Nucleus-Encoded, Plastid-Targeted Glycerolaldehyde-3-Phosphate
Dehydrogenase (GAPDH) Indicates a Single Origin for
Chromalveolate Plastids

James T. Harper and Patrick J. Keeling

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

Plastids (the photosynthetic organelles of plants and algae) originated through endosymbiosis between a cyanobacterium
and a eukaryote and subsequently spread to other eukaryotes by secondary endosymbioses between two eukaryotes.
Mounting evidence favors a single origin for plastids of apicomplexans, cryptophytes, dinoflagellates, haptophytes, and
heterokonts (together with their nonphotosynthetic relatives, termed chromalveolates), but so far, no single molecular
marker has been described that supports this common origin. One piece of evidence comes from plastid-targeted
glycerolaldehyde-3-phosphate dehydrogenase (GAPDH), which originated by a gene duplication of the cytosolic form.
However, no plastid GAPDH has been characterized from haptophytes, leaving an important piece of the puzzle missing.
We have sequenced genes encoding cytosolic, mitochondrion-targeted, and plastid-targeted GAPDH proteins from
a number of haptophytes and heterokonts and found haptophyte homologs that branch within a strongly supported clade
of chromalveolate plastid-targeted genes, being more closely related to an apicomplexan homolog than was expected.
The evolution of plastid-targeted GAPDH supports red algal ancestry of apicomplexan plastids and raises a number of questions
about the importance of plastid loss and the possibility of cryptic plastids in nonphotosynthetic lineages such as ciliates.

Introduction

Our understanding of the evolutionary history of plastids has been significantly clarified by the recognition
that plastids have moved between distantly related eukaryotic groups by the process of secondary endosymbiosis.
Glaucophytes, green algae, plants, and red algae are the only descendants of the original endosymbiotic partnership
with a cyanobacterium. All other eukaryotic algae owe their photosynthetic nature to the capture and
retention of one of these primary algae: euglenids and chlorarachniophytes have green algal secondary plastids,
whereas heterokonts, haptophytes, cryptophytes, and dinoflagellates all possess red algal secondary endosymbionts (for review see Archibald and Keeling 2002). The recently discovered plastid in apicomplexan parasites (McFadden et al. 1996; Wilson et al. 1996) appears to be of red algal ancestry (for review see Wilson 2002), although this remains somewhat contentious (Palmer 2003).

Although our understanding of plastid diversity and evolution has improved, the number of secondary endosymbiotic events required to account for the diversity of plastids in extant algae is unknown. There are contrasting views as to whether the two green plastid lineages are related (Cavalier-Smith 1999; Archibald and Keeling 2002), and the debate over the origin of red plastids is thornier still. Although the latter share a number of common characteristics (e.g., the possession of chlorophyll c), they display a tremendous diversity of morphological, ecological, and behavioral traits. Early molecular phylogenetic studies failed to demonstrate a relationship among these groups (Helmchen, Bhattacharya, and Melkonian 1995; Daugbjerg and Andersen 1997; Oliveira and Bhattacharya 2000). One exception is the well-supported relationship between apicomplexans and dinoflagellates, which, together with ciliates, represent the alveolates (Wolters 1991; Fast et al. 2002). Even considering this relationship, the apicomplexan plastid was generally not believed to be related to that of dinoflagellates (e.g., Taylor 1999). Nevertheless, based on the idea that endosymbiotic mergers are very complex and thus extremely rare, it has been proposed that apicomplexan, cryptophyte, dinoflagellate, haptophyte, and heterokont plastids all share a common endosymbiotic origin (Cavalier-Smith 1998, 1999, 2003). These organisms and their nonphotosynthetic relatives were termed the chromalveolates.

Despite the early lack of evidence for a common origin for chromalveolates, molecular data are now beginning to reveal some affinities. Analyses of small subunit rRNA weakly support a relationship between heterokonts and alveolates (Van de Peer and De Wachter 1997; Ben Ali et al. 2001), multiple protein-coding genes weakly support a grouping of heterokonts, apicomplexans, and ciliates (Baldauf et al. 2000), and RNA polymerase II gene phylogenies give moderate support for grouping apicomplexans and heterokonts (Dacks et al. 2002).

To date, the best support for the single origin of chromalveolate plastids comes from two recent analyses of plastid-targeted and plastid-encoded genes. Glycerolaldehyde-3-phosphate dehydrogenase (GAPDH) is an enzyme involved in glycolysis, gluconeogenesis, and the Calvin cycle, whose evolution has been studied extensively (e.g., Henze et al. 1995; Liaud et al. 2000). Photosynthetic eukaryotes have two different nucleus-encoded forms of GAPDH, one cytosolic and one plastid-targeted. The plastid-targeted homologs in *Euglena*, green algae, land plants, *Pyrocystis*, and red algae are related to cyanobacterial GAPDH (Henze et al. 1995; Fagan and Hastings 2002). In contrast, the plastid-targeted GAPDH sequences in apicomplexans, cryptophytes, dinoflagellates, and...
heterokonts are closely related to eukaryotic cytosolic GAPDH (Fagan, Hastings, and Morse 1998; Liaud et al. 2000; Fast et al. 2001). This relationship has been interpreted as the result of a duplication of the cytosolic GAPDH, and the replacement of the cyanobacterial plastid-targeted GAPDH by one of these copies (Fast et al. 2001). This event is an important marker of plastid evolution in these organisms, since their common possession of the gene replacement indicates that their plastids most likely originated from a single endosymbiotic event. Similarly, a recent concatenated analysis of five plastid-encoded genes showed strong support for a single origin of plastids in cryptophytes, haptophytes, and heterokonts (Yoon et al. 2002). Between these two studies, there is now evidence for a single origin for all chromalveolate plastids, but neither study included data from all chromalveolate plastid lineages. Here, we report the sequences of cytosolic, mitochondrion-targeted, and plastid-targeted GAPDH genes from several haptophyte and heterokont taxa. The characterization of haptophyte plastid-targeted genes makes GAPDH the first protein-coding molecular data set to include representatives of all chromalveolate plastids, and the resulting phylogenies strongly support a single origin.

Materials and Methods

Strains and Culture Conditions

Unialgal axenic cultures of the heterokont *Mallomonas rasilis* (strain CCMP 478) and the haptophytes *Isochrysis galbana* (strain CCMP 1323), *Pavlova lutheri* (strain CCMP 1325), and *Prymnesium parvum* (strain CCMP 1926) were obtained from the Provasoli-Guillard National Centre for Culture of Marine Phytoplankton and CCMP 1926) were obtained from the Provasoli-Guillard National Centre for Culture of Marine Phytoplankton and grown in f/2-Si medium at 16°C (12:12 light:dark cycle). Genomic DNAs from the oomycete heterokonts *Aphochlyla brachynema* (strain CBS 557.69), *Phytophthora palmivora* (strain CBS 236.30), *Plectospora myriandra* (strain CBS 523.87), *Pythium graminicola* (strain CBS 327.62), and *Thraustotheca clavata* (strain CBS 343.33) were kindly donated by A. W. DeCock.

DNA and RNA Isolation and Amplification of GAPDH Genes

Algal cultures were harvested by centrifugation, and cell pellets were lysed by grinding in a Knots Duall 20 tissue homogenizer. Genomic DNAs (gDNAs) were extracted from *M. rasilis* and *P. parvum* lysates using the DNeasy Plant Mini Kit (Qiagen). Total RNA was isolated from *I. galbana* and *P. lutheri* lysates using Trizol reagent (Invitrogen).

GAPDH genes were PCR-amplified from gDNAs using primers 5′-CAAGGCTCAGNATHAYGNNTTY-GG-3′ and 5′-CGAGTAGCCCCAYT CRTTRCTTCA-CA-3′ under the following conditions: 95°C for 2 min; 40 cycles of 92°C for 45 s, 48°C for 45 s, and 72°C for 1 min and 30 s; and 72°C for 5 min. GAPDH cDNAs were amplified by RT-PCR from isolated total RNA under the following conditions: 42°C for 20 min; 95°C for 5 min; 36 cycles of 92°C for 45 s, 45°C for 45 s, and 72°C for 1 min and 30 s; and 72°C for 5 min. PCR-products were gel-purified and cloned into the TOPO-TA vector pCR2.1 (Invitrogen), and multiple clones of each were sequenced on both strands with ABI BigDye terminator chemistry (Applied Biosystems).

Phylogenetic Analyses

New GAPDH sequences were added to an existing amino acid alignment. Positions of insertions were located on the crystal structure of the *E. coli* GapA protein (PDB accession 1DC5). Distance and maximum-likelihood (ML) analyses were performed on two data sets: a global GAPDH alignment of 91 sequences and a reduced alignment of 57 GapC sequences (both alignments included 269 amino acid characters [alignments available from the authors]). ML distances were calculated using Tree-Puzzle version 5.0 (Strimmer and von Haeseler 1996), using eight rate categories plus invariable sites with parameters described elsewhere (Keeling 2003). Distance trees were constructed with weighted neighboring-joining using Neighbor version 1.0.1a, (Bruno, Socci, and Halpern 2000) and Fitch-Margoliash using Fitch version 3.6a (Felsenstein 1993). Fitch-Margoliash trees were inferred using the global rearrangements option and 10 input order jumbles. Bootstrapping was carried out as previously described (Keeling 2003). Protein maximum-likelihood analysis was performed using ProML version 3.6a (Felsenstein 1993) using parameters described elsewhere (Keeling 2003). For GapC analyses, site-to-site rate variation was modeled on a gamma curve using the –r option with four variable rate categories and invariable sites (rates and frequencies were estimated by Tree-Puzzle). ProML bootstrap trees were constructed for the 57-taxon data set as above, but with no site-to-site rate variation and a single randomized input order.

Results and Discussion

Characterization of GAPDH Genes from Haptophytes and Heterokonts

Twelve novel GAPDH genes were amplified from three haptophytes and six heterokonts (GenBank accession numbers AY292371 to AY292382). All new GAPDH genes were similar to the various forms characterized from other chromalveolates: a cytosolic form, a plastid-targeted form, and the GAPDH moiety of a mitochondrial-targeted TPI-GAPDH fusion protein described in oomycetes and diatoms (Unkles et al. 1997; Liaud et al. 2000). A full-length GAPDH from the haptophyte *I. galbana* most similar to the cytosolic form was also characterized from an ongoing *I. galbana* EST survey and, as expected, did not possess a leader (data not shown).

GAPDH uses either NADH or NADPH as a cofactor, and the cofactor specificity differs between the cytosolic protein involved in glycolysis and gluconeogenesis and the plastid form. In general, the cytosolic GAPDH has three highly conserved residues, D32, G187, and P188 (numbered according to Clermont et al. 1993), that confer NADH specificity, but in plastid-targeted GAPDH, these residues are not conserved. Among the new sequences characterized here, all three residues are conserved in all...
sequences except for *P. palmivora* (where D32 is not conserved) and one paralog from *M. rasillis* and one paralog from each of the haptophytes *I. galbana* and *P. lutheri*. In *I. Galbana*, all three residues are substituted, whereas in *M. rasillis* and *P. lutheri*, only G187 is conserved. This is also seen in the plastid-targeted GAPDH from *Toxoplasma gondii* (Fast et al. 2001), and this residue is known to be the least important for cofactor specificity (Clermont et al. 1993).

Phylogeny of GAPDH

The overall features of the global GAPDH phylogeny (fig. 1) are quite similar to previous results (Henze et al. 1995; Liaud et al. 1997; Fagan, Hastings, and Morse 1998; Fast et al. 2001). GAPDH trees generally consist of GapC, which is predominantly composed of eukaryotic cytosolic genes, and GapA/B, which is predominantly composed of eubacterial genes, including the cyanobacterial and plastid-targeted GAPDH genes from green algae, land plants, red algae, *P. lumula*, and *E. gracilis*. Branching between these two groups are lineages that cannot easily be ascribed to either major clade (e.g., several bacterial homologs and euglenozooan genes). All GAPDH genes from chromalveolates branch in the GapC clade, and none are related to the plastid-targeted GapA/B genes found in plants and other algae. The single exception to this is the plastid-targeted GAPDH from the dinoflagellate *P. lumula*, which branches with the *E. gracilis* plastid GapA/B (Fagan and Hastings 2002). This relationship is not expected if this were the ancestral dinoflagellate plastid-targeted GAPDH, making its origin unclear.

Most importantly, GAPDH paralogs from the haptophytes *I. galbana* and *P. lutheri* fell within the very strongly supported chromalveolate plastid-targeted GAPDH clade (also supported at 99% to 100% in all analyses of GapC alone, see Supplementary Material online). These genes are clearly descended from a plastid-targeted ancestor, and both genes lack the highly conserved triplet of residues that confer a strict specificity for NADH, suggesting that the proteins also function in the Calvin cycle. The two haptohyte sequences grouped together with very strong support, specifically with plastid-targeted GAPDH from the apicomplexan *T. gondii*. With the addition of these sequences, this clade now includes representatives from all chromalveolate groups known to retain a plastid, or in other words, all eukaryotes with secondary plastids of red algal origin.

In general, the nonplastid GAPDH sequences from chromalveolates formed a number of groups whose positions were equivocal. The heterokont cytosolic GAPDH sequences do form a weakly supported clade with alveolate cytosolic GAPDH, and this clade is weakly but consistently related to the chromalveolate plastid-targeted GAPDH clade. However, the cytosolic GAPDH from cryptophytes and haptophytes show no real affiliation with other eukaryotic lineages. This could be taken to indicate that these genes belong to a different functional subfamily of enzyme than the heterokont and alveolate cytosolic proteins, but it is more likely that the position of these sequences is simply not resolved. We infer these to be the cytosolic GAPDH since they maintain the highly conserved NADH-specific cofactor-binding triplet, and the full-length cryptophyte and *I. galbana* sequences do not encode leaders. The origin of the GAPDH portion of the heterokont mitochondrion-targeted TPI-GAPDH fusion protein was also not resolved, nor was the nonplastid gene from *M. rasillis*. In distance analyses of both alignments, the gene branched weakly with the mitochondrion-targeted form from other heterokonts, suggesting it is mostly likely a member of this clade.

A two–amino acid insertion found in an exterior loop of GAPDH has been interpreted as a possible signature for chromalveolate cytosolic GAPDH (Fast et al. 2001) This insertion was found in all new heterokont cytosolic GAPDH genes but interestingly not in those of cryptophytes or haptophytes (see Supplementary Material online). The plastid-targeted GAPDH sequences for the haptophyte taxa are also novel with respect to this insertion, as *I. galbana* and *P. lutheri* both lack only one residue in this region. It is interesting to note that those chromalveolate cytosolic GAPDH sequences, which do not possess the two–amino acid insertion (cryptophytes and haptophytes), are those same taxa that do not branch in the chloroplast-targeted cytosolic GAPDH clade.

The specific grouping of the haptophyte plastid-targeted GAPDH sequences with that of the apicomplexan *T. gondii* to the exclusion of the remaining chromalveolates is unexpected since apicomplexans are much more closely related to dinoflagellates than to haptophytes. Nevertheless, the apicomplexan-haptophyte relationship is extremely well supported (94% to 100% in all methods in analysis of GapC, see Supplementary Material online). One region of the protein (fig. 2) contains a two–amino acid insertion shared by *P. lutheri* and *T. gondii*, and the second contains a one–amino acid insertion shared by *T. gondii* and both haptophytes. These insertions are also located in an external loop and are likely of consequence structurally, but do suggest a specific relationship between these sequences. We tested these sequences for signs of recombination, but none were detected (not shown). The significance of the haptophyte-apicomplexan relationship is not yet clear, but it should be considered carefully in light of the fact that no plastid-targeted GAPDH has been found in the nearly complete genome of the apicomplexan *Plasmodium* (Gardner et al. 2002). There is no obvious reason why it should have been retained in *T. gondii* and lost in *Plasmodium*, but if it is absent it suggests that the metabolism of the *Plasmodium* plastid has lost whatever pathway in which GAPDH played a role or that *T. gondii* acquired its GAPDH gene secondarily. To distinguish between these alternatives, an increased sampling of apicomplexan and haptophyte plastid homologs would be highly desirable to determine whether the *T. gondii* GAPDH was inherited vertically or by some other means.

Concluding Remarks: A Single Red Algal Secondary Endosymbiosis

A single origin for chromalveolate plastids has several interesting implications for plastid evolution. First,
there has been a debate over the red versus green nature of the apicomplexan plastid ever since its discovery (Williamson et al. 1994; Köhler et al. 1997; McFadden and Waller 1997; Blanchard and Hicks 1999; Zhang, Green, and Cavalier-Smith 2000; Fast et al. 2001; Funes et al. 2002). The shared GAPDH gene replacement event suggests that the ancestor must have possessed a red plastid. Second, a common origin for chromalveolate
plastids obviously means the host lineages should also be related. This may seem trivial, but genes for host cytosolic proteins have so far failed to provide strong support for this relationship. In most cases, these phylogenies do not support a clear alternative either, much as with cytosolic GAPDH. An analogous situation was seen previously with plastid-encoded genes from chromists, even when a large number of concatenated genes were analyzed (Martin et al. 2002). Only by concatenating a substantial number of genes from a broad range of taxa was the relationship recovered with high support (Yoon et al. 2002). In the case of GAPDH, we suspect that the conversion of a NADH-specific cytosolic enzyme to a NADH/NADPH-specific plastid-localized enzyme may have precipitated a short period of rapid adaptation, which has allowed the chromalveolate plastid-targeted clade to be recovered with high support. The prediction from these results and those of plastid-encoded genes (Yoon et al. 2002) is that analyses of multiple nuclear protein–coding genes will also be required to reveal a close relationship among the host component of chromalveolates.

Lastly, a single origin for chromalveolate plastids has a significant impact on how we regard the process of plastid loss. A great deal of work has focused on plastid origins, but the reverse process is far more mysterious. Indeed, it is even difficult to say whether any of the nonphotosynthetic lineages that are predicted to have evolved from a plastid-bearing ancestor have lost the organelle or whether we simply have not yet discovered it. Refining our thinking on “plastid loss” versus “loss of photosynthesis” is clearly important. Moreover, proving that cryptic organelles may be found in many lineages that cryptic organelles may be found in many lineages where they have not traditionally been looked for (e.g., ciliates), exactly as they were discovered in the apicomplexans (McFadden et al. 1996; Wilson 2002). Alternatively, if the organelle has been lost altogether, molecular relics of a plastid ancestry should exist, as has been suggested in oomycetes (Andersson and Roger 2002). Ironically, only when the processes of plastid loss and plastid degradation are better understood will we begin to fully understand how they originated.

Acknowledgments

This work was supported by a grant (MOP-42517) from the Canadian Institutes for Health Research (CIHR). Isochrysis EST sequencing is part of the Protist EST Program supported by Genome Canada via Genome Atlantic. We thank A.W.A.W. DeCock for providing gDNAs from oomycetes, J. M. Archibald, N. M. Fast, M. B. Rogers, B. A. P. Williams, and B. S. Leander for critical reading of the manuscript. P.J.K. is a scholar of the Canadian Institute of Advanced Research and a New Investigator of the Michael Smith Foundation for Health Research and the CIHR.

Literature Cited

A Single Origin for Chromalveolate Plastids 1735


Franz Lang, Associate Editor

Accepted June 10, 2003