J Mol Evol (1997) 44(Suppl 1):S28–S37

JOURNAL OF MOLECULAR EVOLUTION



# Evolutionary Origin of Cryptomonad Microalgae: Two Novel Chloroplast/Cytosol-Specific GAPDH Genes as Potential Markers of Ancestral Endosymbiont and Host Cell Components

**Marie-Françoise Liaud,[1] Ulrike Brandt,[1] Margitta Scherzinger,[2,*] Rüdiger Cerff[1]**

[1] Institut für Genetik, Universität Braunschweig, D-38106 Braunschweig, Germany
[2] Institut für Biologie II, Universität Freiburg, D-79104 Freiburg, Germany

Received: 3 June 1996 / Accepted: 27 August 1996

**Abstract.** Cryptomonads are complex microalgae which share characteristics of chromophytes (chlorophyll c, extra pair of membranes surrounding the plastids) and rhodophytes (phycobiliproteins). Unlike chromophytes, however, they contain a small nucleus-like organelle, the nucleomorph, in the periplastidial space between the inner and outer plastid membrane pairs. These cellular characteristics led to the suggestion that cryptomonads may have originated *via* a eukaryote–eukaryote endosymbiosis between a phagotrophic host cell and a unicellular red alga, a hypothesis supported by rRNA phylogenies. Here we characterized cDNAs of the nuclear genes encoding chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenases (GAPDH) from the two cryptomonads *Pyrenomonas salina* and *Guillardia theta*. Our results suggest that in cryptomonads the classic Calvin cycle GAPDH enzyme of cyanobacterial origin, GapAB, is absent and functionally replaced by a photosynthetic GapC enzyme of proteobacterial descent, GapC1. The derived GapC1 precursor contains a typical signal/transit peptide of complex structure and sequence signatures diagnostic for dual cosubstrate specificity with NADP and NAD. In addition to this novel *GapC1* gene a cytosol-specific *GapC2* gene of glycolytic function has been found in both cryptomonads showing conspicuous sequence similarities to animal GAPDH. The present findings support the hypothesis that the host cell component of cryptomonads may be derived from a phototrophic rather than a organotrophic cell which lost its primary plastid after receiving a secondary one. Hence, cellular compartments of endosymbiotic origin may have been lost or replaced several times in eukaryote cell evolution, while the corresponding endosymbiotic genes (e.g., *GapC1*) were retained, thereby increasing the chimeric potential of the nuclear genome.

**Key words:** *Pyrenomonas salina, Guillardia theta* — Glyceraldehyde-3-phosphate dehydrogenase — Endosymbiotic gene transfer — Signal peptide — Enzyme engineering — Cell evolution

## Introduction

Photosynthetic eukaryotes, algae and land plants, can be classified into four major categories on the basis of their plastid complexity. Land plants and green, red, and glaucophyte algae have plastids surrounded by two membranes. *Euglenozoa* and most dinoflagellates contain three membrane plastids, while algae with chlorophyll c such as brown algae, diatoms, and other heterokont chromophytes have plastid envelopes composed of

The nucleotide sequence data (cDNAs) reported will appear in the DDBJ/EMBL/GenBank International Nucleotide Sequence Database under the accession numbers U40032 (GapC1, *Guillardia theta*), U40033 (CapC1, *Pyrenomonas salina*), U39873 (GapC2, *Guillardia theta*), and U39897 (GapC2, *Pyrenomonas salina*). Until recently *Guillardia theta* (original name *Cryptomonas theta*) has been erroneously called *Cryptomonas phi* due to a confusion of the two different organisms prior to the distribution of cultures to the international cryptomonad community (G. McFadden, publication in preparation)

* *Present address:* Lehrstuhl Pflanzenökologie und Systematik, Universität Bayreuth, D-95440 Bayreuth, Germany
*Correspondence to:* R. Cerff

four distinct membranes (Gibbs 1981). The unicellular cryptophytes and chlorarachniophytes may be considered as a separate fourth category. They also possess four membranes around their plastids which, however, are split into two pairs of membranes enclosing a periplastidial compartment which harbors 80S ribosomes, starch grains, and a vestigial nucleus, the nucleomorph (Greenwood et al. 1977).

It has been suggested for some time that the four categories of photosynthetic eukaryotes mentioned above represent phylogenetic chimeras with different evolutionary histories (for reviews see Gibbs 1981; Douglas 1992; Gray 1992; Sitte 1993; McFadden and Gilson 1995). Two membrane plastids have been proposed to result from one or several independent primary endosymbioses between a eukaryotic host cell and an ancient cyanobacterium. Phylogenetic analyses based on plastid and nuclear gene sequences from red algae and green plants support this view and furthermore suggest that red algae and green plants may be sister groups of monophyletic descendance (Liaud et al. 1993, 1994, 1995; Zhou and Ragan 1994; Ragan and Gutell 1995). Complex plastids with four membranes are thought to reflect eukaryote–eukaryote endosymbioses between a phagotrophic host cell and an ancient microalga. Cryptophytes and chlorarachniophytes, on the other hand, may be considered intermediate forms of complex algae which still contain remnants of the cytosol and nucleus of the photosynthetic secondary endosymbiont corresponding to the periplastidial compartment and nucleomorph, respectively.

In support of the hypothesis that cryptophytes are evolutionary chimeras of two distinct eukaryotic cells, it has been demonstrated that *Pyrenomonas salina* and *Guillardia theta* contain two phylogenetically separate 18S rRNA genes which are localized and actively transcribed in the nucleomorph and nucleus, respectively (Douglas et al. 1991; Maier et al. 1991). Incorporation of the two relatively distant rRNA sequences into phylogenetic trees suggested that the complex plastid together with the nucleomorph may be derived from an ancient red alga, while the host cell component of cryptomonads and its nucleus have been tentatively allied to green algae (Douglas et al. 1991) and zooflagellates (McFadden et al. 1994), respectively. To obtain more information concerning the elementary organisms ancestral to the complex cryptomonad cell we cloned and sequenced cDNAs encoding chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenases (GAPDH) from *Pyrenomonas salina* and *Guillardia theta*. We demonstrate that GAPDH genes of cryptomonads differ fundamentally from those encoding chloroplast and cytosolic GAPDH (GapAB and GapC) in green plants, red algae, and *Euglena*. Nuclei of cryptomonads harbor two novel *GapC* genes, one of which, *GapC1*, encodes an apparent chloroplast-specific enzyme which probably operates in the Calvin cycle as a functional equivalent of the classic photosynthetic enzyme GapAB.

## Materials and Methods

*Algae Material, Isolation of Poly (A)$^+$ mRNA, and Construction of cDNA Libraries.* *Pyrenomonas salina* and *Guillardia theta* were cultivated and harvested as previously described (Hansmann and Eschbach 1990; Rensing et al. 1994). For the purification of mRNAs from *Pyrenomonas salina* and *Guillardia theta* and the construction of the *Pyrenomonas salina* cDNA library see Müller et al. (1994). *Guillardia theta* cDNA was synthesized and cloned into the *Eco*RI site of NM1149 in essentially the same way as described by Liaud et al. (1994).

*Isolation and Sequencing of cDNA Clones.* The cDNA library from *Pyrenomonas salina* was screened by hybridization with the end-labeled 16-fold degenerate 16-mer oligodeoxynucleotide encoding the amino acid motif WYDNE(W/Y/F) (Martin et al. 1993a) conserved among all GAPDH genes known so far. Hybridization was performed at 32°C overnight in 6 × SSPE, 0.02% PVP, 0.02% Ficoll, 0.05% SDS, and the labeled oligonucleotide. Filters were washed twice at 32°C in the same buffer without radioactive probe. The cDNA library from *Guillardia theta* was screened using the random-prime-labeled cDNAs encoding GapC1 and GapC2 from *Pyrenomonas salina* previously identified by oligonucleotide screening. Hybridizations were performed at 55°C overnight in 3 × SSPE, 0.02% PVP, 0.02% Ficoll, 0.1% SDS, and the labeled probe. Filters were washed once in 6 × SSPE, 0.1% SDS and once in 2 × SSPE, 0.1% SDS at 55°C. Both libraries were also screened with the cDNA coding for chloroplast GAPDH (GapA) of *Chondrus crispus* (Liaud et al. 1994). Hybridizations were performed at 55°C as described above. cDNA insertions obtained from the purified lambda phage after digestion were subcloned into the *Eco*RI or *Not*I site of pBluescript-SK. The sequencing was performed by the dideoxy chain termination method (Pharmacia protocol) on double-stranded DNA using universal pSK or gene-specific primers.

*Primer Extension Experiments by a Modified RACE Technique.* Reverse transcription followed by dG-tailing and PCR amplification were performed essentially as described (Hirzmann et al. 1993) with the following modifications: 1 μg poly (A)$^+$-mRNA was heated to 90°C for 2 min prior to reverse transcription at 42°C for 1.5 h using an 18-mer oligod(T) primer or gene-specific primers fused to an *Eco*RI site and located downstream the initiation codon of the cDNAs encoding GapC1 of both cryptomonads (CTAGAATTCAGAGCCAGCGTTGAC for *Pyrenomonas salina* and CTAGAATTCGCGACCGTGGATGGAAT for *Guillardia theta*, respectively). The reaction was performed in the PCR buffer containing 40 mM Tris-HCl, pH 8.3, 47.5 mM KCl, 2.2 mM $MgCl_2$, 0.4 mM each of dNTPs. After dG-tailing, the first round of PCR (35 cycles) was performed using an *Eco*RI-oligod$(C)_{11}$ primer ($CTAGAATTC_{12}$) and the gene-specific distal GapC1 primers from *Pyrenomonas salina* and *Guillardia theta*, respectively (s. sequences above). A second round of PCR was performed using the same oligod(C) primer and proximal gene-specific primers (CATCTTAGCGGCCTTGG for *Pyrenomonas salina* and CTAGAATTCTTGATCATGATGC for *Guillardia theta*).

*Phylogenetic Data Analysis.* Glyceraldehyde-3-phosphate dehydrogenase sequences were obtained from the GenBank. The phylogenetic tree was constructed with the neighbor-joining method (Saitou and Nei 1987) from a matrix of amino acid distances based on an alignment of 322 amino acids. The alignment was done with the program CLUSTAL V (Higgins et al. 1992). The calculation of the distance matrix (Kimura's method), the phylogenetic analyses, and bootstrapping (100 cycles) were performed with the programs PROTDIST, NEIGHBOR, and SEQBOOT, respectively, of the program package PHYLIP 3.5C distributed by the editor (Felsenstein 1989).

## Results

### *Isolation and Characterization of cDNAs Encoding GapC1 and GapC2 from* Pyrenomonas salina *and* Guillardia theta

In an initial attempt to isolate cDNA clones encoding cytosolic and chloroplast GAPDH, the library from *Pyrenomonas salina* was screened with an oligonucleotide encoding the universal GAPDH motif WYDNE(W/Y/F) (see Materials and Methods). Ten hybridizing cDNA clones were purified and subcloned into the vector pBlueScript-SK. A first sequence analysis of these clones using the WYDNE oligonucleotide as sequencing primer revealed two groups of cDNAs encoding two different types of GapC proteins, GapC1 and GapC2. The terminal sequences from cDNA inserts larger than 1 kb were determined. Since none of these inserts represented a full-size cDNA, the library was rescreened under stringent conditions with homologous probes encoding GapC1 and GapC2, respectively. The longest positive clones for each GapC-type were subcloned and sequenced.

The cDNA encoding GapC1 of *Pyrenomonas salina* is 1,204 bp long; it contains the entire mature subunit coding region and a 3′ trailer of 61 bp without poly(A) tail. In addition, the GapC1 reading frame has a conspicuous in-frame 5′ extension encoding an N-terminal presequence. The cDNA for GapC2 of *Pyrenomonas salina* is 1,063 bp long; it contains the entire coding region of the mature subunit downstream codon No. 5 and a 3′ trailer of 58 bp with a potential poly(A) tail. Hybridization experiments with filters bearing total DNA and pure nucleomorph DNA (three minichromosomes of 195, 225, and 250 kb separated by pulsed field gel electrophoresis; Maier et al. 1991) of *Pyrenomonas salina,* respectively, confirmed that both cDNAs, GapC1 and GapC2, are encoded in the nucleus and not in the nucleomorph (data not shown).

The cDNA library from *Guillardia theta* was screened at low stringency with the heterologous cDNAs encoding GapC1 and GapC2 identified for *Pyrenomonas salina.* The longest positive clones obtained with each probe were subcloned and submitted to sequence analysis. The GapC1 cDNA of *Guillardia theta* contains the entire coding region including a 5′ extension homologous to that of *Pyrenomonas salina* GapC1 and a 3′ trailer of 64 nucleotides without poly(A) tail. The GapC2 cDNA of *Guillardia theta* is 1,000 bp long, starts with codon No. 9, and ends 2 nucleotides after the stop codon.

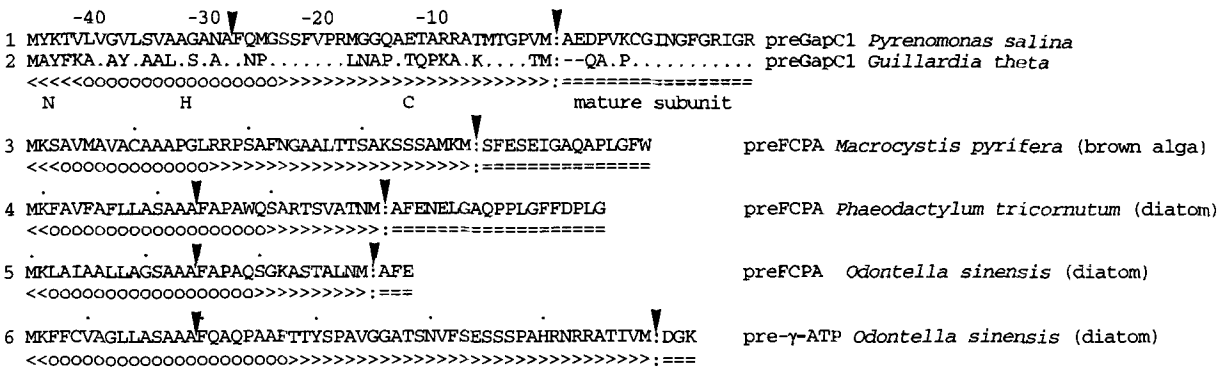
### *Characterization of 5′ Regions of mRNAs Encoding GapC1 and GapC2 from* Pyrenomonas salina *and* Guillardia theta

To obtain the 5′ terminal sequences of GapC1 and GapC2 mRNAs a modified RACE technique (reverse transcription, dG-tailing, PCR amplification; Hirzmann et al. 1993) was employed by using poly(A)$^+$–mRNA preparations from *Pyrenomonas salina* and *Guillardia theta.* For each mRNA and cryptomonad species several independent PCR products were sequenced, all of which had very short leader regions not longer than 15 nucleotides. Similar short 5′ untranslated regions had previously been described for nuclear genes of the diatom *Phaeodactylum tricornutum* (Bhaya and Grossman 1993).

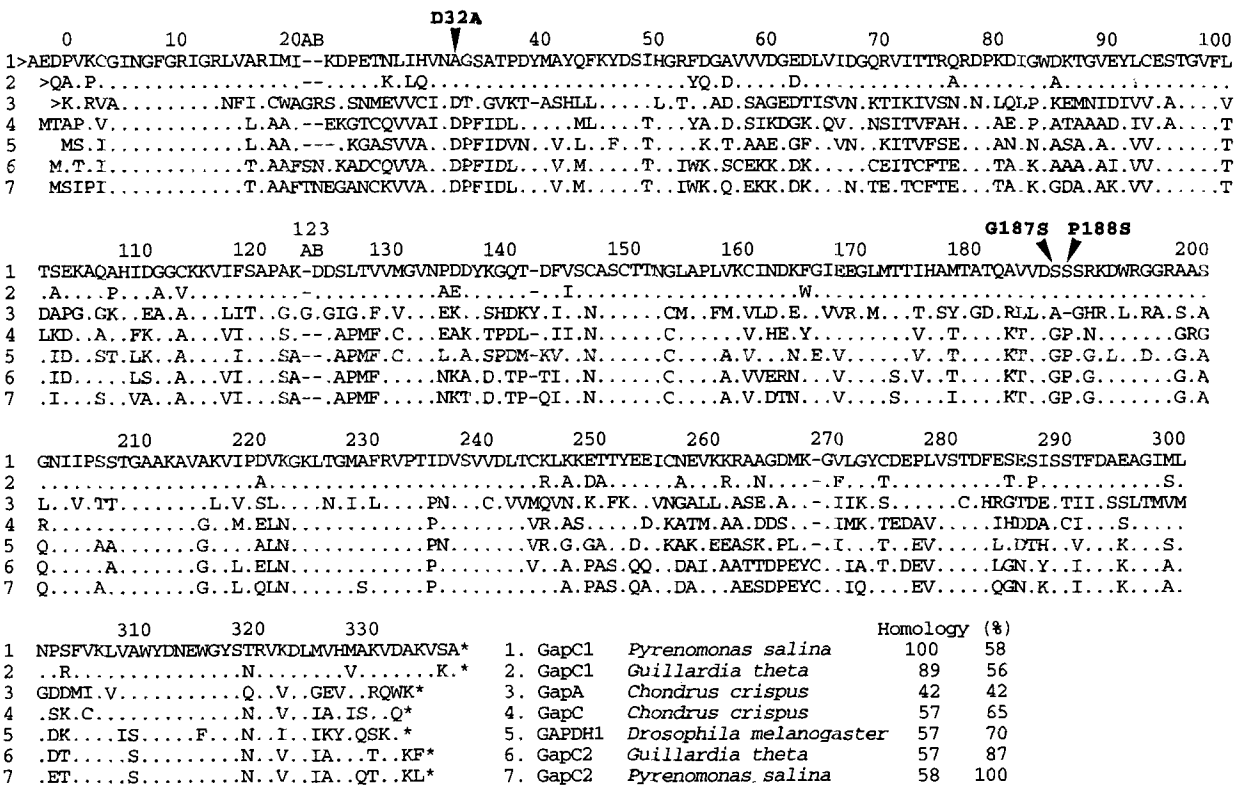
As shown in Fig. 1A, the derived N-terminal presequences of GapC1 from *Pyrenomonas salina* and *Guillardia theta* are identical in length (46 residues), homologous in sequence (41% identity), and, in addition, display some typical features of both secretory signal peptides and chloroplast transit peptides. As reviewed previously (von-Heijne 1988), signal peptides of secretory proteins from both prokaryotes and eukaryotes have a common design. They are typically between 15 and 30 residues long, have a hydrophobic central domain surrounded by a positively charged N-terminal region and a polar C-terminal domain, and a cleavage site conforming to the (−3, −1) rule predicting that residues −1 and −3 should be small and uncharged. The N-terminal extensions of GapC1 show a similar tripartite structure (elements N, H, and C in Fig. 1A). However, region C of the GapC1 presequence is at least twice as long as that of typical signal peptides and displays features which are characteristic of chloroplast transit peptides such as a clear preference for positively charged residues (K, R), hydroxy amino acids (S, T), and residues conferring structural flexibility (P, G) (von-Heijne and Nishikawa 1991). Considering that the N-terminal part of all native GAPDH enzymes is highly conserved at the level of primary and secondary structure (see Fig. 1B) the sequence–GPT(V)M:Q(A)– would seem to be a good candidate for the putative cleavage site of GapC1 from cryptomonads. It is similar to the previously suggested cleavage site –GPTM:KV– of the cytosolic precursor of chloroplast GapA from the marine red alga *Chondrus crispus* (Liaud et al. 1994). Apart from this, probably convergent, similarity, no further sequence homology is found relative to chloroplast GAPDH transit peptides from red algae (preGapA) and green plants (preGapA, preGapB, and preGapCp) (Brinkmann et al. 1989; Liaud et al. 1994; Meyer-Gauen et al. 1994).

Presequences of this bipartite architecture have previously been described for the cytosolic precursors of fucoxanthin chlorophyll proteins (FCPs) of diatoms (*Phaeodactylum tricornutum* and *Odontella sinensis*) and brown algae (*Macrocystis pyrifera*) and for the γ-subunit of ATPase from *Odontella sinensis* (sequences 3–6 in Fig. 1A) (Bhaya and Grossman 1991; Apt et al. 1995; Kroth-Pancic 1995). Diatoms and brown algae resemble cryptomonads in that their plastids are surrounded by an extra pair of ER membranes (although they do not contain a periplastidial compartment with

## A) GapC1 presequences




## B) GapC1 and GapC2 mature subunits




| | | | Homology | (%) |
|---|---|---|---|---|
| 1. | GapC1 | *Pyrenomonas salina* | 100 | 58 |
| 2. | GapC1 | *Guillardia theta* | 89 | 56 |
| 3. | GapA | *Chondrus crispus* | 42 | 42 |
| 4. | GapC | *Chondrus crispus* | 57 | 65 |
| 5. | GAPDH1 | *Drosophila melanogaster* | 57 | 70 |
| 6. | GapC2 | *Guillardia theta* | 57 | 87 |
| 7. | GapC2 | *Pyrenomonas salina* | 58 | 100 |

**Fig. 1.** **A** Alignment of derived preGapC1 signal/transit peptides from *Pyrenomonas salina* and *Guillardia theta*. The presequences are compared to those of the precursors of the fucoxanthin chlorophyll protein (preFCPA) and the ATPase gamma subunit (pre-γ-ATP) from diatoms and brown algae as indicated in the figure. Symbols ⟨⟨⟨, ooo, ⟩⟩⟩, and = = = designate N-terminal, hydrophobic, and C-terminal domains of the presequences and the beginning of the mature subunits, respectively. *Filled arrowheads* indicate the putative cleavage sites for signal peptides (motif–SAAA:F–) and transit peptides. Sources of sequences: *Pyrenomonas salina* and *Guillardia theta*, sequences 1 and 2, this paper; preFCPA of *Macroscystis pyrifera*, sequence 3 (Apt et al. 1995); preFCPA of *Phaeodactylum tricornutum*, sequence 4 (Bhaya and Grossman 1991); preFCPA and pre-γ-ATP of *Odontella sinensis*, sequences 5 and 6 (Kroth-Pancic 1995). **B** Alignment of amino acid sequences of mature subunits GapC1 and GapC2 from *Pyrenomonas salina* and *Guillardia theta*. The sequences of GapC1 and GapC2 are compared to those of GapA and GapC of the red alga *Chondrus crispus* (Liaud et al. 1994) and to GAPDH-1 of *Drosophila melanogaster* (Tso et al. 1985). Only amino acids not identical to the reference sequence (GapC1 of *Pyrenomonas salina*) are shown; identical residues are indicated by *dots*, indels by *dashes*. *Open arrowheads* at the beginning of sequences 1–3 indicate the presence of amino-terminal presequences (see Fig. 1A). D32A, G187S, and P188S indicate substitutions Asp32 → Ala32, Gly187 →Ser187, and Pro188 → Ser188, respectively, which are diagnostic for dual cosubstrate specificity with NADP and NAD in GAPDH enzymes (Clermont et al. 1993). Amino acid identities in percent relative to sequences GapC1 and GapC2 are given in *two columns* at the bottom of the figure.

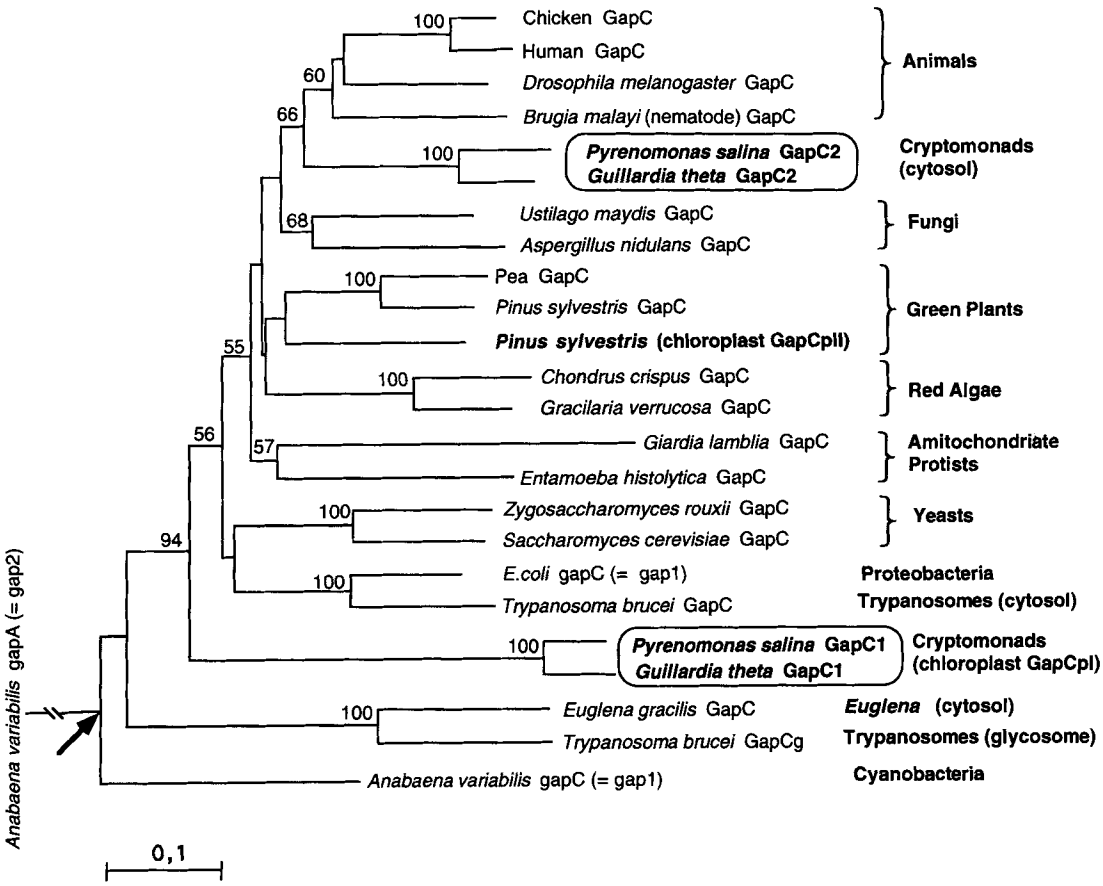



**Fig. 2.** Neighbor-joining tree for GapC protein sequences constructed from pairwise estimates of percent amino acid substitutions. For technical details see Materials and Methods. The tree was rooted using the amino acid sequence of the paralogous gene *gapA* (= *gap2*) from *Anabaena variabilis*. Bootstrap values above 50% are indicated at internal nodes. *Scale bar* indicates amino acid substitutions per site. Sources of sequences: NCBI database (see also Kersanach et al. 1994; Cerff 1995). The *arrow* indicates the point of divergence of cyanobacteria and proteobacteria.

nucleomorph, see Introduction) which polypeptide precursors have to traverse before entering the plastid. Four of the six presequences shown in Fig. 1A have in their hydrophobic domain the common sequence motif –SAAAF– corresponding to a putative signal peptide cleavage site (–SAAA:F–, see Discussion).

### *GapC1 and GapC2 Sequences of* Pyrenomonas salina *and* Guillardia theta *and Their Phylogenetic Positions on the GAPDH Tree*

The deduced polypeptide sequences GapC1 and GapC2 of the two cryptomonads are aligned together with the GapA and GapC sequences from the red alga *Chondrus crispus* and the GapC sequence of *Drosophila melanogaster* (see Fig. 1B). As shown at the bottom of the figure, the two cryptomonads are closely related to each other with respect to both GapC1 and GapC2 (89 and 87% sequence identity, respectively) while the differences between sequences GapC1 and GapC2 are considerable with only 56 to 58% sequence identity both for intra- and interspecific comparisons. However, if GapC1 and GapC2 are compared to the GapC sequences of red algae and animals a remarkable dichotomy becomes apparent. While GapC1 is equally distant to *Chondrus crispus* and *Drosophila* GapC (57%), GapC2 shows 65 and 70% sequence identity, respectively, indicating a possible relationship of cryptomonad GapC2 to higher eukaryotes *viz* ancestral animals. As expected, both genes show the largest differences compared to the paralogous gene *GapAB* encoding chloroplast GAPDH in red algae (42% sequence identity), green plants and *Euglena*.

In Fig. 2 a phylogenetic tree has been constructed on the basis of 24 GapC amino acid sequences from 19 different species (*E. coli, Anabaena variabilis,* and 17 eukaryotes) by using the neighbor-joining method (Saitou and Nei 1987). The tree has been rooted with the amino acid sequence of the paralogous gene *gapA* (= *gap2*) from *Anabaena variabilis*. The most striking feature of this tree is that GapC1 and GapC2 from cryptomonads specify two widely separated branches, probably corresponding to distinct organismal lineages (see Discussion). While GapC2 forms a common branch together with animals after their separation from fungi and plants (66% bootstrap coherence), GapC1 represents an early lineage on the universal GapC tree emerging at an

intermediate position, e.g., below the eukaryotic crown taxa plus *E. coli* (bootstrap value: 56%) and above *Euglena*/trypanosomes (bootstrap value: 94%) and the cyanobacterial lineage represented by *Anabaena variabilis*.

## Discussion

Phosphorylating chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenases (class I GAPDH; Cerff 1995) are marker enzymes of two metabolic pathways, the photosynthetic Calvin cycle and glycolysis, sharing a number of reactions that are catalyzed by distinct (iso-)enzymes unique to each. The two "heterotopic" GAPDH enzymes, GapAB and GapC, have been isolated from various higher plant species and were characterized with respect to protein structure and catalytic function (Cerff 1978, 1982; Cerff and Chambers 1979; Pupillo and Faggiani 1979). Both enzymes were shown to be encoded in the nucleus (Cerff and Kloppstech 1982) and the first GAPDH cDNA sequences from higher plants were reported over 10 years ago (Martin and Cerff 1986; Shih et al. 1986). Since then, cDNAs and genomic sequences encoding chloroplast and cytosolic GAPDH have been characterized from many photosynthetic eukaryotes including land plants, green and red algae, and recently also *Euglena gracilis* (Brinkmann et al. 1987, 1989; Quigley et al. 1988, 1989; Shih et al. 1988, 1992; Martinez et al. 1989; Liaud et al. 1990, 1993, 1994; Kersanach et al. 1994; Meyer-Gauen et al. 1994; Henze et al. 1995; for reviews see Martin et al. 1993b and Cerff 1995). Taken together, these investigations showed that all photosynthetic eukaryotes harboring plastids surrounded by two membranes and *Euglena* with three membrane plastids have a basically similar GAPDH gene composition of two highly divergent nuclear genes, *GapAB* and *GapC*, specifying chloroplast and cytosolic enzyme functions, respectively. Both types of genes are also present in cyanobacteria (*gapA* and *gapC*) and proteobacteria (only *gapC*), suggesting that their closely related eukaryotic counterparts are endosymbiotic acquisitions from chloroplasts/mitochondria (*GapAB/GapC* of higher eukaryotes) and from ancient eubacteria (*GapA* and *GapC* of *Euglena*/trypanosomes), respectively (Martin et al. 1993a; Henze et al. 1995; Cerff 1995).

Here we show that cryptomonads are the first photosynthetic eukaryotes which do not fit into this general scheme. Apparently in these complex microalgae the classic photosynthetic *GapAB* gene of cyanobacterial origin is functionally replaced by a chloroplast-specific *GapC* gene of proteobacterial descent, *GapC1*. In addition to this novel *GapC1* gene a cytosol-specific *GapC2* gene has been found showing conspicuous sequence similarities to animal GAPDH. The characterization of GapC1 and GapC2 cDNAs from two independent species, *Pyrenomonas salina* and *Guillardia theta*, documents their authentic cryptomonad origin and excludes the possibility that we are dealing with gene transcripts from associated organisms possibly contaminating our cDNA libraries.

### *Gene* GapC1 *of Cryptomonads Probably Encodes a Chloroplast Enzyme with Dual Cosubstrate Specificity Operating in the Calvin Cycle as a Functional Equivalent of the Classic Photosynthetic* GapAB

There are three lines of evidence suggesting that the deduced GapC1 enzyme of cryptomonads is a functional equivalent of chloroplast GapAB found in green plants, red algae, and *Euglena*. First, we were unable to identify GapAB transcripts in cryptomonad cDNA libraries. Screening experiments with 100,000 cDNA clones under nonstringent hybridization conditions with heterologous GapAB probes from red algae and green plants remained unsuccessful. This indicates that GapAB transcripts are either absent in cryptomonads or extremely rare, while GapAB mRNAs are expressed abundantly in all other photosynthetic eukaryotes so far analyzed.

Second, both deduced GapC1 polypeptides contain presequences typical for plastid proteins of complex algae displaying a bipartite structure composed of an apparent signal peptide followed by a chloroplast transit peptide (see Fig. 1A). This bipartite structure probably reflects a two-step mechanism of protein transport across the ER and the plastid envelope linked to a consecutive cleavage of the signal and transit peptide domains, respectively (Grossman et al. 1995). The contranslational import of *Phaeodactylum* preFCPA into the ER of canine microsomes has been demonstrated (Bhaya and Grossman 1991). Although intermediate processing has not yet been shown directly, it seems very likely to occur at the putative signal peptide cleavage site –SAAA:F– completely conserved in four of the six presequences shown in Fig. 1A.

Third, native GapC1 enzymes from cryptomonads are probably capable of using NADPH (in addition to NADH) for the generation of triosephosphate from 1,3-bisphosphoglycerate, which is the natural biochemical function of phosphorylating Calvin cycle GAPDH in all photosynthetic eukaryotes so far investigated including cryptomonads and other complex algae (Mateos and Serrano 1992). This conclusion is based on the striking observation that in GapC1 the three usually invariant residues Asp32, Gly187, and Pro188 known to determine NAD specificity in GapC enzymes by preventing NADP binding through steric interference with the 2′-$PO_4$ group have been substituted by Ala32, Ser187, and Ser 188. This natural "triple mutant" D32A-G187S-P188S (see Fig. 1B) is very similar to the in vitro engineered triple mutant D32A-L187A-P188S of the GAPDH-enzyme from *Bacillus stearothermophilus* displaying dual speci-

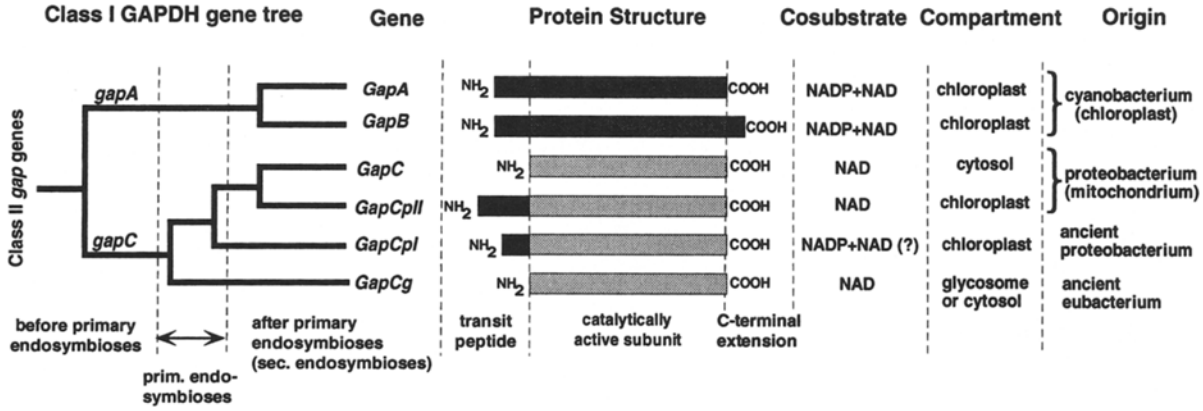



**Fig. 3.** Schematic representation of structure, cosubstrate specificity, cellular localization, and origin of eukaryotic GAPDH enzymes and their genes, respectively. Presequences of polypeptides GapA, GapB, GapCpI, and GapCpII and the carboxy-terminal extension of GapB are shown as *solid boxes*. Catalytic subunits are *shaded* to emphasize their common descendance from ancestral bacterial genes *gapA* and *gapC*, respectively. There are four different types of nuclear *GapC* genes, two of which (*GapCg* and *GapCpI*) may have originated by endosymbiotic gene translocations or by separations of bacterial lineages preceding these "cryptic" endosymbioses; the two others (*GapC/GapCpII*) are derived from a gene duplication event in green algae (see text).

ficity with NADP and NAD, in contrast to the *Bacillus* wild-type enzyme which is strictly NAD specific (Clermont et al. 1993).

### *Chloroplast-Specific GapC Enzymes Originated Twice Independently During Algae Evolution*

Until recently eukaryotic GapC enzymes were thought to be located exclusively in the cytosolic compartment of the cell as functional component of the glycolytic pathway (Fothergill-Gillmore and Michels 1993). The first exception was reported for trypanosomes which contain two largely different GapC enzymes, one located in the cytosol and the other in a microbody-like organelle, called the glycosome, harboring the major part of glycolysis from glucose to 3-phosphoglycerate in these organisms (Michels et al. 1991). Surprisingly, this glycosomal GapC of trypanosomes is closely related to *cytosolic* GapC of *Euglena gracilis* (Henze et al. 1995), suggesting a differential compartmentalization of GapC enzymes after the separation of trypanosomes and *Euglena*. The second exception was reported for *Pinus sylvestris* (Meyer-Gauen et al. 1994), which contains a nuclear-encoded chloroplast GapC enzyme, called GapCp, transported across the envelope with the aid of a typical transit peptide composed of 94 amino acids. This nuclear *GapCp* gene is highly expressed and has a wide distribution in gymnosperms and ferns (Jens Pahnke, personal communication). It encodes a NAD-specific GAPDH enzyme of unknown function, which, however, may play a potential role in the carbohydrate catabolism of the chloroplast, e.g., chlororespiration (Vermiglio et al. 1990; Singh et al. 1993). Incorporation of GapCp into the universal GapC tree (see Fig. 2 and Meyer-Gauen et al. 1994) suggests that it originated early in plant evolution, probably in green algae, in striking parallel to the *GapA/GapB* gene duplication.

Here we report a third exception, that of chloroplast GapC1 of cryptomonads which may be called GapCpI to distinguish it from GapCpII of gymnosperms and ferns. GapCpI and GapCpII are clearly independent acquisitions of chloroplasts as evidenced by their separate phylogenetic positions on the universal GapC tree corresponding to an early and late origin, respectively (see Fig. 2); their apparent different cosubstrate specificities; and their different presequences, which vary in architecture and length (46 vs 94 amino acids). These results clearly suggest that GapC enzymes which usually function in cytosolic glycolysis and which are probably of proteobacterial or ancient eubacterial origin (see below) have been recruited at least twice independently by chloroplasts during algae evolution to fulfill different metabolic functions. Hence, chloroplasts may be considered as biochemical chimeras with functional components of diverse evolutionary, e.g., cyanobacterial and proteobacterial, origin.

In Fig. 3 the schematic structures of the six different types of GAPDH polypeptides encoded in nuclear genomes of eukaryotes are shown as descendants of six separate gene lineages, four of which (*GapA, GapB, GapC, GapCpII*) arose by gene duplications before (*gapA/gapC*) and after (*GapA/GapB* and *GapC/GapCpII*) the endosymbiotic emergence of mitochondria and chloroplasts, while the other two may either represent endosymbiotic gene translocations from ancient eubacteria (*GapCg*) and ancient proteobacteria (*GapCpI*), respectively, or separations of bacterial lineages preceding these "cryptic" primary endosymbioses. GAPDH transit peptides for chloroplast import have been acquired three times independently, once for GapAB and twice for GapC, leading to GapCpI and GapCpII with apparent anabolic and catabolic functions, respectively. These multiple molecular adaptations of GapC raise interesting questions concerning the origin and evolution

of the Calvin cycle in complex algae and are particularly informative concerning questions of enzyme engineering and protein design with NAD(P)-dependent dehydrogenases (Clermont et al. 1993).

### *Genes* GapC1 *and* GapC2 *Probably Represent Separate Phylogenetic Lineages Ancestral to the Complex Cryptomonad Cell*

The apparent mitochondrial origin of *GapC* genes from higher eukaryotes is indicated in Fig. 2 by their close relationship to the *gapC* gene of *E. coli* and other proteobacteria (not shown in Fig. 2) generally considered as the free-living relatives of mitochondria (Gray 1993). On the other hand, *GapCpI* (= *GapC1*) seems to derive from a more ancient proteobacterial *GapC* gene, since it separates below this phylogenetic assemblage (bootstrap value 56%) but well above the division of cyanobacteria from proteobacteria (indicated by an arrow in Fig. 2) and also above the divergence of the lineage leading to *Euglena*/trypanosomes (bootstrap value of 94%). This interpretation proposing an independent endosymbiotic origin of the nuclear gene *GapCpI* unrelated to the emergence of mitochondria from higher eukaryotes is in agreement with the observation that gene *GapCpI* is absent in all eukaryotic crown taxa including green and red algae. It is also in agreement with theoretical predictions (Häuber et al. 1994) suggesting that the host cell component of cryptomonads may be derived from a phototrophic rather than an organotrophic cell which lost its primary plastid after receiving a secondary one. This model assumes that the protein import machinery for most essential photosynthetic functions was already established in the phototrophic host cell and was subsequently taken over and modified by the secondary endosymbiont. In this way a phototrophic host would be much better preadapted for the evolution of a complex plastid than a organotrophic protist which would need to evolve a four-membrane transfer machinery "from scratch," e.g., *via* translocation of multiple chloroplast-specific gene functions from the nucleus of the secondary endosymbiont into that of the host. An internuclear transfer of chloroplast-specific functional genes within the chimeric cell seems to be a rather unlikely possibility, in particular because nuclear genes, as opposed to plastid genes, do usually not occur in large copy numbers (for details see Häuber et al. 1994).

The present findings support the view that cryptomonads are, indeed, phylogenetic chimeras of two ancestral phototrophic microalgae, one of which, the secondary endosymbiont, may be related to present-day red algae (Douglas et al. 1991; Cavalier-Smith et al. 1994). For the presumptive phototrophic host cell no free-living relatives have so far been described, and it seems possible that some plastid-specific genes of this ancestral cell survived exclusively as components of present-day cryptomonads and perhaps other chromist algae with four-membrane plastids. Gene *GapC1* (*GapCpI*) suggests that the lost primary plastids of the host cell were related to ancient photosynthetic proteobacteria and, hence, may have performed nonoxygenic photosynthesis mediated by bacteriochlorophyll rather than by chlorophyll a. But where did the nucleus-cytosol component of this hypothetic photosynthetic host, the primary host cell, come from? It is surprising in this context that gene *GapC2* encoding cytosolic GAPDH in cryptomonads branches together with animals (66% bootstrap value, see Fig. 2), suggesting the involvement of an early animal protist in this chimeric conundrum. Whether or not *GapC2* may represent the primary host cell is difficult to decide at the present time, since we cannot exclude the possibility that this gene, and possibly also nuclear rRNA genes from cryptomonads (McFadden et al. 1994), are derived from yet another, tertiary, endosymbiosis (Cavalier-Smith et al. 1994). In this case gene markers of the primary host cell have yet to be discovered.

As summarized in Figs. 2 and 3, endosymbioses with subsequent intracellular translocations of *GapC* genes may have occurred at least three times independently during eukaryote evolution, thereby establishing three separate major lineages (leading to trypanosomes /*Euglena* (*GapCg;* Michels et al. 1991; Henze et al. 1995), cryptomonads (*GapCpI;* this study), and the eukaryotic crown taxa (*GapC-GapCpII;* Cerff 1995 and refs. therein), respectively. The gene encoding *cytosolic* GapC of trypanosomes with its extraordinary close affinity to *E. coli gapC* (see Fig. 2) is yet another case of a rather recent, possibly nonendosymbiotic, prokaryote → eukaryote *GapC* gene transfer (Michels et al. 1991; Martin et al. 1993a). In this context the protozoan parasites *Giardia lamblia* and *Entamoeba histolytica* merit particular attention, since they do not contain mitochondria but harbor relatively modern proteobacterial *GapC* genes (see Fig. 2), suggesting that they may have lost mitochondria (but retained mitochondrial *GapC*) as an adaptation to their parasitic lifestyle. From these examples it may be inferred that cellular compartments of endosymbiotic origin may have been lost or replaced many times in eukaryote cell evolution, while the corresponding endosymbiotic genes were retained, thereby increasing the chimeric potential of the nuclear genome.

*Acknowledgments.* We thank B. Martin (Braunschweig) for his help with the construction of the cDNA library from *Pyrenomonas salina*. This work was financially supported by the Deutsche Forschungsgemeinschaft (Ce 1/15-2).

## References

Apt KE, Clendennen SK, Powers DA, Grossman AR (1995) The gene family encoding the fucoxanthin chlorophyll proteins from the brown alga *Macrocystis pyrifera*. Mol Gen Genet 246:455–464

Bhaya D, Grossman AR (1991) Targeting proteins to diatom plastids

involves transport through an endoplasmic reticulum. Mol Gen Genet 229:400–404

Bhaya D, Grossman AR (1993) Characterization of gene clusters encoding the fucoxanthin chlorophyll proteins of the diatom *Phaeodactylum tricornutum*. Nucleic Acids Res 21:4458–4466

Brinkmann H, Martinez P, Quigley F, Martin W, Cerff R (1987) Endosymbiotic origin and codon bias of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize. J Mol Evol 26:320–328

Brinkmann H, Cerff R, Salomon M, Soll J (1989) Cloning and sequence analysis of cDNAs 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

Cavalier-Smith T, Allsopp MTEP, Chao EE (1994) Chimeric conundra: are nucleomorphs and chromists monophyletic or polyphyletic? Proc Natl Acad Sci USA 91:11368–11372

Cerff R (1978) Glyceraldehyde-3-phosphate dehydrogenase (NADP) from *Sinapis alba*: steady state kinetics. Phytochemistry 17:2061–2067

Cerff R (1982) Separation and purification of NAD- and NADP-linked glyceraldehyde-3-phosphate dehydrogenases from higher plants. In: Edelman M, Hallick RB, Chua NH (eds) Methods in chloroplast molecular biology. Elsevier Biochemical Press, Amsterdam, pp 683–694

Cerff R (1995) The chimaeric 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*. Proceedings of the 20th Taniguchi International Symposium, Division of Biophysics, held in Nagoya, Japan, 31 October–4 November 1994. Elsevier Science, Amsterdam, pp 205–227

Cerff R, Chambers SE (1979) Subunit structure of higher plant glyceraldehyde-3-phosphate dehydrogenases (EC 1.2.1.12 and EC 1.2.1.13). J Biol Chem 254:6094–6098

Cerff R, Kloppstech K (1982) Structural diversity and differential light control of mRNAs coding for angiosperm glyceraldehyde-3-phosphate dehydrogenases. Proc Natl Acad Sci USA 79:7624–7628

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

Douglas SE (1992) Eukaryote-eukaryote endosymbioses: insights from studies of a cryptomonad alga. Biosystems 28:57–68

Douglas SE, Murphy CA, Spencer DF, Gray MW (1991) Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. Nature 350:148–151

Felsenstein J (1989) PHYLIP: phylogeny interference package (version 3.2). Cladistics 5:164–165

Fothergill-Gilmore LA, Michels PAM (1993) Evolution of glycolysis. Prog Biophys Mol Biol 59:105–235

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

Gibbs SP (1981b) The plastids of some algal groups may have evolved from endosymbiotic algae. Ann N Y Acad Sci 36:193–207

Gray MW (1992) The endosymbiont hypothesis revisited. Int Rev Cytol 141:233–357

Gray MW (1993) Origin and evolution of organelle genomes. Curr Opin Gen Dev 3:884–890

Greenwood AD, Griffiths HB, Santore UJ (1977) Chloroplasts and cell compartments in Cryptophyceae. Br Phycol J 12:119

Grossman AR, Bhaya D, Apt KE, Kehoe DM (1995) Light-harvesting complexes in oxygenic photosynthesis: diversity, control and evolution. Annu Rev Genet 29:231–288

Hansmann P, Eschbach S (1990) Isolation and preliminary characterization of the nucleus and the nucleomorph of a cryptomonad. *Pyrenomonas salina*. Eur J Cell Biol 52:373–378

Häuber MM, Müller SB, Speth V, Maier U-G (1994) How to evolve a complex plastid?—A hypothesis. Bot Acta 107:383–386

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

Higgins DG, Bleasby AJ, Fuchs R (1992) CLUSTAL V: improved software for multiple sequence alignment. Comput Appl Biosci 8:189–191

Hirzmann J, Luo D, Hahnen J, Hobom G (1993) Determination of messenger RNA 5′-ends by reverse transcription of a cap structure. Nucleic Acids Res 21:3597–3598

Kersanach R, Brinkmann H, Liaud M-F, Zhang D-X, Martin WF, Cerff R (1994) Five identical intron position in ancient duplicated genes of eubacterial origin. Nature 367:398–389

Kroth-Pancic PG (1995) Nucleotide sequence of two cDNAs encoding fucoxanthin chlorophyll a/c proteins in the diatom *Odontella sinensis*. Plant Mol Biol 27:825–828

Liaud M-F, Zhang CX, Cerff R (1990) Differential intron loss and endosymbiontic transfer of chloroplast glyceraldehyde-3-phosphate dehydrogenase genes to the nucleus. Proc Natl Acad Sci USA 87:8918–8922

Liaud M-F, Valentin C, Brandt U, Bouget F-Y, Kloareg B, Cerff R (1993) The GAPDH gene system of the red alga *Chondrus crispus*: promotor structures, intron/exon organization, genomic complexity and differential expression of genes. Plant Mol Biol 23:981–984

Liaud M-F, Valentin C, Martin W, Bouget F-Y, Kloareg B, Cerff R (1994) The evolutionary origin of red algae as deduced from the nuclear genes encoding cytosolic and chloroplast glyceraldehyde-3-phosphate dehydrogenase from *Chondrus crispus*. J Mol Evol 38:319–327

Liaud M-F, Brandt U, Cerff R (1995) The marine red algae *Chondrus crispus* has a highly divergent beta-tubulin gene with a characteristic 5′-intron: functional and evolutionary implications. Plant Mol Biol 28:313–325

Maier U-G, Hofmann CJB, Eschbach S, Wolters J, Igloi GL (1991) Demonstration of nucleomorph encoded eukaryotic small subunit ribosomal RNA in cryptomonads. Mol Gen Genet 230:155–160

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

Martin W, Brinkmann H, Savona C, Cerff R (1993a) 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, Lydiate D, Brinkmann H, Forkmann G, Saedler H, Cerff R (1993b) Molecular phylogenies in angiosperm evolution. Mol Biol Evol 10:140–162

Martinez P, Martin W, Cerff R (1989) Structure, evolution and anaerobic regulation of a nuclear gene encoding cytosolic glyceraldehyde-3-phosphate dehydrogenase from maize. J Mol Biol 208:551–565

Mateos MI, Serrano A (1992) Occurrence of phosphorylating and non-phosphorylating NADP-dependent glyceraldehyde-3-phosphate dehydrogenases in photosynthetic organisms. Plant Sci 163–170

McFadden G, Gilson P (1995) Something borrowed, something green: lateral transfer of chloroplasts by secondary endosymbiosis. TREE 10:12–17

McFadden GI, Gilson PR, Hill DRA (1994) Eur J Phycol 29:29–32

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

Michels PA, Marchand M, Kohl L, Allert S, Wierenga RK, Opperdoes FR (1991) The cytosolic and glycosomal isoenzymes of glyceraldehyde-3-phosphate dehydrogenase in *Trypanosoma brucei* have a distant evolutionary relationship. Eur J Biochem 198:421–428

Müller S, Rensing SA, Martin WF, Maier U-G (1994) cDNA cloning of a Sec61 homologue from the cryptomonad alga *Pyrenomonas salina*. Curr Genet 26:410–414

Pupillo P, Faggiani R (1979) Subunit structure of three glyceraldehyde 3-phosphate dehydrogenases of some flowering plants. Arch Biochem Biophys 194:581–592

Quigley F, Martin WF, Cerff R (1988) Intron conservation across the prokaryote-eukaryote boundary: structure of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize. Proc Natl Acad Sci USA 85:2672–2676

Quigley F, Brinkmann H, Martin WF, Cerff R (1989) Strong functional GC pressure in light-regulated maize gene encoding subunit GapA of chloroplast glyceraldehyde-3-phosphate dehydrogenase: implications for the evolution of GapA pseudogenes. J Mol Evol 29: 412–421

Ragan MA, Gutell RR (1995) Are red algae plants? Bot J Linn Soc 118:81–105

Rensing SA, Goddemeier M, Hofmann CJB, Maier UG (1994) The presence of a nucleomorph *hsp*70 gene is a common feature of Cryptophyta and Chlorarachniophyta. Curr Genet 26:451–455

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

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

Shih M-C, Heinrich P, Goodman HM (1988) Intron existence predated the divergence of eukaryotes and prokaryotes. Science 242:1164–1166

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

Singh KK, Chen C, Epstein DK, Gibbs M (1993) Respiration of sugars in spinach (*Spinacia oleracea*), maize (*Zea mays*), and *Chlamydomonas reinhardtii* F-60 chloroplasts with emphasis on the hexose kinases. Plant Physiol 102:587–593

Sitte P (1993) Symbiogenic evolution of complex cells and complex plastids. Eur J Protist 29:131–143

Tso JY, Sun X-H, Wu R (1985) Structure of two *Drosophila melanogaster* glycerinaldehyde-3-phosphate dehydrogenase genes. J Biol Chem 260:8220–8228

Vermiglio A, Ravenel J, Peltier G (1990) Chlororespiration: a respiratory activity in the thylakoid membrane of microalgae and higher plants. In: Wiessner W, Robinson DG, Starr RC (eds) Cell walls and surfaces, reproduction, photosynthesis. Springer-Verlag, Berlin, pp 188–205

von-Heijne G (1988) Transcending the impenetrable: how proteins come to terms with membranes. Biochim Biophys Acta 947:307–333

von-Heijne G, Nishikawa K (1991) Chloroplast transit peptides. The perfect random coil? FEBS Lett 278:1–3

Zhou Y-H, Ragan MA (1994) Cloning and characterization of the nuclear gene encoding glyceraldehyde-3-phosphate dehydrogenase from the marine red alga *Gracilaria verrucosa*. Curr Genet 26:79–86