Nuclear-Encoded, Plastid-Targeted Genes Suggest a Single Common Origin for Apicomplexan and Dinoflagellate Plastids

Naomi M. Fast,* Jessica C. Kissinger,† David S. Roos,† and Patrick J. Keeling*

*Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada; and †Department of Biology, University of Pennsylvania

The phylum Apicomplexa encompasses a large number of intracellular protozoan parasites, including the causative agents of malaria (Plasmodium), toxoplasmosis (Toxoplasma), and many other human and animal diseases. Apicomplexa have recently been found to contain a relic, nonphotosynthetic plastid that has attracted considerable interest as a possible target for therapeutics. This plastid is known to have been acquired by secondary endosymbiosis, but when this occurred and from which type of alga it was acquired remain uncertain. Based on the molecular phylogeny of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) genes, we provide evidence that the apicomplexan plastid is homologous to plastids found in dinoflagellates—close relatives of apicomplexa that contain secondary plastids of red algal origin. Surprisingly, apicomplexan and dinoflagellate plastid-targeted GAPDH sequences were also found to be closely related to the plastid-targeted GAPDH genes of heterokonts and cryptomonads, two other groups that contain secondary plastids of red algal origin. These results address several outstanding issues: (1) apicomplexan and dinoflagellate plastids appear to be the result of a single endosymbiotic event which occurred relatively early in eukaryotic evolution, also giving rise to the plastids of heterokonts and perhaps cryptomonads; (2) apicomplexan plastids are derived from a red algal ancestor; and (3) the ancestral state of apicomplexan parasites was photosynthetic.

Introduction

Apicomplexa are obligate intracellular parasites that are responsible for a number of serious diseases affecting a broad range of animal hosts, including humans. Malaria, which is among the 10 most deadly infectious diseases today, is caused by the apicomplexan *Plasmodium falciparum.* Other common apicomplexan infections include toxoplasmosis, cryptosporidiosis, and Texas cattle fever (babesiosis) (see Levine [1988] for a general overview). Apicomplexa are members of the Alveolata (or alveolates), a group of organisms that also includes dinoflagellates and ciliates (Van de Peer and De Wachter 1997).

In recent years, apicomplexa have been found to harbor a relic, nonphotosynthetic plastid homologous to the chloroplasts of plants and algae. This plastid, or apicoplast, was first identified in Plasmodium and Toxoplasma (McFadden et al. 1996; Wilson et al. 1996; Köhler et al. 1997) but now appears to be widespread throughout the phylum (Denny et al. 1998; Lang-Unnasch et al. 1998). Ultrastructural and molecular data have confirmed that this organelle originated by secondary endosymbiosis, whereby a heterotrophic eukaryote engulfed and retained a photosynthetic eukaryote, resulting in a plastid with more than two bounding membranes (Waller et al. 1998). Although the apicomplexan plastid has been postulated to be surrounded by three membranes (Hopkins et al. 1999), other studies show it to be surrounded by four membranes (Köhler et al.

1997): two inner membranes corresponding to the plastid membranes of the primary endosymbiont, a third membrane which is thought to be the plasma membrane of the secondary endosymbiont, and an outermost membrane which is homologous to the host endomembrane system. Protein trafficking to secondary plastids is complicated by these extra membranes, so proteins encoded by genes in the host nucleus that are targeted to the secondary plastid require a signal peptide to direct the product to the endomembrane system and a transit peptide to direct the protein into the plastid, unlike plastid-targeted products in plants, which require only a transit peptide (Delwiche 1999; McFadden 1999). Nuclear-encoded, plastid-targeted genes from both Toxoplasma and Plasmodium have been found to contain such bipartite leader sequences, and targeting studies with green fluorescent protein in Toxoplasma and Plasmodium also indicate that the bipartite leader is necessary and sufficient to target proteins to the plastid (Waller et al. 1998, 2000).

The presence of a plastid in apicomplexa has attracted considerable attention, primarily for two reasons. First, the plastid is a remarkable target for therapeutics, since plastid metabolic pathways are prokaryotic in origin, allowing herbicides and antibiotics to specifically disrupt the parasite with minimal effect on the animal host (Jomaa et al. 1999; McFadden and Roos 1999; Zuther et al. 1999). Second, the origin of this plastid has been something of an evolutionary enigma, since it is unclear why obligate intracellular parasites would acquire or retain a plastid when they are clearly not photosynthetic. Arguments based on gene order in plastid genomes and on some plastid gene phylogenies have been used to support a red algal origin for the apicomplexan plastid, but other phylogenies (based on the translation factor *tuf*A) have alternatively been used to support the notion that the secondary symbiont was a green alga (Williamson et al. 1994; Köhler et al. 1997;

McFadden and Waller 1997). To complicate matters further, the apicomplexa are now well known to share a very close relationship with dinoflagellates, and dinoflagellates also contain a plastid (here we refer specifically to the peridinin-containing plastid of dinoflagellates) derived from secondary endosymbiosis in this case involving a red alga (Zhang, Green, and Cavalier-Smith 1999). The apicomplexan-dinoflagellate relationship has raised questions as to whether the apicoplast shares a common origin with the dinoflagellate plastid; however, a solution to this problem has proved difficult, since plastid gene sequences from both groups are extremely divergent and not phylogenetically informative (Zhang, Green, and Cavalier-Smith 2000).

We addressed the evolutionary origin of the apicomplexan plastid using a novel approach. In both groups, nuclear-encoded, plastid-targeted gene sequences are generally less divergent than plastid-encoded sequences and are therefore more amenable to phylogenetic analysis. We accordingly characterized the first comparable plastid-targeted genes from apicomplexa and dinoflagellates, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), and used these genes to test the relationship between their plastids. We sequenced cytosolic copies of GAPDH from the apicomplexan *Toxoplasma gondii,* the ciliates *Tetrahymena thermophila, Paramecium tetraurelia, Halteria grandinella,* and *Blepharisma intermedium,* and the heterokont alga *Heterosigma akashiwo.* Plastid-targeted GAPDH sequences were also determined from *T. gondii* and *H. akashiwo.*

GAPDH is a central metabolic enzyme involved in both glycolysis and the Calvin cycle. Phylogenetic analyses of GAPDH sequences have proven to be notoriously complex, reflecting a complicated history of gene duplications, lateral transfers, and endosymbiotic gene replacements (e.g., Martin et al. 1993; Henze et al. 1995; Liaud et al. 1997; Figge et al. 1999). Previously characterized plastid-targeted GAPDH genes from a dinoflagellate were found to be most closely related to cryptomonad plastid-targeted genes, but together these plastid genes are clearly related to eukaryotic cytosolic genes (Fagan, Hastings, and Morse 1998). In contrast, the plastid GAPDH genes from plants, green algae, and red algae are closely related to cyanobacteria and other eubacteria, as one would expect of a plastid-derived gene, suggesting that the dinoflagellate and cryptomonad plastid-targeted GAPDH genes originated by lateral gene transfer or endosymbiotic gene replacement (Liaud et al. 1997; Fagan, Hastings, and Morse 1998). Because of this unusual evolutionary history, we can test whether the apicoplast and dinoflagellate plastids are related simply by determining whether the apicomplexan plastid-targeted GAPDH is specifically related to the peculiar dinoflagellate gene or to the cyanobacteria-plastid lineage.

Phylogenetic trees constructed with our new sequences from apicomplexa and their more distant relatives ciliates and heterokonts show that the apicomplexan plastid-targeted GAPDH gene is indeed related to dinoflagellate plastid-targeted GAPDH, supporting the conclusion that their plastids arose from a single sec-ondary endosymbiotic event involving a red alga. In addition, both genes are part of a very highly supported clade that also includes plastid-targeted GAPDH from heterokonts and cryptomonads, two other groups with red algal secondary endosymbionts, altogether suggesting that the plastids of all these groups are homologous and arose from a single engulfment of a red alga. This evidence confirms the red algal origin of the apicoplast and shows that the apicoplast is related to the dinoflagellate plastid. Moreover, it also suggests that this endosymbiosis took place relatively early in eukaryotic evolution, leading to the provocative possibility that ciliates at one time also harbored a plastid.

Materials and Methods
New GAPDH Sequences

An examination of the *T. gondii* EST database (http://www.cbil.upenn.edu/ParaDBs/Toxoplasma/index.html) identified several clones for GAPDH. Clones TgZy46f03.r1, TgZy74f06.r1, and TgZy85b05.r1 were ordered from Genome Systems, and full-length double-stranded sequence was obtained from each (ABI BigDye chemistry). Clone TgZy46f03.r1 contained the complete *T. gondii* cytosolic GAPDH sequence. The remaining two clones overlapped and provided a nearly complete second GAPDH gene. In an attempt to clone the missing N-terminus of the second GAPDH gene, 5′ RACE cDNA pools were prepared (Clontech Smart RACE) from total and polyA purified *T. gondii* RNA (RH strain) using nested reverse primers rtGSP1 (5′-GGCC-GCGGCTTCCGTGGTGCAGAAG) and rtNGSP1 (5′-GAAGACGCCAGTGGATTCGCAGACG). Products were cloned into the TOPO-TA vector pCR2.1 (Invitrogen). In addition, cDNA libraries were screened with the 5′ end of the GAPDH cDNA fragment isolated from clone TgZy74f06.r1. Five clones were isolated; three showed sequence extensions and all were sequenced on both strands. Results from both 5′ RACE and the additional library screening indicated that this second GAPDH possesses an apparent N-terminal extension of >700 nt, as no in-frame methionine has yet been encountered.

Near-full-length GAPDH sequences were amplified from *H. akashiwo, B. intermedium, P. tetraurelia* (strain 51.s), and *H. grandinella* by PCR using the degenerate primers 5′-CCAAGGTCGGNATHAAYGGNTTYGG and 5′-CGAGTAGCCCCAYTCRTTRTCRTACCA. All amplification reactions consisted of 35 cycles of 92°C for 12 s, 45°C for 12 s, and 72°C for 1 min with a slope of 3.0 in an Idaho Rapidcycler. Products were cloned into the TOPO-TA vector pCR2.1 (Invitrogen), and multiple clones of each were sequenced on both strands with ABI BigDye terminator chemistry. *Heterosigma akashiwo, B. intermedium, P. tetraurelia,* and *H. grandinella* DNAs were generous gifts from K. Ishida, J. Berger, J. Preer, and D. Lynn, respectively. The *T. thermophila* GAPDH sequence was amplified from an excised UniZap cDNA library generously provided by A. Turkewitz. The 5′ end of the *T. thermophila* GAPDH was amplified from the same cDNA preparation using the

M13 Reverse primer in conjunction with an exact-match primer designed to the 5′ end of the *T. thermophila* GAPDH.

New sequences have been deposited in GenBank under accession numbers AF319448–AF319456, AF265361, and AF265362.

Phylogenetic Analysis

Amino acid alignments were created with PIMA (Smith and Smith 1992), and the alignment was edited manually. Distances were calculated from 290 unambiguously aligned characters by PUZZLE, version 4.0.1 (Strimmer and von Haeseler 1996), using the JTT substitution matrix with the frequency of amino acid usage calculated from the data. Distances were calculated assuming both a constant rate of substitution and modeling site-to-site variation on a gamma curve with eight rate categories, estimating invariable sites and the shape parameter from the data. Trees were constructed with BioNJ (Gascuel 1997) and Fitch-Margoliash (Felsenstein 1993), with the latter using global rearrangements and 10 input order jumbles. The same well-supported groups were seen in analyses with both constant and gamma-corrected rates, in both BioNJ and Fitch-Margoliash trees. All trees shown are BioNJ trees. Bootstrap trees were constructed using both BioNJ and Fitch-Margoliash (with no global rearrangements or input order jumbles) from 100 resampled data sets, with gamma-corrected distances (with the rate category parameters above) calculated using Puzzleboot (shell script by M. Holder and A. Roger).

Kishino-Hasegawa tests (Kishino and Hasegawa 1989) to assess alternative tree topologies were performed using PUZZLE, version 4.0.1. Rate category parameters were as above, and alternatives were deemed significantly worse at a 5% or 1% level.

Results
New GAPDH Gene Sequences

The Toxoplasma expressed sequence tag (EST) sequencing project identified two distinct GAPDH genes. These cDNA clones were isolated, and double-stranded sequences were obtained from both. One clone contained a clear 5′ untranslated region (UTR) followed by a putative start codon at a position corresponding to the amino terminus of other mature GAPDH genes. In contrast, the second clone possessed an N-terminal extension upstream of the mature GAPDH-coding region but was truncated and contained no initiator codon. In an attempt to identify the beginning of this transcript, 5′ RACE was carried out; however, the leader appears to be extremely long, as no methionine start codon could be identified in over 700 bp of leader sequence. Leader sequences are characteristic of plastid-targeted genes from Toxoplasma and generally encode long signal and transit peptides (Waller et al. 1998).

As ciliates are members of the alveolates, and therefore close relatives of apicomplexa and dinoflagellates (Van de Peer and De Wachter 1997), GAPDH gene sequences were also amplified from the ciliates Halteria,

Blepharisma, Paramecium, and Tetrahymena. The ciliate alternative genetic code (Caron and Meyer 1985; Preer et al. 1985) is evident by the presence of in-frame TAR codons in the GAPDH genes of Halteria, Paramecium, and Tetrahymena. Paramecium genes are also typified by short spliceosomal introns (Russell, Fraga, and Hinrichsen 1994), and 23–27-bp introns were found in the GAPDH sequences of Paramecium. The Paramecium GAPDH sequence is very similar to a short (279 bp) sequence fragment in the database, and the *T. thermophila* gene sequence is very similar to the partial (515 bp) *Tetrahymena pyriformis* sequence in GenBank (Hafid et al. 1998) (two other *T. pyriformis* GAPDH fragments in the database appear to arise from bacterial contaminants).

To determine if the ciliate GAPDH genes possessed a leader sequence, the 5′ end of the GAPDH gene was amplified from *T. thermophila* cDNA. Several different cDNA ends from the same gene were sequenced and none encoded a leader; in all cases, transcripts contained a 50–60-bp 5′ UTR followed by a methionine codon corresponding to the amino terminus of the mature GAPDH coding sequence. These data are congruent with previous analysis of a different Tetrahymena GAPDH cDNA that is described in the literature but is not available in GenBank (Zhao et al. 1997).

Molecular data indicate that heterokonts are also potential relatives of alveolates (Van de Peer and De Wachter 1997; Tengs et al. 2000), and since heterokonts also contain secondary endosymbiotic plastids, GAPDH was amplified from the genomic DNA of *H. akashiwo,* resulting in two different GAPDH sequences. Comparing these sequences with those available in the database indicated that each was similar to either the plastid-targeted GAPDH or the cytosolic GAPDH of diatoms, which were reported during the course of this work (see Liaud et al. [2000], which also reports the presence of a mitochondrion-targeted TPI-GAPDH fusion protein in heterokonts that is not related to the GAPDH described here).

Phylogenetic Analysis

New GAPDH sequences were aligned with available GAPDH sequences in GenBank, and a global phylogenetic analysis was undertaken based on a data set of 290 unambiguously aligned characters. The resulting tree (fig. 1) recovered the relationships typical of GAPDH phylogeny; eukaryotic cytosolic GAPDH (GapC) is clearly distinct from the main lineage of eubacterial GAPDH (GapA/B), which includes the plastid-targeted GAPDH genes of land plants, green algae, and red algae. The apicomplexan plastid-targeted GAPDH does not branch with these plastid sequences, but instead branches with eukaryotic GapC genes, specifically in the clade containing the dinoflagellate, heterokont, and cryptomonad plastid-targeted GAPDH (supported at 100% by bootstrap analysis). The inclusion of these genes in the eukaryotic GapC clade at the exclusion of the "typical" plastid-targeted GapA/B genes is very strongly supported by bootstrap analysis (at several
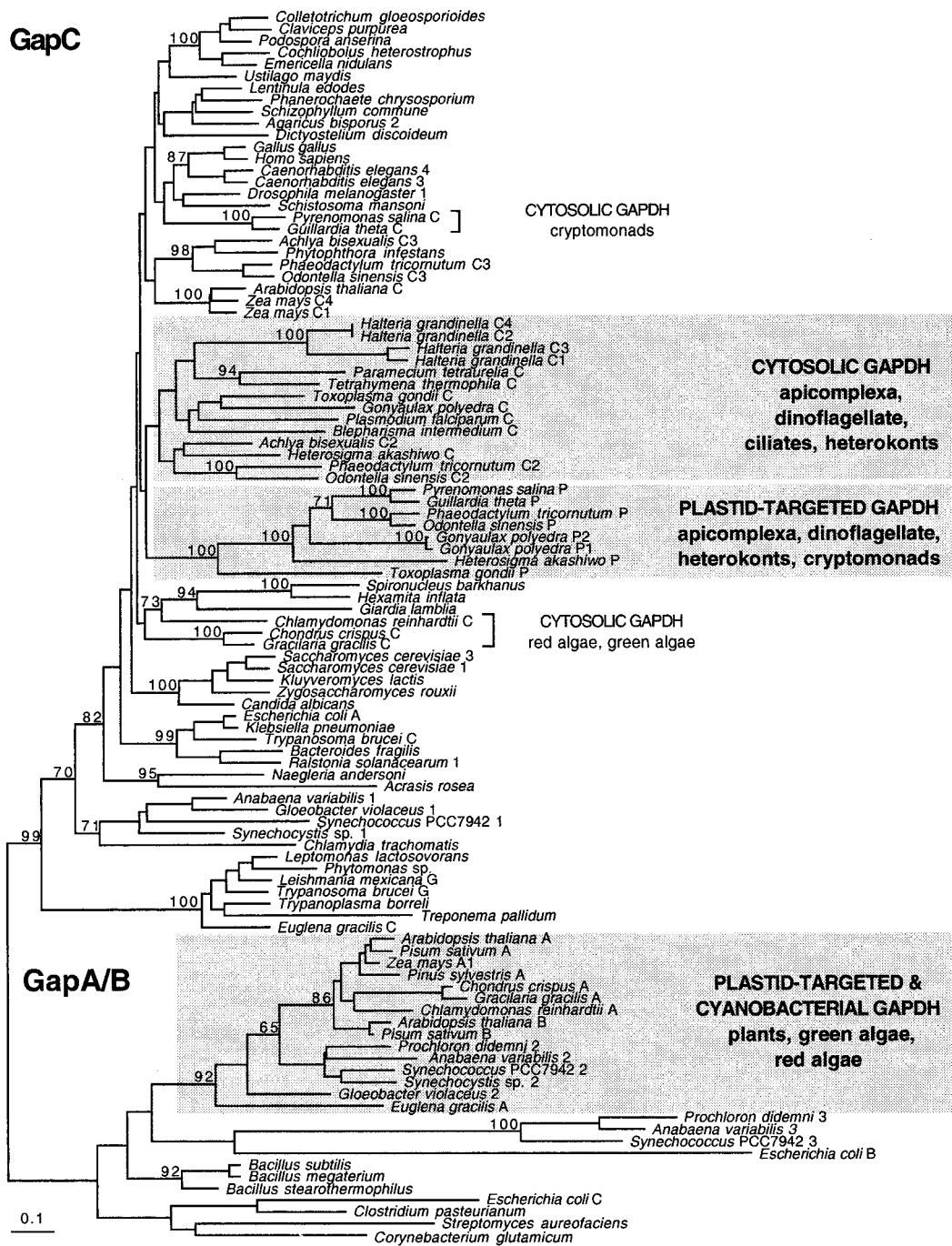
Fig. 1.—Global GAPDH phylogeny including eukaryotic cytosolic (GapC), as well as eubacterial (GapA/B), homologs. Numbers at nodes indicate neighbor-joining bootstrap support >50% for major nodes. Plastid-targeted GAPDH from land plants, green algae, and red algae branch in the eubacterial GapA/B group with cyanobacterial Gap2s, which are the paralogs most closely related to the plastid lineage. In contrast, the apicomplexan, dinoflagellate, heterokont, and cryptomonad plastid-targeted GAPDH genes branch in the eukaryotic cytosolic GapC group, specifically related to the cytosolic homologs from these same organisms, except cryptomonads. Cytosolic genes from green algae, red algae, and cryptomonads are also indicated in the figure.

highly supported nodes), altogether suggesting that the apicomplexan plastid-targeted GAPDH is indeed derived from the same endosymbiotic event as that of dinoflagellates. Interestingly, these plastid-targeted GAPDH genes are most closely related to the cytosolic GAPDH genes from these same organisms (with the exclusion of cryptomonads), which is consistent with the notion that the plastid-targeted GAPDH originated through an endosymbiotic gene replacement involving a duplication of the cytosolic homolog. Although it is a formal possibility that the plastid-targeted GAPDH could have arisen from the algal nuclear lineage (i.e., a transfer of the nuclear GAPDH of the secondary symbiont), this is not supported by the phylogeny (see the
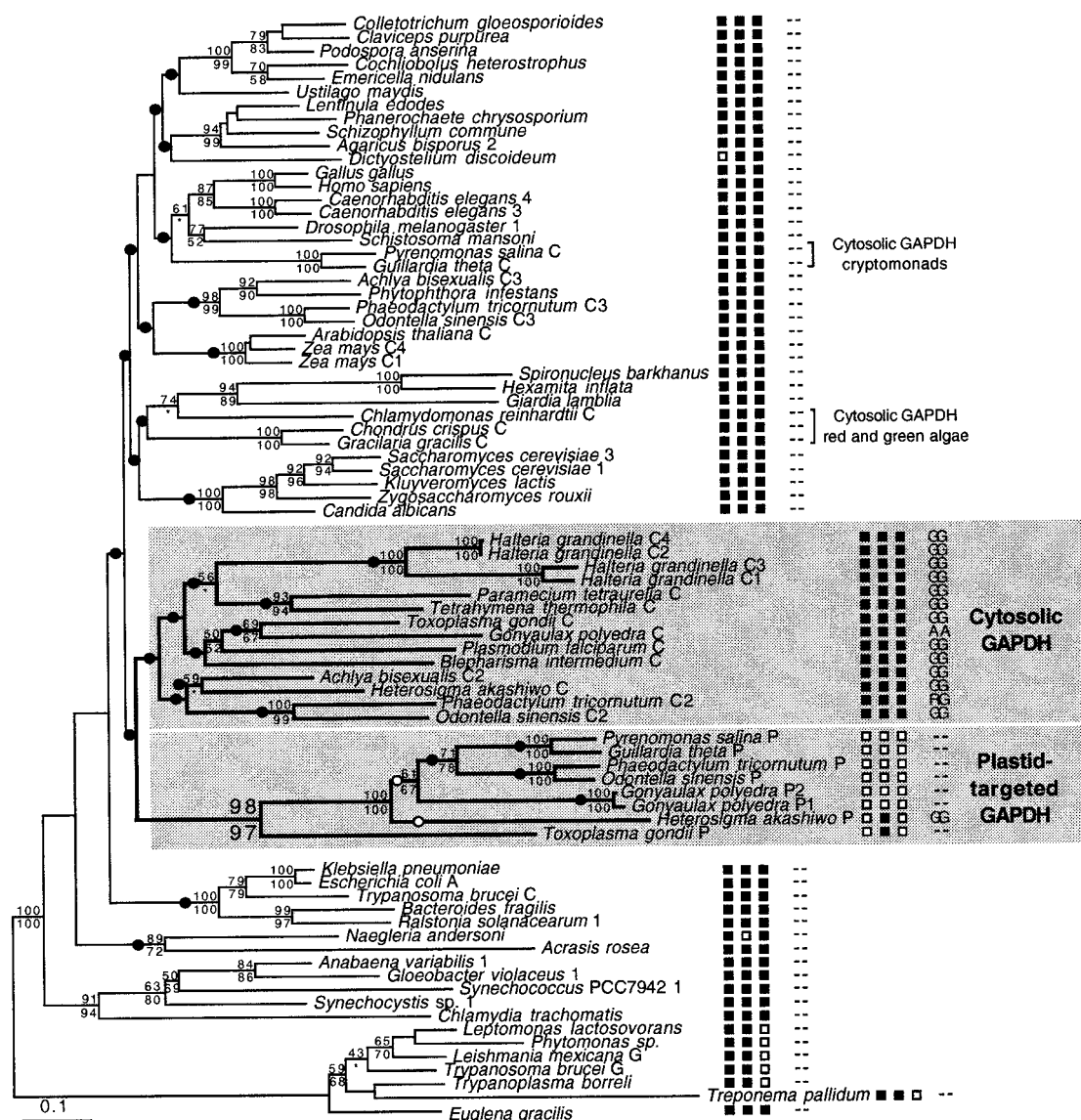
Fig. 2.—Phylogeny of GapC including new plastid-targeted and cytosolic genes. Numbers at nodes correspond to bootstrap support >50% for major nodes from neighbor joining (above node) and Fitch-Margoliash (below node). Asterisks at nodes indicate bootstrap support <50%. Circles indicate positions at which the Toxoplasma plastid-targeted gene was constrained in Kishino-Hasegawa tests: filled circles indicate positions that were rejected at 5%, while open circles indicate positions that were not rejected. Boxes adjacent to taxon names represent three residues (D32, G187, and P188) that are highly conserved for NAD-specific GAPDH (filled box) or substituted for NAD/NADP-specific GAPDH (open boxes). Also shown is the distribution of a two-amino-acid insertion (typically GG) found in the cytosolic GAPDHs of alveolates and heterokonts and the Heterosigma plastid-targeted gene. Cytosolic and plastid-targeted GAPDH genes are shaded and labeled; genes from green algae, red algae, and cryptomonads are also labeled.

positions of the algal cytosolic sequences in figs. 1 and 2).

To allow a more comprehensive analysis of the apicomplexan and dinoflagellate genes and their relationship to heterokont and cryptomonad genes, we focused on the eukaryotic cytosolic (GapC) clade. Phylogenies were constructed using GapC; plastid-targeted GAPDH sequences from alveolates, heterokonts, and cryptomonads; and a handful of the closely related bacterial sequences (fig. 2). As with other GAPDH phylogenies, some eukaryotic groups were recovered with strong support, but overall the backbone of this tree was only weakly supported, so the relationships among various

groups could not be readily resolved. One striking exception was the highly supported node uniting the plastid-targeted GAPDH sequences from apicomplexa, dinoflagellates, heterokonts, and cryptomonads. The bootstrap support for this clade was very strong, although the relationships within the clade were not well resolved and did not reflect well-documented relationships (e.g., the heterokonts did not branch together). The inclusion of the apicomplexan plastid-targeted sequence within this clade was further examined with Kishino-Hasegawa tests (Kishino and Hasegawa 1989). The Toxoplasma sequence was moved to 32 alternative positions, including other positions within the plastid-targeted clade,

within the cytosolic clade, and with other eukaryotic groups. The best tree was that shown in figure 2, and only positions at which the Toxoplasma sequence was either specifically related to the Heterosigma plastid GAPDH or sister to the dinoflagellate plastid GAPDHs were not rejected at a 5% confidence level. Indeed, almost all placements of the apicomplexan plastid-targeted sequence with cytosolic sequences were rejected at a <1% level, indicating that its placement within the plastid clade was robust.

Support for the inclusion of the Heterosigma sequence within the plastid-targeted clade was also tested, since this was a PCR product and therefore lacked any targeting information. Again, the best tree found was that in figure 2, and every position outside of the plastid clade was rejected at the 5% level (although any position within the plastid-targeted clade was not found to be significantly worse). In addition, a plastid role for this Heterosigma GAPDH was evident from its amino acid sequence. There are three amino acids that generally indicate substrate specificity for GAPDH: cytosolic GAPDH utilizes NAD and possesses conserved D32, G187, and P188 residues, whereas plastid GAPDH sequences have substitutions at these residues and can utilize both NADP and NAD (residues numbered following Clermont et al. 1993). The Heterosigma GAPDH that branches in the plastid-targeted clade has just such substitutions at D32 and P188, as does the plastid-targeted Toxoplasma sequence (open and filled boxes in fig. 2).

In contrast to the strong support for the plastid-targeted clade, there is little bootstrap support for the node uniting the alveolate and heterokont cytosolic GAPDH sequences shown in figure 2; nor is there support for the internal relationships within this clade. However, this clade is supported by the presence of a two-amino-acid insertion which is unique to this group (fig. 2), with the single exception of the Heterosigma plastid homolog, which appears to be the result of a recombination event between the plastid and cytosolic paralogs in Heterosigma. There is no other evidence of recombination from examining the alignment, but this is not unexpected given the presumed ancient nature of this event. Also in contrast to the plastid-targeted clade, the cytosolic clade does not include the cryptomonad cytosolic GAPDHs, which branch with animals in figures 1 and 2. The relationship of the cryptomonad and dinoflagellate plastid-targeted GAPDH genes has previously been attributed to lateral transfer (Fagan, Hastings, and Morse 1998), although a relationship between cryptomonads, alveolates, and heterokonts has also been proposed (Cavalier-Smith 1999; Liaud et al. 2000). Kishino-Hasegawa tests were used to compare 12 alternative placements of the cryptomonad cytosolic GAPDHs, and these tests showed that their exact position among GapCs was not clear, although their placement with the cytosolic GAPDHs from alveolates and heterokonts was rejected at the 5% level. Clearly, the plastid-targeted and cytosolic homologs of cryptomonads are telling conflicting stories, but presently it is not possible to discern which of the two genes, if either, was involved in a transfer.

Discussion

Targeting sequence information and the presence of more than two membranes around the apicomplexan plastid indicate a secondary origin, but several important questions regarding the origin of the apicomplexan plastid remain unanswered despite vigorous debate. In particular, no strong evidence has been forthcoming to distinguish whether the secondary symbiont was a red or a green alga, or whether the dinoflagellate plastid is related to the apicoplast. Since only a few plastid genes have been characterized from both apicomplexa and dinoflagellates, and these are uniformly poor phylogenetic markers, we turned to the more conserved nuclear-encoded, plastid-targeted GAPDH gene.

Previously characterized dinoflagellate plastid-targeted GAPDH genes are not specifically related to the plastid GAPDH genes of plants, green algae, or red algae, which are all of cyanobacterial origin, but are instead derived from eukaryotic cytosolic GAPDH (Fagan, Hastings, and Morse 1998). This eukaryotic origin of the dinoflagellate plastid-targeted gene is indicative of endosymbiotic gene replacement, in this case, the plastid homolog being replaced by a nuclear, cytosolic homolog. Endosymbiotic gene replacements are not uncommon in plastid-bearing lineages (either in cases such as this or in cases in which plastid genes replace cytosolic homologs) and have affected several different genes, including plastid GAPDH genes among gymnosperms and ferns (Meyer-Gauen et al. 1994). The eukaryotic origin of the dinoflagellate plastid GAPDH allowed us to test the relationship between the apicoplast and dinoflagellate plastids by determining whether the apicomplexan plastid-targeted GAPDH shares this distinctive position with the dinoflagellates (which would show that the plastids shared a common origin) or whether it branches with the plastid sequences of plants, green algae, and red algae (which would show that the apicomplexan and dinoflagellate plastids arose independently).

In addition to characterizing plastid and cytosolic GAPDH sequences from Toxoplasma, we also sequenced plastid and cytosolic genes from the heterokont, Heterosigma, and cytosolic GAPDHs from several ciliates. Phylogenetic trees with these new sequences showed that the apicomplexan plastid-targeted sequence did indeed branch very strongly with the clade that included the dinoflagellate plastid-targeted GAPDH sequences, along with the plastid-targeted GAPDHs of heterokonts and cryptomonads, both of which also contained red algal secondary endosymbionts. Moreover, this clade was weakly but specifically related to the cytosolic GAPDH genes of apicomplexa, dinoflagellates, ciliates, and heterokonts, lending further support to the notion that the plastid-targeted GAPDH in these organisms arose by a duplication of the cytosolic gene, at least in the ancestor of heterokonts and alveolates (how the cryptomonads fit into this picture is not certain, and lateral transfer has been evoked to explain their plastid-targeted GAPDH). Altogether, the simplest explanation for this phylogeny is depicted in figure 3A. Both the

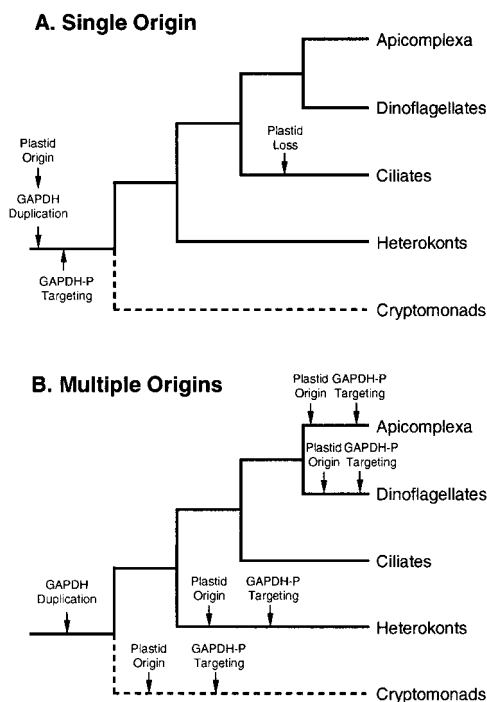**A. Single Origin**

B. Multiple Origins

FIG. 3.—Implications of two alternative explanations for GAPDH phylogeny. Cryptomonads are shown with a dashed line to indicate their uncertain role due to the position of the cryptomonad cytosolic GAPDH in the phylogeny. *A,* The plastid arose from the endosymbiotic uptake of a red alga in the ancestor of alveolates, heterokonts, and perhaps cryptomonads. The cytosolic GAPDH of the host lineage was duplicated, and one copy targeted to the plastid, replacing the original, cyanobacterial, GAPDH. Ciliates then lost their plastid. Independent losses have also subsequently occurred in individual members of each lineage. *B,* If one assumes that the apicomplexan, dinoflagellate, heterokont, and cryptomonad plastids all arose independently, then the duplication of the host cytosolic GAPDH must still have predated their divergence, but then each lineage successively and independently acquired a plastid by engulfing a red alga, and the same copy of the duplicated gene was subsequently and independently targeted to the plastid. There is no evidence for alternative B, which is also highly unlikely.

plastid-targeted and the cytosolic clades support a close relationship between alveolate and heterokont nuclear lineages (which is also supported by several other molecular phylogenies, e.g., Van de Peer and De Wachter 1997). The gene duplication that resulted in the cytosolic and plastid-targeted clades *must* have occurred prior to the divergence of alveolates and heterokonts (and perhaps cryptomonads). Since all of the plastid-targeted GAPDH genes are closely related, it follows that the plastid also originated before this divergence and that one of the duplicate GAPDH genes was targeted to the plastid. The lineages would then have diverged, and ciliates later lost their plastid. For illustrative purposes, an alternative explanation, whereby these secondary plastids arose independently, is outlined in figure 3*B*. Here, the gene duplication still has to predate the divergence of the groups; however, in this scenario, cryptomonads, heterokonts, apicomplexa, and dinoflagellates *all* have to independently undergo a secondary symbiotic event (most or all with a red alga), and *all* have to independently "choose" to target the same GAPDH paralog to

the secondary plastid. We reject this alternative not simply on the grounds of parsimony, but also on the grounds that the shear coincidence of independent symbioses with red algae and independent targeting of the *same* paralog is entirely unlikely. Similarly, the retention of both paralogs through a considerable span of evolutionary history, presumably with no function in a non–plastid-containing cell (since ciliates apparently do not retain this gene) is also extremely implausible.

In terms of the origin of the apicomplexan plastid, figure 3*A* has important implications. First, apicomplexan and dinoflagellate plastids do share a common endosymbiotic origin, which demands that the ancestral state of both dinoflagellates and apicomplexa was photosynthetic. Therefore, the basal apicomplexan and dinoflagellate lineages that seem to lack plastids (such as *Cryptosporidium* and *Perkinsus*) (Siddall et al. 1997; Zhu, Keithly, and Philippe 2000; Zhu, Marchewka, and Keithly 2000) must have a cryptic plastid or must have lost their plastids—a frequent occurrence among dinoflagellates. The second implication for the apicoplast is that it must be derived from a red alga. Based on pigmentation and plastid gene relationships, heterokont, dinoflagellate, and cryptomonad secondary plastids are all very clearly red algal (Martin et el. 1998; Douglas and Penny 1999; Durnford et al. 1999; Zhang, Green, and Cavalier-Smith 1999). The source of the apicomplexan plastid has been controversial (Williamson et al. 1994; Köhler et al. 1997; McFadden and Waller 1997; Blanchard and Hicks 1999); however, a common origin of the apicoplast with these red algal plastids settles this "red-versus-green" debate.

While these issues regarding the apicoplast are important, implications of plastid-targeted GAPDH phylogeny are also more far-reaching, since the origin of this secondary plastid likely far predates the divergence of apicomplexa and dinoflagellates. A single origin for the heterokont, alveolate, and perhaps cryptomonad plastid means that plastid loss must have occurred extensively among the heterokonts, and intriguingly suggests that ciliates have a photosynthetic ancestry. There is the provocative possibility that other evidence for a plastid still exists in ciliates, perhaps in the form of transferred genes or even an as-yet-unrecognized organelle. Regardless, it is clear that this single endosymbiotic event has had a tremendous effect on eukaryotic evolution.

Conclusions

Two outstanding controversies regarding the origin of the apicomplexan plastid are addressed by the phylogeny of GAPDH. We have shown that the apicomplexan and dinoflagellate plastids are indeed the product of a single endosymbiotic event and that this event likely took place surprisingly early in eukaryotic evolution, also giving rise to the plastids of heterokonts and possibly cryptomonads. This endosymbiosis was a pivotal event in eukaryotic history, affecting a large proportion of eukaryotic algae. This early origin of the plastid also leads to the provocative suggestion that ciliates were

ancestrally photosynthetic and may even retain a relic plastid, much as the apicomplexa do. The single origin of these plastids also reveals that the apicoplast must be derived from a red alga, since there is clear evidence that heterokont, dinoflagellate, and cryptomonad plastids are of red algal ancestry.

Acknowledgments

LITERATURE CITED

BLANCHARD, J. L., and J. S. HICKS. 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.

CARON, F., and E. MEYER. 1985. Does *Paramecium primaurelia* use a different genetic code in its macronucleus? Nature **314**:185–188.

CAVALIER-SMITH, T. 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryotic family tree. J. Eukaryot. Microbiol. **46**:347–366.

CLERMONT, S., C. CORBIER, Y. MELY, D. GERARD, A. WONACOTT, and G. BRANLANT. 1993. Determinants of coenzyme specificity in glyceraldehyde-3-phosphate dehydrogenase: role of the acidic residue in the fingerprint region of the nucleotide binding fold. Biochemistry **32**:10178–10184.

DELWICHE, C. F. 1999. Tracing the thread of plastid diversity through the tapestry of life. Am. Nat. **154**:S164–S177.

DENNY, P., P. PREISER, D. WILLIAMSON, and I. WILSON. 1998. Evidence for a single origin of the 35 kb plastid DNA in apicomplexans. Protist **149**:51–59.

DOUGLAS, S. E., and S. L. PENNY. 1999. The plastid genome of the cryptophyte alga, *Guillardia theta*: complete sequence and conserved synteny groups confirm its common ancestry with red algae. J. Mol. Evol. **48**:236–244.

DURNFORD, D. G., J. A. DEANE, S. TAN, G. I. McFADDEN, E. GANTT, and B. R. GREEN. 1999. A phylogenetic assessment of the eukaryotic light-harvesting antenna proteins, with implications for plastid evolution. J. Mol. Evol. **48**:59–68.

FAGAN, T., J. W. HASTINGS, and D. MORSE. 1998. The phylogeny of glyceraldehyde-3-phosphate dehydrogenase indicates lateral gene transfer from cryptomonads to dinoflagellates. J. Mol. Evol. **47**:633–639.

FELSENSTEIN, J. 1993. PHYLIP (phylogeny inference package). Distributed by the author, Department of Genetics, University of Washington, Seattle.

FIGGE, R. M., M. SCHUBERT, H. BRINKMANN, and R. CERFF. 1999. Glyceraldehyde-3-phosphate dehydrogenase gene diversity in eubacteria and eukaryotes: evidence for intra- and inter-kingdom gene transfer. Mol. Biol. Evol. **16**:429–440.

GASCUEL, O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol. Biol. Evol. **14**:685–695.

HAFID, N., F. VALVERDE, E. VILLALOBO, M. S. ELKEBBAJ, A. TORRES, A. SOUKRI, and A. SERRANO. 1998. Glyceraldehyde-3-phosphate dehydrogenase from *Tetrahymena pyriformis*: enzyme purification and characterization of a gapC gene with primitive eukaryotic features. Comp. Biochem. Physiol. B Biochem. Mol. Biol. **119**:493–503.

HENZE, K., A. BADR, M. WETTERN, R. CERFF, and W. MARTIN. 1995. A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. Proc. Natl. Acad. Sci. USA **92**:9122–9126.

HOPKINS, J., R. FOWLER, S. KRISHNA, I. WILSON, G. MITCHELL, and L. BANNISTER. 1999. The plastid in *Plasmodium falciparum* asexual blood stages: a three-dimensional ultrastructural analysis. Protist **150**:283–295.

JOMAA, H., J. WIESNER, S. SANDERBRAND et al. (12 co-authors). 1999. Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. Science **285**:1573–1576.

KISHINO, H., and M. HASEGAWA. 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.

KÖHLER, S., C. F. DELWICHE, P. W. DENNY, L. G. TILNEY, P. WEBSTER, R. J. WILSON, J. D. PALMER, and D. S. ROOS. 1997. A plastid of probable green algal origin in apicomplexan parasites. Science **275**:1485–1489.

LANG-UNNASCH, N., M. E. REITH, J. MUNHOLLAND, and J. R. BARTA. 1998. Plastids are widespread and ancient in parasites of the phylum Apicomplexa. Int. J. Parasitol. **28**:1743–1754.

LEVINE, N. D. 1988. The protozoan phylum Apicomplexa. CRC Press, Boca Raton, Fla.

LIAUD, M. F., U. BRANDT, M. SCHERZINGER, and R. CERFF. 1997. Evolutionary origin of cryptomonad microalgae: two novel chloroplast/cytosol-specific GAPDH genes as potential markers of ancestral endosymbiont and host cell components. J. Mol. Evol. **44**:S28–S37.

LIAUD, M. F., C. LICHTLE, K. APT, W. MARTIN, and R. CERFF. 2000. Compartment-specific isoforms of TPI and GAPDH are imported into diatom mitochondria as a fusion protein: evidence in favor of a mitochondrial origin of the eukaryotic glycolytic pathway. Mol. Biol. Evol. **17**:213–223.

McFADDEN, G. I. 1999. Plastids and protein targeting. J. Eukaryot. Microbiol. **46**:339–346.

McFADDEN, G. I., M. E. REITH, J. MUNHOLLAND, and N. LANG-UNNASCH. 1996. Plastid in human parasites. Nature **381**:482.

McFADDEN, G. I., and D. S. ROOS. 1999. Apicomplexan plastids as drug targets. Trends Microbiol. **7**:328–333.

McFADDEN, G. I., and R. F. WALLER. 1997. Plastids in parasites of humans. BioEssays **19**:1033–1040.

MARTIN, W., H. BRINKMANN, C. SAVONNA, and R. CERFF. 1993. Evidence for a chimeric nature of nuclear genomes: eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. Proc. Natl. Acad. Sci. USA **90**:8692–8696.

MARTIN, W., B. STOEBE, V. GOREMYKIN, S. HANSMANN, M. HASEGAWA, and K. V. KOWALLIK. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. Nature **393**:162–165.

MEYER-GAUEN, G., C. SCHNARRENBERGER, R. CERFF, and W. MARTIN. 1994. Molecular characterization of a novel, nuclear-encoded, NAD(+)-dependent glyceraldehyde-3-phos-

phate dehydrogenase in plastids of the gymnosperm *Pinus sylvestris* L. Plant Mol. Biol. **26**:1155–1166.

PREER, J. R. JR., L. B. PREER, B. M. RUDMAN, and A. J. BARNETT. 1985. Deviation from the universal code shown by the gene for surface protein 51A in *Paramecium.* Nature **314**:188–190.

RUSSELL, C. B., D. FRAGA, and R. D. HINRICHSEN. 1994. Extremely short 20–33 nucleotide introns are the standard length in *Paramecium tetraurelia.* Nucleic Acids Res. **22**:1221–1225.

SIDDALL, M. E., K. S. REECE, J. E. GRAVES, and E. M. BURRESON. 1997. 'Total evidence' refutes the inclusion of *Perkinsus* species in the phylum Apicomplexa. Parasitology **115**:165–176.

SMITH, R. F., and T. F. SMITH. 1992. Pattern-induced multi-sequence alignment (PIMA) algorithm employing secondary structure-dependent gap penalties for use in comparative protein modelling. Protein Eng. **5**:35–41.

STRIMMER, K., and A. VON HAESELER. 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13**:964–969.

TENGS, T., O. J. DAHLBERG, K. SHALCHIAN-TABRIZI, D. KLAVENESS, K. RUDI, C. F. DELWICHE, and K. S. JAKOBSEN. 2000. Phylogenetic analyses indicate that the 19′hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. Mol. Biol. Evol. **17**:718–729.

VAN DE PEER, Y., and R. DE WACHTER. 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, R. F., P. J. KEELING, R. G. DONALD, B. STRIEPEN, E. HANDMAN, N. LANG-UNNASCH, A. F. COWMAN, G. S. BESRA, D. S. ROOS, and G. I. MCFADDEN. 1998. Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum.* Proc. Natl. Acad. Sci. USA **95**:12352–12357.

WALLER, R. F., M. B. REED, A. F. COWMAN, and G. I. MCFADDEN. 2000. Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. EMBO J. **19**:1794–1802.

WILLIAMSON, D. H., M. J. GARDNER, P. PREISER, D. J. MOORE, K. RANGACHARI, and R. J. WILSON. 1994. The evolutionary origin of the 35 kb circular DNA of *Plasmodium falciparum*: new evidence supports a possible rhodophyte ancestry. Mol. Gen. Genet. **243**:249–252.

WILSON, R. J., P. W. DENNY, P. R. PREISER et al. (11 co-authors). 1996. Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum.* J. Mol. Biol. **261**:155–172.

ZHANG, Z., B. R. GREEN, and T. CAVALIER-SMITH. 1999. Single gene circles in dinoflagellate chloroplast genomes. Nature **400**:155–159.

———. 2000. Phylogeny of ultra-rapidly evolving dinoflagellate chloroplast genes: a possible common origin for sporozoan and dinoflagellate plastids. J. Mol. Evol. **51**:26–40.

ZHAO, Y., S. NAKASHIMA, M. ANDOH, and Y. NOZAWA. 1997. Cloning and sequencing of a cDNA encoding glyceraldehyde-3-phosphate dehydrogenase from *Tetrahymena thermophila*: growth-associated changes in its mRNA expression. J. Eukaryot. Microbiol. **44**:435–437.

ZHU, G., J. S. KEITHLY, and H. PHILIPPE. 2000. What is the phylogenetic position of *Cryptosporidium*? Int. J. Syst. Evol. Microbiol. **50**:1673–1681.

ZHU, G., M. J. MARCHEWKA, and J. S. KEITHLY. 2000. *Cryptosporidium parvum* appears to lack a plastid genome. Microbiology **146**:315–321.

ZUTHER, E., J. J. JOHNSON, R. HASELKORN, R. MCLEOD, and P. GORNICKI. 1999. Growth of *Toxoplasma gondii* is inhibited by aryloxyphenoxypropionate herbicides targeting acetyl-CoA carboxylase. Proc. Natl. Acad. Sci. USA **96**:13387–13392.