Letter to the Editor

Phylogenetic Analysis Indicates Multiple Origins of Chloroplast Glyceraldehyde-3-Phosphosphate Dehydrogenase Genes in Dinoflagellates

Thomas F. Fagan and J. Woodland Hastings

Department of Molecular and Cellular Biology, Harvard University

Although the endosymbiotic theory of evolution (Margulis 1970) is widely accepted, the series of events that led to the permanent inclusion of mitochondria and chloroplasts in eukaryotic cells are poorly understood. In particular, the diverse biochemical and morphological properties of chloroplasts have led to suggestions that these organelles have been acquired through multiple primary endosymbiotic events (for a review see Delwiche 1999). Furthermore, the presence of three and sometimes four (Gibbs 1962) membranes around certain chloroplasts, some in association with a second nucleus, suggests that secondary endosymbioses—the incorporation within a eukaryote of a heritable organelle from another eukaryote—has also occurred (Gibbs 1981). Secondary endosymbioses may also have occurred more than once (Delwiche and Palmer 1997; Delwiche 1999); in fact, the possibility that some autotrophic eukaryotes lost their photosynthetic organelles only to regain them in a later endosymbiotic event cannot be excluded. The occurrence of such multiple, sequential endosymbiotic events can only be proven if some remnant of the first endosymbiont was retained by the host cell. Nuclear-encoded genes for chloroplast proteins, for example GAPDH, could provide such evidence.

Although phylogenetic analyses have resolved cytosolic (GapC) and chloroplast (GapA) sequences into two distinct clades (Martin et al. 1993; Liaud et al. 1997), recent studies of gapdh isoforms from marine algae have complicated the simple cytosol-chloroplast tree structure. In the dinoflagellate Lingulodinium polyedrum (Stein) Dodge, formerly Gonyaulax polyedra and Morse 1998) and in two cryptomonads (Guillardia theta and Pyrenomonas salina) the role of GAPDH in the Calvin cycle is filled by a modified cytosolic isoform (GapC-I), which has a signal sequence at the N-terminal end for intracellular translocation to the chloroplast and, consistent with its anabolic role, amino acid substitutions that allow binding of NADPH (Liaud et al. 1997; Fagan, Hastings, and Morse 1998). Phylogenetic analysis places GapC-I firmly in the cytosolic clade (Fagan, Hastings, and Morse 1998). Homologous isoforms have been identified subsequently in heterokonts and apicomplexans (Liaud et al. 2000; Fast et al. 2001).

How the GapC-I isoform was acquired and how pervasive it is among dinoflagellates is unknown. We proposed that it was obtained through lateral transfer (Fagan, Hastings, and Morse 1998) because dinoflagellates are known to engulf and harbor organelles of other algae (Larsen 1988; Lewitus, Glasgow, and Burkholder 1999). In order to check the validity of this hypothesis, we undertook the cloning and analysis of gapdh sequences from other dinoflagellates. Surprisingly, we have found that two species of Pyrocystis (P. lunula and P. noctiluca) possess a GapA isoform, indicating that acquisition of GAPDH genes in the dinoflagellates may have occurred more than once and by different mechanisms.

gapdh sequences were amplified by RT-PCR using total RNA from P. lunula as template. After reverse transcription, degenerate oligonucleotide primers coding for the peptide SNASCTT, which is found in the active site of almost all GAPDH isoforms, were used in initial amplifications together with an antisense primer directed to the polyA region of the first strand cDNA. Potential gapdh products, as determined by size, were cloned; sequences identified as homologous to gapdh were used to design nondegenerate primers for subsequent 5' rapid amplification of cDNA ends (RACE) experiments.

A set of overlapping sequences totaling 1,305 bp in length was obtained from P. lunula. The cDNA has a GC content of 59.5% and contains an open reading frame of 1,032 bp, encoding a predicted protein sequence of 344 amino acids (36.6 kDa). A BLAST (Altschul et al. 1990) search of the protein sequence against the NCBI nonredundant database showed that the protein is homologous to GAPDH, but surprisingly the most statistically significant alignments were with chloroplast GAPDH enzymes (GapA) from Euglena gracilis and the red alga Chondrus crispus. This suggested that the P. lunula sequence is a homolog of the GapA isoforms, corresponding to those found in most other photosynthetic eukaryotes.

Pairwise alignments showed that the predicted protein sequence of this isoform is 37% and 41% similar to the cytosolic (GapC) and chloroplast (GapC-I) isoforms, respectively, from L. polyedra. This is in striking contrast to the greater similarity (~68%) found between the GapC-I isoforms from L. polyedra and the cryptomonads G. theta and P. salina (Fagan, Hastings, and Morse 1998). In addition, the P. lunula isoform and the GapA isoform from the green alga E. gracilis show much greater similarity (~72%), suggesting the idea that the sequence isolated is a GapA homolog.

In order to confirm that this gene was amplified from P. lunula, Northern blots of total RNA were probed using the P. lunula GapA gene and a P. lunula luciferase (LCF) probe as control; dinoflagellate luciferases are unique to these algae and provide an excellent diagnostic tool to distinguish between dinoflagellate and non-dinoflagellate nucleic acids. Under stringent hybrid-
GAPDH gene, comprising 760 nucleotides covering the C-terminal end of the coding region (192 amino acids) and the 3'UTR. Although the UTR (184 bp) is only 27% identical to the P. lunula gene 3'UTR (174 bp), the coding sequences are 87% identical, indicating that the two genes are homologous.

We investigated the relationship between the P. lunula sequence and those of other GAPDH isoforms further by phylogenetic analyses. Alignments of protein sequences from 36 representative taxa were used, including photosynthetic and nonphotosynthetic eukaryotes and prokaryotes. Phylogenetic trees generated by parsimony and distance analysis gave similar topologies (fig. 2).
Using the sequence from the archaea *Sulfolobus solfataricus* as an out-group, two major clades were resolved by both analytical treatments. The first comprises mostly cytosolic isoforms, including dinoflagellate and cryptophyte GapC sequences. In addition, proteobacterial (*Escherichia coli*) and cyanobacterial (*Anabaena 2*) sequences are found in this clade, as are the dinoflagellate and cryptophyte GapC-I sequences, which branch deeply within it; these form a single clade with the GapC-I sequences from heterokonts and the apicomplexan *Toxoplasma gondii*, previously shown to contain vestigial chloroplasts. The second major clade consists of isoforms that are involved in the Calvin cycle of photosynthetic organisms, cyanobacterial (*Anabaena 1*) and chloroplast GapA sequences. There is robust support for these two clades in both types of analyses, and the tree topologies are consistent with many published previously (Martin and Cerff 1986; Liaud et al. 2000; Fast et al. 2001). Both parsimony and distance analyses place the *P. lunula* sequence firmly within the GapA clade. It branches deeply within the clade and forms a sister group with the GapA sequence from *E. gracilis*.

Support for this topology comes from phylogenetic analysis of a subset of characters, comprising 167 amino acids of the C-termini of the proteins. This analysis also places the *P. noctiluca* sequence in the GapA clade, with both *Pyrocystis* species forming a sister group with *Euglena* (fig. 2 inset).

If the proteins coded by these two *Pyrocystis* genes are truly chloroplast GAPDHs, then they should contain N-terminal transit sequences. The 5' end of the *P. lunula* GapA, though incomplete, does have such characteristics, being rich in hydroxylated residues and long hydrophobic stretches. Indeed, ChloroP (Emanuelsson, Nielsen, and Von Heijne 1999), an algorithm that has proven to be reliable in predicting transit sequences and their cleavage sites, indicates that the N-terminus of this gene does contain a transit sequence and predicts a cleavage site compatible with the expected start of a mature enzyme (fig. 1C).

The identification of GapA genes in *Pyrocystis* species and a GapC-I gene in *L. polyedrum* raises questions about the origin and distribution of gapdh genes in dinoflagellates. The recent identification of GapC-I in the diatoms *Odontella sinensis* and *Phaeodactylum tricornutum* (Liaud et al. 2000) and in the Apicomplexa (Fast et al. 2001) indicate that this isoform may be more widely distributed than previously thought. Cryptomonad, di-
atom, and apicomplexan plastids have been acquired through secondary endosymbiosis, with several lines of evidence, such as ultrastructural, pigmentation, and phylogenetic (Durnford, Aebersold, and Green 1996), indicating that the plastid donor was a red alga. Although dinoflagellate plastids differ in some respects (e.g., organization of plastid genes on minicircles [Zhang, Green, and Cavalier-Smith 1999] and lack of phycobilisomes, Jeffrey 1980), there is phylogenetic evidence from psbA photosystem sequences (Takishita and Uchida 1999; Zhang, Green, and Cavalier-Smith 2000) and ribosomal sequences (Zhang, Green, and Cavalier-Smith 1999) supporting a red algal origin for dinoflagellate plastids also. Thus, the presence of the GapC-I isoform in these phyla could be explained by its acquisition during a single secondary endosymbiotic event in which an ancestor of present-day photosynthetic dinoflagellates, cryptomonads, diatoms, and apicomplexans engulfed a red alga-type eukaryote. As proposed by Fast et al. (2001), a duplication of the cytosolic GapC gene in the ancestor, followed by addition of a targeting sequence to one of the copies, could explain the origin of the GapC-I isoform in all these phyla. This theory is supported by phylogenetic analyses, which show GapC-I and GapC clans, comprising taxa from these phyla, as sister groups (Fast et al. 2001). Indeed, a precedent for such a duplication occurs in the extant gymnosperm Pinus sylvestris (Meyer-Gauen et al. 1994). On the basis of parsimony alone the single ancestor theory seems highly likely.

But what of the GapA isoform? GapAs are present in red algae, raising the possibility that dinoflagellates acquired it from the red algal endosymbiont in the single ancestor. If there was such an origin, GapA should have been transferred to the nucleus of the single ancestor and then selectively replaced by the duplicated GapC-I in only some dinoflagellates. This seems unlikely. Moreover, the Pyrocystis genes form a clade with Euglena (see fig. 2) and are thus closely related to the green algal lineage because the position of Euglena in the tree basal to the divergence of the red and green lineages has been shown to be caused by long branch attraction (Figge et al. 1999). A red algal origin for the Pyrocystis GapA is, therefore, unlikely.

It is possible that these genes arose through lateral transfer from an engulfed photosynthetic prey. But this would surely require that the host, in this case the ancestor to Pyrocystis, be a heterotroph, thus indicating that a second endosymbiotic event has taken place within the dinoflagellates. We propose that the two chloroplast-targeted gapdh isoforms arose from independent endosymbiotic events. It has been suggested that the ancestor to all dinoflagellates was photosynthetic (Cavalier-Smith 1999), an idea that has received phylogenetic support (Zhang, Green, and Cavalier-Smith 2000). Extant heterotrophic dinoflagellates, or their ancestors, must therefore have lost their chloroplasts. It is worth noting, however, that phylogenetic analysis of small ribosomal subunit sequences places several heterotrophs (Noctiluca scintillans, Cryptophyllum colurn) basal to many photosynthetic species (Saunders et al. 1997), supporting the possibility that a second endosymbiotic event has taken place.

Sequences reported here have been deposited in GenBank and can be retrieved using accession numbers AF406628 (P. lunula) and AF406629 (P. noctiluca).

Acknowledgments

This work was supported by grants from the NIH (MH 46660) and NSF (9982880). We would like to thank Deb Robertson and Thérèse Wilson for their helpful comments and suggestions, and Keith Okamoto for his help in propagating Pyrocystis colonies.

LITERATURE CITED


WILLIAM MARTIN, reviewing editor

Accepted March 14, 2002