J Mol Evol (1998) 47:633–639

JOURNAL OF MOLECULAR EVOLUTION

The Phylogeny of Glyceraldehyde-3-Phosphate Dehydrogenase Indicates Lateral Gene Transfer from Cryptomonads to Dinoflagellates

Thomas Fagan,[1] J. Woodland Hastings,[1] David Morse[2]

[1] Department of Molecular & Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138-2020, USA
[2] Department of Biology, University of Montreal, Montreal, Canada H1X2B2

Received: 29 April 1998 / Accepted: 7 July 1998

Abstract. Sequence analysis of two nuclear-encoded glyceraldehyde-3-phosphate dehydrogenase (GAPDH) genes isolated from the dinoflagellate *Gonyaulax polyedra* distinguishes them as cytosolic and chloroplastic forms of the enzyme. Distance analysis of the cytosolic sequence shows the *Gonyaulax* gene branching early within the cytosolic clade, consistent with other analyses. However, the plastid sequence forms a monophyletic group with the plastid isoforms of cryptomonads, within an otherwise cytosolic clade, distinct from all other plastid GAPDHs. This is attributed to lateral gene transfer from an ancestral cryptomonad to a dinoflagellate, providing the first example of genetic exchange accompanying symbiotic associations between the two, which are common in present day cells.

Key words: Glyceraldehyde-3-phosphate dehydrogenase — Gene transfer — Cryptomonads — Dinoflagellates

Introduction

Dinoflagellates have features that make their phylogenetic affinities and evolutionary history enigmatic. Their nuclei contain up to 200 pg DNA per nucleus (Holm-Hansen 1969) and lack histones (Rizzo 1981), while the chromosomes remain permanently condensed through the S phase (Grasse and Dragesco 1957). Mitosis is distinct from that of other eukaryotes (Spector et al. 1981); there is no spindle and chromosomes are segregated by attachment to the nuclear envelope, which remains intact throughout.

Phylogenetic analysis based on large nuclear-encoded ribosomal RNA sequences has placed the dinoflagellates with ciliates and apicomplexans (Baroin et al. 1988; Gajadhar et al. 1991; Lenaers et al. 1991; Woese 1987). However, about 50% of dinoflagellate species are photosynthetic, with chloroplasts surrounded by a triple membrane; whether this is explained by capture of prokaryotic or eukaryotic endosymbionts is debated (Gibbs 1981a). In addition, most contain the carotenoid peridinin, which to date has been found only in dinoflagellates (Jeffrey et al. 1975; Loeblich 1976).

We have obtained the sequence of two glyceraldehyde-3-phosphate dehydrogenase (GAPDH) genes from *Gonyaulax polyedra*, a photosynthetic dinoflagellate. GAPDH has been extremely useful in phylogenetic analysis, notably in demonstrating a symbiotic origin of chloroplasts in vascular plants (Shih et al. 1986; Martin and Cerff 1986; Martin et al. 1993). It is highly conserved and its genes have been sequenced from a large number of different taxa. All known plastid-containing eukaryotes have two different forms of the enzyme (Cerff 1995; Martin and Schnarrenberger 1997). One, a glycolytic enzyme, specific for NAD^+, is typically found in the cytosol. A second form, found in chloroplasts, can utilize both NAD^+ and $NADP^+$ and is involved in Calvin cycle reactions. We report here that the phylogeny of the

Correspondence to: J.W. Hastings; *e-mail:* hastings@fas.harvard.edu

cytosolic form of (GAPC) is consistent with previous nuclear ribosomal RNA analysis. However, as in the case of the recently reported cryptomonads (Liaud et al. 1997), the chloroplast form (GAPCp) is different from that found in all other plastids.

Materials and Methods

Clones for GAPDH were obtained by polymerase chain reaction and by screening of a *G. polyedra* cDNA library. Sequences were determined using the dideoxy chain termination method (Sanger et al. 1977), and their identities determined by BLAST searches of the NCB1 sequence data banks (Altschul et al. 1990). The forward degenerate primer GAPP 5a (TCSAACGCNTCSTGYACBAC), which codes for the conserved peptide sequence SNASCTT, surrounding the GAPDH catalytic cysteine residue, was used with the vector primer T7 to amplify a 765-bp fragment from a *Gonyaulax* cDNA library directionally cloned into the λ Zap vector. This fragment had significant homology with mammalian GAPDH genes and allowed design of the nondegenerate antisense primer GAPP 27 (CACAAAACAAGCGAGCTTCACTC), which, together with the vector primer T3, amplified a 1.3-kb product from the same library. This product contained a complete open reading frame which, based on sequence, was predicted to encode a GAPDH enzyme. GAPP 5a was also used together with the reverse degenerate primer GAPP 4a (RATSGGGTTVGTCTCSARSTC), the complement of which encodes the peptide DLETNPI, originally identified by peptide sequences of proteins isolated from 2D gels (Markovic et al. 1996), to amplify a 430-bp fragment. This fragment was then used to create the reverse primer GAPP 4b (TAGTCTCGAAGTCTGTGGACAC), which, together with the vector primer T3, amplified a 1.1-kb clone. This was used to isolate a 1.4-kb cDNA clone from the library, which, though different in sequence from the 1.3-kb clone, also contained a complete open reading frame and showed sequence similarities to GAPDH genes from other organisms.

Sequences from a wide range of species were utilized in order to obtain a comprehensive picture of GAPDH phylogeny. Sequences were aligned using ClustalW (Thompson et al. 1994) and adjusted manually; the N- and C-terminal ends of the sequences were trimmed to yield a common block of sequences 339 amino acids long, and equivocal residues were removed. Parsimony analysis was carried out using PAUP Version 3.1.1 written by David L. Swofford, and the 50% majority rule consensus tree was retained after a heuristic search of 100 bootstrap resamplngs. Distance analysis was carried out using software in the Phylogeny Inference Package (PHYLIP) Version 3.572, by Joseph Felsenstein, University of Washington. The Dayhoff PAM matrix was used to generate distance matrices (Protdist) from 100 bootstrap resamplings (Seqboot). The majority-rule consensus tree was computed (Consense) after unrooted trees were generated from the distance matrices by the neighbor joining method (Neighbor). Branch lengths were obtained from an unbootstrapped distance matrix using the unrooted consense tree as a user tree (Fitch). The consense tree is graphically displayed using TreeView 1.5 (Page 1996). The sequences used can be obtained from GENBANK using the following accession numbers: *Anabaena* 1, P34916; *Anabaena* 2, P34917; *Arabidopsis* A, P25857; *Arabidopsis* C, P25858; *Chlamydomonas* A, P50362; *Chlamydomonas* C, P49644; *Chondrus* A, P34919; *Chondrus* C, P34920; *Drosophila,* P07486; *E. coli* 1, P06977; *E. coli* 2, P11603; *Euglena* A, P21904; *Euglena* C, L21903; *Giardia,* M88062; *Gonyaulax C,* AF028562; *Gonyaulax* Cp, AF028560; *Gracillaria* C, P54270; *Gracillaria* A, P30724; *Guillardia* C, U39873; *Guillardia* Cp, U40032; *Methanothermus,* P19314; *Neurospora,* P54118; Pinus A, P12859; *Pinus* Cp, L07501; *Pisum* A, P12859; *Pyrenomonas* C, U39897; *Pyrenomonas* Cp, U40033; *Pyrococcus,* P20286; *Rhodobacter,* P29272; *Saccharomyces,* P00360; *Sulfolobus,* P39460; *Synechocystis,* P49433; *Trypanosoma* C, P10097; and *Xenopus,* P51469. The alignment is available upon request.

Results and Discussion

Two cDNAs encoding GAPDH were isolated from *Gonyaulax polyedra,* and though slightly different in size (~1.3 and ~1.4 kb), both show homology to cytosolic forms of GAPDH from other eukaryotes. The 1.3-kb form (GAPC) encodes a cytosolic enzyme (see below); it contains a 1029-bp open reading frame, predicting a protein of 36 kDa, similar to the 32 to 36 kDa reported for other cytosolic GAPDH enzymes.

Two lines of evidence indicate that the 1.4-kb form encodes a protein (GAPCp) localized to the chloroplast. First, it contains a presequence indicative of chloroplast targeting. Its 1275-bp open reading frame encodes a ~45-kDa protein, considerably larger than the cytosolic form, but similar in mass to the 42 and 48 kDa for chloroplast GAPDH precursors from *Arabidopsis* and *Pisum,* respectively (Brinkmann et al. 1989; Shih et al. 1991, 1992) (Fig. 1a). Although the cleavage site has not been determined, the molecular mass of the mature protein (38–40 kDa; data not shown) is consistent with the removal of a presequence. In addition, the N-terminal region of the *G. polyedra* GAPCp is similar to those found in other chloroplast-localized proteins in dinoflagellates (Le et al. 1997) and *Euglena* (Chan et al. 1990; Henze et al. 1995; Sharif et al. 1989) (Fig. 1b), These N-terminal regions have in common a typical signal peptide structure (von Heijne 1986) with a charged N terminus, followed by a hydrophobic core and peptidase cleavage site, and a second region rich in hydroxylated and charged residues, containing another hydrophobic stretch (Fig. 1b). The similarity to the *Euglena* sequences, which are involved in directing proteins to the chloroplast via the endoplasmic reticulum (Kishore et al. 1993), suggests that proteins may be likewise targeted to plastids in dinoflagellates. Also, the chloroplasts in both groups are surrounded by a triple membrane, which in *Euglena* has been shown to be associated with the endoplasmic reticulum (Gibbs 1981b).

Second, a chloroplast location of the GAPCp is indicated by the identity of three key amino acid residues involved in nucleotide binding, which differ in the cytosolic and chloroplastic forms of the enzyme (Fig. 2) (Clermont et al. 1993). Eukaryotic cytosolic forms, which typically fulfill a catabolic role in glycolysis and are NAD^+ specific, have conserved aspartate, glycine, and proline residues at positions 32, 187, and 188, respectively, numbered as in *Bacillus* (Biesecker et al. 1977). Plastid GAPDHs, which typically have an anabolic role and are able to utilize both NAD^+ and $NADP^+$, have substitutions at these positions. In *G. polyedra* the residues are indeed conserved in GAPC, confirming its identity as a cytosolic form, but not in GAPCp, where alanine occurs at positions 32 and 188, and serine at 187.

Based on the deduced amino acid sequence, *G. polyedra* GAPCp is less than 42% identical to the previously known chloroplast forms of vascular plants, chloro-

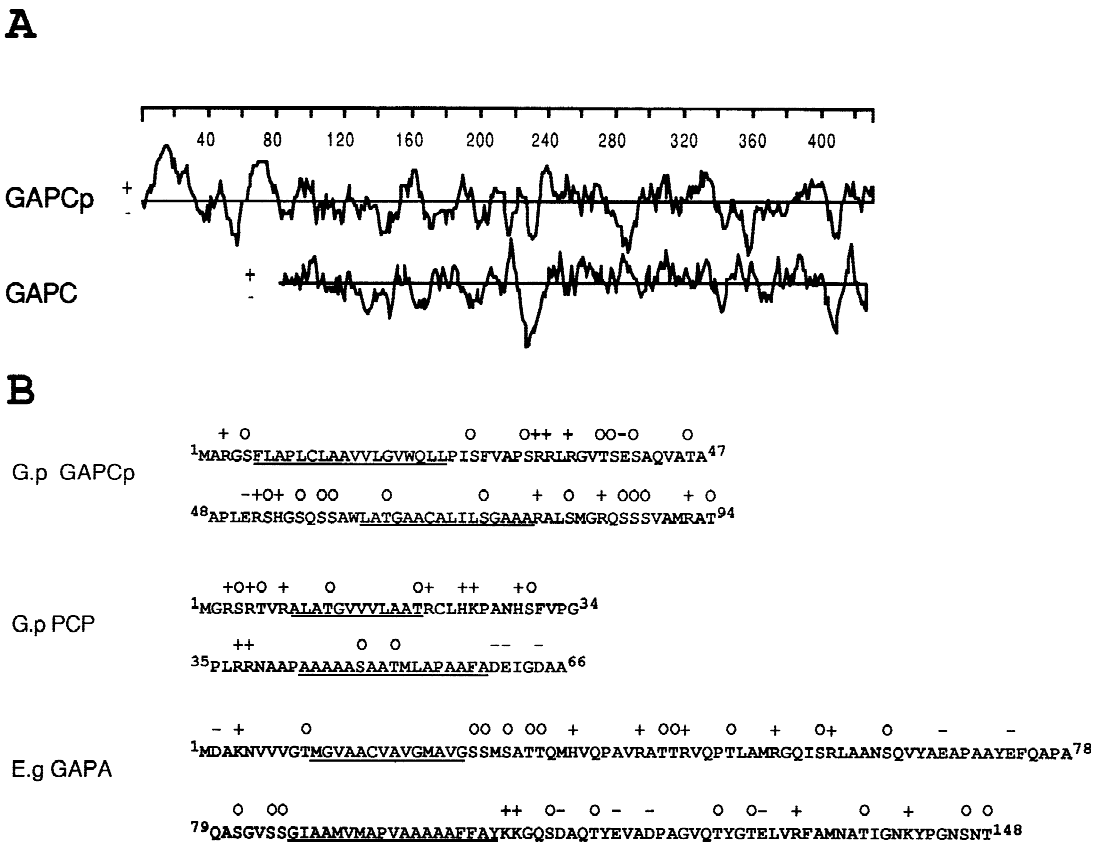

Fig. 1. **A** Hydrophobicity plot of *G. polyedra* GAPDH sequences, obtained using the Protean module of the DNASTAR (Madison, WI) package based on parameters of Kyte–Doolittle (1982). GAPC is the cystolic isoform, while GAPCp, with its additional N-terminal sequence, is the chloroplast form. **B** N-terminal amino acid sequences of nuclear-encoded chloroplast targeted proteins. G.p., *Gonyaulax polyedra:* E.g., *Euglena gracilis;* GAPCp, this paper; PCP, peridinin–chlorophyll a binding protein (Lee et al. 1997); GAPA, chloroplast glyceraldehyde-3-phosphate dehydrogenase (Henze et al. 1995). The symbols +, −, and ○ refer to positively charged, negatively charged, and hydroxylated amino acids, respectively. Hydrophobic regions are *underlined. Superscripted numbers* identify amino acids positions. The N-terminal amino acids of the mature proteins are unknown.

phytes, and rhodophytes, which among themselves are greater than 60% identical (Fig. 3). Indeed, excluding the N-terminal presequence, the *G. polyedra* GAPCp is more identical to the cytosolic than to the chloroplastic form of GAPDH found in other eukaryotes. However, *G. polyedra* GAPCp exhibits a high sequence similarity to the recently reported (Liaud et al. 1997) chloroplast GAPDH sequences from two cryptomonads, *Guillardia theta* and *Pyrenomonas salina* (67 and 68%, respectively). An analysis of the cryptomonad sequences suggested that they may have been derived from proteobacteria (Liaud et al. 1997), as proposed previously for the *G. polyedra* RuBisCO gene (Morse et al. 1995). Irrespective of their origin, it is clear from the amino acid sequence comparisons that dinoflagellate and cryptomonad plastid GAPCp proteins are similar to one another and different from GAPDHs found in other chloroplasts.

In order to obtain a more detailed picture of the evolution of *G. polyedra* GAPDH genes, phylogenetic analyses were performed using sequences from a broad array of taxa. Similar tree topologies were obtained with either maximum-parsimony or distance analysis (Fig. 2). In both analyses the taxa fall into three main clades: (i) the Archaea, which was used as an outgroup in the parsimony analysis; (ii) plastids of most higher plants and algae together with the eubacteria; and (iii) primarily eukaryotic cytosolic sequences.

In the parsimony analysis the cyanobacterial sequences (*Synechocystis* and *Anabaena*) appear as a polytomy with the red alga and vascular plant sequences, while the distance analysis shows modest support for earlier branching of the cyanobacterial lineages. The analyses are thus consistent with a cyanobacterial origin of chloroplasts in vascular plants and rhodophytes. Somewhat unexpectedly (as cyanobacteria are also thought to be the origin of *Euglena* chloroplasts), the *Euglena* lineage is more deeply rooted than the cyanobacterial branches. It may be noted that the *Euglena* sequence contains many insertions not present in GAPDH genes from other taxa.

The topology of the cytosolic clade is less well defined, and not all nodes are well supported. The high degree of similarity between the cytosolic sequences compromises reliable tree resolution and produces the polytomy and poorly supported branches observed in the parsimony and distance analyses, respectively. Neverthe-

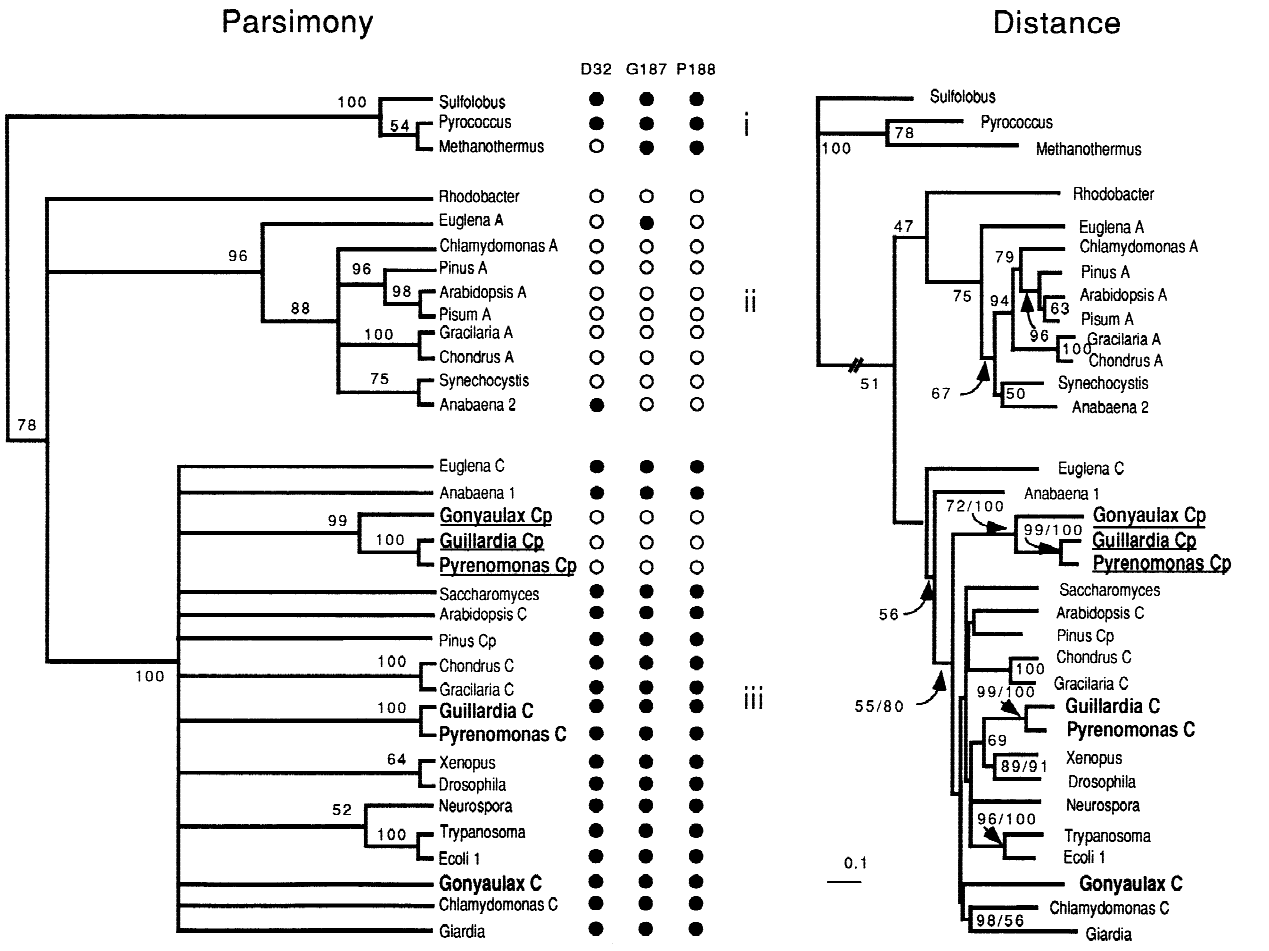

Fig. 2. Phylogenetic analysis of 33 GAPDH sequences generated by maximum parsimony (length = 2493 steps, CI = 0.606, RI = 0.629, RC = 0.381) and by distance analysis. Dinoflagellate and cryptomonad branches are shown in *boldface,* with the chloroplast sequences *underlined.* Both trees separate archaeal (i), plastisd (ii), and cytosolic (iii) sequences with similar topology. Bootstrap support is given as the number of times a given node appeared in 100 reiterations; those with values below 50 are shown as polytomies. Where two values are shown, the second was obtained from an analysis of the eukaryotic subset using *Anabaena* 1 as an outgroup. In the distance analysis the scale bar represents the number of substitutions per site. The branch length between the outgroup (i) and the rest of the tree is greater than can be represented by the diagram, indicated as a *break* in the line. *Filled circles* indicate the presence of aspartate[32], glycine[187], or proline[188] in a given sequence, indicative of a specificity for NAD^+, while substitutions in these positions (*open circles*) are indicative of a dual specificity and chloroplast localization.

less, apart from the *E. coli/T. brucei* clade, which has previously been observed as an outgroup to the vertebrate and plant clades, the branching order is consistent with a previously reported phylogenetic analysis (Martin et al. 1993). In particular, the relatively late branching of the cryptomonad cytosolic sequences has been previously documented (Liaud et al. 1997). Three chloroplast targeted sequences, which unexpectedly fall into the cytosolic clade, have been described previously. Two are the cryptomonad chloroplast GAPDH sequences, described above. The other, from the gymnosperm *Pinus sylvestris,* is thought to be derived from the introduction of a chloroplast-targeting sequence to the N terminal of the plant cytosolic isoform (Meyer-Gauen et al. 1994). The function of this isoform is unclear, as it lacks the necessary modifications for binding $NADP^+$ used in the Calvin cycle; it has been suggested that it may play a role in ''chlororespiration'' (Liaud et al. 1997).

Both parsimony and distance analyses place the NAD^+ specific *G. polyedra* GAPC sequence within the cytosolic clade (Fig. 2). Distance analysis shows branching close to and before *Giardia* and *Chlamydomonas* but the bootstrap support is weak. Reanalysis of the cytosolic clade, with the *Anabaena* 1 sequence as the outgroup, reveals the same overall topology but with better support values.

The phylogeny of the chloroplast form of the enzyme (GAPCp) places it outside the plastid clade, where most of the other plastid forms are found. In both analyses there is a well-supported grouping of the *G. polyedra* and cryptomonad GAPCp sequences within the cytosolic clade, well separated from the chloroplast forms of other eukaryotes (Fig. 2). The distance analysis shows this group branching deeply within this clade. Reanalysis of the cytosolic subset reveals the same branching order, with even higher support values (Fig. 2).

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
1	Sulfolobus	-	0.52	0.52	0.85	0.85	0.83	0.86	0.86	0.86	0.85	0.85	0.84	0.84	0.84	0.85	0.85	0.86	0.86	0.86	0.88	0.89	0.88	0.88	0.87	0.87	0.88	0.87	0.87	0.86	0.86	0.85	0.86	0.87
2	Pyrococcus	168	-	0.43	0.86	0.86	0.86	0.87	0.87	0.87	0.87	0.86	0.86	0.85	0.86	0.86	0.86	0.86	0.87	0.86	0.85	0.87	0.86	0.85	0.86	0.87	0.87	0.87	0.84	0.83	0.84	0.86	0.84	0.85
3	Methanothermus	168	137	-	0.86	0.85	0.85	0.86	0.87	0.87	0.86	0.85	0.85	0.84	0.83	0.85	0.87	0.86	0.87	0.85	0.85	0.86	0.85	0.85	0.86	0.87	0.86	0.86	0.85	0.83	0.86	0.85	0.86	0.84
4	Rhodobacter	274	273	275	-	0.50	0.48	0.48	0.49	0.48	0.51	0.51	0.47	0.48	0.52	0.50	0.58	0.59	0.60	0.53	0.52	0.53	0.53	0.53	0.58	0.57	0.55	0.53	0.54	0.52	0.51	0.56	0.54	0.57
5	Euglena A	276	275	275	162	-	0.35	0.33	0.35	0.34	0.40	0.39	0.38	0.41	0.53	0.51	0.54	0.57	0.56	0.51	0.56	0.54	0.55	0.54	0.58	0.57	0.58	0.55	0.53	0.55	0.52	0.58	0.50	0.59
6	Chlamydomonas A	271	274	276	157	116	-	0.20	0.21	0.21	0.26	0.25	0.29	0.32	0.54	0.51	0.57	0.55	0.55	0.49	0.53	0.54	0.52	0.52	0.57	0.55	0.57	0.55	0.55	0.51	0.51	0.54	0.51	0.58
7	Pinus A	278	277	280	156	110	67	-	0.13	0.11	0.26	0.27	0.27	0.32	0.52	0.48	0.56	0.56	0.55	0.52	0.54	0.53	0.53	0.53	0.57	0.57	0.56	0.54	0.55	0.52	0.52	0.56	0.50	0.57
8	Arabidopsis A	279	277	281	160	114	70	44	-	0.09	0.26	0.26	0.29	0.34	0.53	0.49	0.56	0.56	0.55	0.53	0.55	0.54	0.54	0.53	0.58	0.57	0.57	0.54	0.55	0.53	0.52	0.55	0.52	0.57
9	Pisum A	279	278	283	157	112	70	37	31	-	0.27	0.28	0.27	0.32	0.52	0.49	0.57	0.56	0.54	0.52	0.55	0.54	0.54	0.53	0.57	0.56	0.57	0.54	0.54	0.53	0.51	0.55	0.52	0.58
10	Gracilaria A	275	278	280	168	131	87	87	87	89	-	0.08	0.30	0.33	0.53	0.49	0.59	0.59	0.58	0.52	0.55	0.55	0.55	0.56	0.58	0.58	0.57	0.56	0.57	0.54	0.54	0.58	0.57	0.59
11	Chondrus A	275	275	276	168	129	82	88	86	91	25	-	0.30	0.33	0.52	0.50	0.58	0.59	0.58	0.52	0.54	0.55	0.56	0.56	0.57	0.58	0.57	0.56	0.56	0.53	0.54	0.59	0.56	0.60
12	Synechocystis	271	272	274	154	124	96	87	94	88	99	100	-	0.24	0.51	0.48	0.55	0.54	0.55	0.48	0.53	0.53	0.54	0.52	0.57	0.55	0.55	0.53	0.51	0.50	0.49	0.55	0.51	0.56
13	Anabaena 2	272	271	272	157	135	107	106	111	107	109	108	79	-	0.53	0.47	0.57	0.57	0.57	0.49	0.51	0.53	0.54	0.54	0.56	0.57	0.54	0.54	0.50	0.50	0.51	0.56	0.51	0.57
14	Euglena C	276	278	272	168	174	178	170	174	170	173	170	168	173	-	0.38	0.47	0.46	0.46	0.47	0.43	0.42	0.45	0.45	0.43	0.43	0.45	0.43	0.45	0.44	0.46	0.46	0.45	0.47
15	Anabaena1	273	273	275	160	165	165	157	158	158	158	162	154	154	124	-	0.41	0.44	0.43	0.40	0.40	0.34	0.39	0.37	0.38	0.38	0.37	0.38	0.40	0.38	0.38	0.42	0.39	0.45
16	Gonyaulax Cp	275	273	279	185	176	184	182	183	184	190	189	178	186	154	132	-	0.29	0.30	0.43	0.43	0.40	0.45	0.44	0.44	0.44	0.45	0.45	0.43	0.44	0.42	0.46	0.40	0.47
17	Guillardia Cp	281	278	281	191	187	182	184	185	183	193	195	177	187	151	142	95	-	0.11	0.42	0.43	0.43	0.42	0.41	0.42	0.44	0.44	0.43	0.43	0.44	0.42	0.43	0.43	0.49
18	Pyrenomonas Cp	283	280	285	194	185	182	182	182	179	190	191	180	187	153	141	99	35	-	0.41	0.42	0.44	0.43	0.43	0.41	0.42	0.44	0.43	0.43	0.44	0.41	0.43	0.43	0.50
19	Saccharomyces	282	278	279	171	167	159	171	173	171	172	171	155	160	155	130	141	139	134	-	0.31	0.34	0.32	0.33	0.39	0.36	0.34	0.36	0.37	0.35	0.31	0.44	0.34	0.44
20	Arabidopsis C	286	273	278	170	183	173	178	180	180	181	178	174	166	142	131	139	141	140	103	-	0.27	0.32	0.34	0.33	0.32	0.32	0.32	0.34	0.37	0.33	0.39	0.32	0.42
21	Pinus Cp	291	280	281	174	178	176	173	177	178	182	181	172	174	139	111	132	144	145	114	89	-	0.30	0.29	0.30	0.26	0.26	0.29	0.33	0.34	0.32	0.37	0.31	0.40
22	Chondrus C	288	277	278	174	181	171	174	179	177	180	185	175	179	150	126	147	140	144	107	106	98	-	0.15	0.35	0.31	0.32	0.33	0.36	0.37	0.32	0.39	0.34	0.40
23	Gracilaria C	287	272	277	173	178	170	174	174	175	184	184	171	177	148	120	143	137	142	109	111	97	48	-	0.34	0.32	0.32	0.31	0.35	0.36	0.32	0.38	0.33	0.39
24	Pyrenomonas C	284	279	281	188	190	188	189	191	188	190	189	185	184	143	123	145	141	136	129	110	99	115	114	-	0.13	0.28	0.30	0.33	0.36	0.36	0.39	0.36	0.44
25	Guillardia C	286	282	283	187	186	182	187	188	185	190	190	181	186	141	124	143	145	140	120	105	87	102	107	43	-	0.28	0.28	0.33	0.34	0.33	0.37	0.33	0.43
26	Xenopus	289	281	281	178	190	186	185	188	189	187	187	179	178	150	121	146	147	147	114	106	86	105	105	92	92	-	0.23	0.33	0.35	0.31	0.36	0.36	0.40
27	Drosophila	285	278	280	173	181	180	177	178	177	184	185	173	178	141	124	146	141	141	120	107	96	108	103	99	92	75	-	0.31	0.33	0.33	0.36	0.36	0.38
28	Neurospora	285	271	278	176	174	181	182	182	178	188	185	168	163	150	131	142	142	142	122	111	111	120	117	111	109	109	103	-	0.33	0.33	0.37	0.38	0.42
29	Trypanosoma	277	266	271	169	178	167	171	173	174	175	172	162	163	144	123	143	143	144	113	120	112	122	118	118	112	114	109	109	-	0.19	0.41	0.37	0.42
30	Ecoli1	278	268	278	166	169	167	169	169	167	177	177	158	166	152	123	136	138	135	101	109	106	105	104	117	110	103	109	109	61	-	0.41	0.35	0.41
31	Gonyaulax C	280	278	279	184	190	176	184	181	180	191	193	181	183	153	137	151	142	142	146	129	124	130	127	130	124	120	118	123	135	133	-	0.37	0.43
32	Chlamydomonas C	283	272	280	177	166	169	164	171	170	188	185	168	168	148	127	132	142	142	112	107	102	114	109	121	111	119	120	127	123	116	123	-	0.37
33	Giardia	285	273	275	185	194	191	187	189	190	196	198	185	188	157	147	154	162	166	145	139	131	134	128	147	144	131	126	139	138	135	141	123	-

Fig. 3. Pairwise distance between taxa. Absolute distances (below diagonal) and mean distances (above diagonal) were calculated for unordered characters using PAUP. *Bold boxes* correspond to the major clades resolved by both parsimony and distance analysis, being from top to bottom, the archaea, plastid, and cytosolic.

The position of *G. polyedra* GAPCp within the cytosolic clade raises the question of its origin. It is possible that, as in *Pinus* GAPCp, it arose by targeting a cytosolic-type isoform to the chloroplast, with amino acid changes allowing catalysis using both NAD^+ and $NADP^+$. This is unlikely because gene duplication followed by an extremely rapid divergence would be necessary to explain the distance between the GAPC and the GAPCp sequences. Also, the similarity between cryptomonad and dinoflagellate GAPCp sequences argues against both being derived independently from their respective GAPC homologs.

The origin of GAPCp in cryptomonads is proposed to be proteobacterial, while that of their chloroplasts is evidently cyanobacterial, as indicated by sequences of several plastid-encoded genes, including 16S rRNAs and *tuf*A (Delwiche et al. 1995; Douglas 1993; Giovannoni et al. 1993). In dinoflagellates RuBisCO is also reported to be proteobacterial in origin (Morse et al. 1995; Rowan et al. 1996; Watson and Tabita 1997), while the intrinsic peridinin–chlorophyll binding protein (iPCP) is of cyanobacterial origin [part of a larger gene family encompassing the fucoxanthin–chlorophyll proteins of the chromophytes and the chlorophyll a/b-binding proteins of the land plants (Durnford et al. 1996)]. What could explain the presence of proteobacterial-type isoforms in plastids of cyanobacterial origin?

For cryptomonads it was proposed that the GAPCp gene was laterally transferred from a proteobacterial-type cell to the nucleus during a primary endosymbiosis (Liaud et al. 1997). GAPCp was then targeted to the cyanobacterial-type plastid of a red algal eukaryote acquired during a second endosymbiotic event. It is also possible that the primary endosymbiont was a cyanobacterium that harbored two (or more) GAPDH isoforms. As reported by Martin et al. (1993), and is evident from our phylogenetic analysis, a GAPC isoform, which could be the GAPCp ancestor, occurs within the cyanobacteria (*Anabaena* 1 in Fig. 2).

While either mechanism (or any similar one) could also explain the presence of a proteobacterial GAPCp in dinoflagellates, it is unlikely to have occurred independently in both, given the sequence similarities. In addition, GAPC isoforms and rDNA phylogenies (Bhattacharya et al. 1992) indicate that dinoflagellates and cryptomonads are only distantly related, thus ruling out a common ancestral origin of the two groups as an explanation for the similarity between the GAPCp forms. A more likely possibility is that the *G. polyedra* GAPCp gene was acquired from a cryptomonad symbiont through lateral transfer.

Lateral gene transfer should be favored in situations where two genomes are closely associated. Such associations are commonly found in both heterotrophic and phototrophic dinoflagellates, many members of which are known to ingest and retain cells of other species. *Peridinium balticum* and *P. foliaceum*, for example, harbor diatoms with distinct nuclei and chloroplasts in a stable association (Chesnick et al. 1997). A more common, if more transient, association occurs in kleptochloroplasty, in which chloroplasts of an ingested cell are retained and used as temporary photosynthetic organelles by the dinoflagellate after endocytotic engulfment of prey. Indeed, in the dinoflagellates *Amphidinium poecilochroum*, *Gymnodinium aeruginosum*, and *Amphidinium wigrense*, the kleptochloroplasts are thought to be of cryptomonad origin (Larsen 1988; Schnepf 1992; Wilcox and Wedemayer 1985). We postulate that the GAPCp gene in the phototroph *G. polyedra* is derived from such an association, in which the integration of a cryptomonad-like cell or chloroplast into a heterotrophic dinoflagellate led to the transfer of a cryptomonad gene(s) to the dinoflagellate nucleus. However, our data are also consistent with such a transfer from a dinoflagellate to a cryptomonad.

Phylogenetic analysis of the three dinoflagellate nuclear-encoded chloroplast-targeted proteins so far examined, RuBisCO, iPCP, and GAPCp, suggests that the genome may be more chimeric than previously anticipated (Morse et al. 1995; Rowan et al. 1996; Watson and Tabita 1997). Lateral transfer from a cryptomonad in the course of an ancestral association with a dinoflagellate would account for the acquisition of two of these genes, GAPCp and iPCP, which have homologues in the cryptomonads, as the consequence of a single endosymbiosis.

Acknowledgments. This research was supported in part by grants from the NIH (GM RO1 19536) and ONR (00014-96-1-1118) (J.W.H.) and from the NSERC of Canada (D.M.). We thank Dr. Deborah Robertson for expert assistance in the preparation of the manuscript and Dr. T. Wilson for helpful comments.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Baroin A, Perasso R, Qu LH, Brugerolle G, Bachellerie JP, Adoutte A (1988) Partial phylogeny of the unicellular eukaryotes based on rapid sequencing of a portion of 28S ribosomal RNA. Proc Natl Acad Sci USA 85:3474–3478

Bhattacharya D, Medlin L, Wainwright PO, Ariztia EV, Bibear C, Stickel SK, Sogin ML (1992) Algae containing chlorophylls a + c are paraphyletic: Molecular evolutionary analysis of the chromophyta. Evolution 46:1801–1817

Biesecker G, Ieuan Harris J, Thierry JC, Walker JE, Wonacott AJ (1977) Sequence and structure of d-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus.* Nature 266:328–333

Brinkmann H, Cerff R, Salomon M, Soll J (1989) Cloning and sequence analysis of complementary DNAs encoding the cytosolic precursors of subunits GapA and GapB of chloroplast glyceraldehyde-3-phosphate dehydrogenase from pea and spinach. Plant Mol Biol 13:81–94

Cerff R (1995) The chimeric nature of nuclear genomes and the antiquity of introns as demonstrated by the GAPDH gene system. In: Go M, Schimmel P (eds) Tracing biological evolution in protein and gene structures. Elsevier Science, Amsterdam, pp 205–227

Chan RL, Keller M, Canaday J, Weil JH, Imbault P (1990) Eight small subunits of *Euglena* ribulose 1-5 biphosphate carboxylase/oxygenase are translated from a large mRNA as a polyprotein. EMBO J 9:333–338

Chesnick JM, Kooistra WHCF, 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

Clermont S, Corbier C, Mely Y, Gerard D, Wonacott A, Branlant G (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 CF, Kuhsel M, Palmer JD (1995) Phylogenetic analysis of tufA sequences indicates a cyanobacterial origin of all plastids. Mol Phylogenet Evol 4:110–128

Douglas SE (1993) Probable evolutionary history of cryptomonad algae. In: Lewin RA (ed) Origins of plastids. Chapman and Hall, New York, pp 265–290

Durnford DG, Aebersold R, Green BR (1996) The fucoxanthin-chlorophyll proteins from a chromophyte alga are part of a large multigene family: Structural and evolutionary relationships to other light harvesting antennae. Mol Gen Genet 253:377–386

Gajadhar A, Marquardt W, Hall R, Gunderaon J, Ariztiacarmona E, Sogin M (1991) Ribosomal-RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Crypthecodinium cohnii* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. Mol Biochem Parasit 45:147–154

Gibbs S (1981a) The chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae. Ann NY Acad Sci 361:193–208

Gibbs SP (1981b) The chloroplast endoplasmic reticulum: Structure, function, and evolutionary significance. Int Rev Cytol 72:49–99

Giovannoni SJ, Wood N, Huss V (1993) Molecular phylogeny of oxygenic cells and organelles based on small subunit ribosomal RNA sequences. In: Lewin RA (ed) Origins of plastids. Chapman and Hall, New York, pp 159–170

Grasse P, Dragesco J (1957) L'ultrastructure du chromosome des peridiniens et ses consequences genetiques. CR Herbd Seances Sci 245:2447–2452

Henze K, Badr A, Wettern M, Cerff R, Martin W (1995) A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. Proc Natl Acad Sci USA 92: 9122–9126

Holm-Hansen O (1969) Algae: amounts of DNA and carbon in single cells. Science 163:87–88

Jeffrey S, Seilicki M, Haxo F (1975) Chloroplast pigment patterns in dinoflagellates. J Phycol 11:374–384

Kishore R, Muchhal US, Schwartzbach SD (1993) The presequence of *Euglena* LHCPII, a cytoplasmically synthesized chloroplast protein, contains a functional endoplasmic reticulum-targeting domain. Proc Natl Acad Sci USA 90:11845–11849

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132

Larsen J (1988) An ultrastructural study of *Amphidinium poecilochroum* (Dinophyceae), a phagotrophic dinoflagellate feeding on small species of cryptophytes. Phycologia 27:366–377

Le QH, Markovic P, Hastings JW, Jovine RVM, Morse D (1997) Structure and organization of the peridinin chlorophyll protein gene in *Gonyaulax polyedra.* Mol Gen Genet 255:595–604

Lenaers G, Scholin C, Bhaud Y, Saint-Hillaire D, Herzog M (1991) A molecular phylogeny of dinoflagellate protists (*Pyrrophyta*) inferred from the sequence of 24S rRNA divergent domains D1 and D8. J Mol Evol 32:53–63

Liaud M-F, Brandt U, Scherzinger M, Cerff R (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 Suppl 44:S28–S37

Loeblich AI (1976) Dinoflagellate evolution: Speculation and evidence. J Protozool 23:13–28

Markovic P, Roenneberg T, Morse D (1996) Phased protein synthesis at several circadian times does not change protein levels in *Gonyaulax.* J Biol Rhythms 11:57–67

Martin W, Cerff R (1986) Prokaryotic features of a nucleus encoded enzyme: cDNA sequences for chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenases from mustard. Eur J Biochem 159:323–331

Martin W, Schnarrenberger C (1997) The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. Curr Genet 32:1–18

Martin W, Brinkmann H, Savonna C, Cerff R (1993) Evidence for a chimerical nature of nuclear genomes: Eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. Proc Natl Acad Sci USA 90:8692–8696

Meyer-Gauen G, Schnarrenberger C, Cerff R, Martin W (1994) Molecular characterization of a novel, nuclear-encoded, NAD^+-dependent glyceraldehyde-3-phosphate dehydrogenase in plastids of the gymnosperm *Pinus sylvestris* L. Plant Mol Biol 26:1155–1166

Morse D, Salois P, Marcovic P, Hastings JW (1995) A nuclear-encoded form II RuBisCo in dinoflagellates. Science 268:1622–1624

Page RDM (1996) TreeView: An application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358

Rizzo PJ (1981) Comparative aspects of basic chromatin proteins in dinoflagellates. Bio Systems 14:433–443

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

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schnepf E (1992) From prey via endosymbiont to plastid: comparative studies in dinoflagellates. In: Lewin RA (ed) Origins of plastids. Chapman and Hall, New York, pp 53–76

Sharif AL, Smith AG, Abell C (1989) Isolation and characterization of a cDNA clone for a chlorophyll synthesis enzyme from *Euglena gracilis.* Eur J Biochem 184:353–359

Shih MC, Lazar G, Goodman HM (1986) Evidence in favor of the symbiotic origin of chloroplasts: Primary structure and evolution of tobacco glyceraldehyde-3-phosphate dehydrogenase. Cell 47:73–80

Shih MC, Heinrich P, Goodman HM (1991) Cloning and chromosomal mapping of nuclear genes encoding chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenase from *Arabidopsis thaliana.* Gene 104:133–138

Shih MC, Heinrich P, Goodman HM (1992) Cloning and chromosomal mapping of nuclear genes encoding chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenase from *Arabidopisis thaliana.* Gene 119:317–319

Spector D, Vasconcelos A, Triemer R (1981) DNA duplication and chromosome structure in the dinoflagellates. Protoplasma 105:185–195

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

von Heijne G (1986) A new method for predicting signal sequence cleavage sites. Nucleic Acids Res 14:4683–4690

Watson GMF, Tabita FR (1997) Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A molecule for phylogenetic and enzymological investigation. FEMS Microbiol Lett 146:13–22

Wilcox LW, Wedemayer GJ (1985) Dinoflagellate with blue-green chloroplasts derived from an endosymbiotic eukaryote. Science 227:192–194

Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271