J Cell Sci* 116:2867–2874
- Palmer JD (2003) The symbiotic birth and spread of plastids: How many times and whodunit? *J Phycol* 39:4–11
- Patron NJ, Waller RF, Archibald JM, Keeling PJ (2005) Complex protein targeting to dinoflagellate plastids. *J Mol Biol* 348:1015–1024
- Patron NJ, Waller RF, Keeling PJ (2006) A tertiary plastid uses genes from two endosymbionts. *J Mol Biol* 357:1373–1382
- Petersen J, Teich R, Brinkmann H, Cerff R (2006) A “green” phosphoribulokinase in complex algae with red plastids: evidence for a single secondary endosymbiosis leading to haptophytes, cryptophytes, heterokonts, and dinoflagellates. *J Mol Evol* 62:143–157
- Rambaut A, Drummond AJ (2005) Tracer: MCMC trace analysis tool, version 1.3, University of Oxford, UK
- Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W, Bohnert HJ, Philippe H, Lang BF (2005) Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr Biol* 15:1325–1330
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574
- Saldarriaga JF, Taylor FJ, Keeling PJ, Cavalier-Smith T (2001) Dinoflagellate nuclear SSU rRNA phylogeny suggests multiple plastid losses and replacements. *J Mol Evol* 53:204–213
- Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihood with applications to phylogenetic inference. *Mol Biol Evol* 16:1114–1116
- Tengs T, Dahlberg OJ, Shalchian-Tabrizi K, Klaveness D, Rudi K, Delwiche CF, Jakobsen KS (2000) Phylogenetic analyses indicate that the 19′hexanoyloxy-fucoanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Mol Biol Evol* 17:718–729
- Van de Peer Y, De Wachter R (1997) Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site rate variation in 18S rRNA. *J Mol Evol* 45:619–630
- Waller RF, McFadden GI (2005) The apicoplast: a review of the derived plastid of apicomplexan parasites. *Curr Issues Mol Biol* 7:57–79
- Waller RF, Keeling PJ, van Dooren GG, McFadden GI (2003) Comment on “A Green Algal Apicoplast Ancestor”. *Science* 301:49; author reply 49
- Waller RF, Patron NJ, Keeling PJ (2006) Phylogenetic history of plastid-targeted proteins in the peridinin-containing dinoflagellate *Heterocapsa triquetra*. *Int J Syst Evol Microbiol* 56:1439–1447
- Wang Y, Morse D (2006) Rampant polyuridylylation of plastid gene transcripts in the dinoflagellate *Lingulodinium*. *Nucleic Acids Res* 34:613–619
- Watanabe M, Tsutomu S, Inouye I, Suda S, Sawaguchi T, Chihara M (1987) A green dinoflagellate with chlorophylls a and b: morphology, fine structure of the chloroplast and chlorophyll composition. *J Phycol* 23:382–389
- Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 18:691–699
- Wolters J (1991) The troublesome parasites—molecular and morphological evidence that Apicomplexa belong to the dinoflagellate-ciliate clade. *Biosystems* 25:75–83
- Yoon HS, Hackett JD, Pinto G, Bhattacharya D (2002) The single, ancient origin of chromist plastids. *Proc Natl Acad Sci USA* 99:15507–15512
- Yoon HS, Hackett JD, Van Dolah FM, Nosenko T, Lidie KL, Bhattacharya D (2005) Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. *Mol Biol Evol* 22:1299–1308
- Zhang Z, Green BR, Cavalier-Smith T (1999) Single gene circles in dinoflagellate chloroplast genomes. *Nature* 400:155–159