J Mol Evol (1994) 38:319–327

Journal of Molecular Evolution

The Evolutionary Origin of Red Algae as Deduced from the Nuclear Genes Encoding Cytosolic and Chloroplast Glyceraldehyde-3-Phosphate Dehydrogenases from *Chondrus crispus*

Marie-Françoise Liaud,[1] Christiane Valentin,[1] William Martin,[1] François-Yves Bouget,[2] Bernard Kloareg,[2] Rüdiger Cerff[1]

[1] Institut für Genetik, Universität Braunschweig, Spielmannstr. 7, D-38106 Braunschweig, Germany
[2] Centre d'Etudes Océanologiques et de Biologie Marine, CNRS-UPR 4601, BP 74, F-29682 Roscoff Cedex, France

Received: 24 February 1993 / Revised: 13 June 1993

Abstract. Algae are a heterogeneous group of photosynthetic eukaryotes traditionally separated into three major subdivisions: rhodophytes, chlorophytes, and chromophytes. The evolutionary origin of rhodophytes or red algae and their links to other photosynthetic and nonphotosynthetic eukaryotes have been a matter of much controversy and speculation. Here we present the first cDNAs of nuclear protein genes from red algae: Those encoding cytosolic and chloroplast glyceraldehyde-3-phosphate dehydrogenases (GAPDH) from *Chondrus crispus*. A phylogenetic analysis including GAPDH gene sequences from a number of eukaryotic taxa, cyanobacteria, and purple bacteria suggests that chloroplasts and rhodoplasts together form a monophyletic group of cyanobacterial descent and that rhodophytes separated from chlorophytes at about the same time as animals and fungi. The composite GAPDH tree further demonstrates that chloroplast and cytosolic GAPDH genes are closely related to their homologs in cyanobacteria and purple bacteria, respectively, the presumptive ancestors of chloroplasts and mitochondria, thereby firmly establishing the endosymbiotic origin of these nuclear genes and their fixation in eukaryotic cells before the rhodophyte/chlorophyte separation. The present data are in conflict with phylogenetic inferences based on plastid-encoded *rbc*L sequences supporting a polyphyletic origin of rhodoplasts and chloroplasts. Comparison of *rbc*L to GAPDH phylogenies suggests that *rbc*L trees may be misleading because they are composed of branches representing ancient duplicated (paralogous) genes.

Key words: Molecular phylogeny of plastids and mitochondria — Endosymbiotic origin — Monophylesis — Cyanobacteria — Purple bacteria — Transit peptide — Neighbor joining method — Bootstrapping

Introduction

Algae are a heterogeneous assemblage of photosynthetic eukaryotes which occur as unicellular protists (microalgae) or multicellular thallophytes (macroalgae) and which are traditionally divided into three major taxonomic groups: Chlorophytes (green algae), rhodophytes (red algae), and chromophytes (chlorophyll a+c–containing plants including the brown algae). From an evolutionary and genetic point of view algae are rather complex organisms because of their "intertaxonic" and polyphyletic origin, comprising multiple intracellular symbioses between one or several ancient eukaryotic host cells and different eubacterial and/or eukaryotic endosymbionts which gave rise to plastids (chloroplasts, rhodoplasts, chromoplasts) and to mitochondria. This implies that molecular phylogenies of algae based on plastid- or mitochondria-specific genes (nucleus or organelle encoded) may be different from

Correspondence to: R. Cerff

those inferred from host-cell-derived genes. (For reviews see Doolittle 1988; Gray 1989; Cavalier-Smith 1991; Martin et al. 1992.)

Red algae present a typical example of this taxonomic dilemma. Based on morphological and cytological criteria rhodophytes seem to be monophyletic descendants of an old eukaryotic lineage with distant relationships to other photosynthetic and nonphotosynthetic eukaryotes (Margulis and Schwartz 1982). They have neither flagella nor centrioles at any stage of their life cycle and their plastids are "primitive"—having unstacked thylakoids which contain no chlorophyll b but which contain phycobilin pigments within phycobilisomes similar to cyanobacteria (Scagel et al. 1982). In a recent cladistic analysis by Lipscomb (1989) based on 36 eukaryotic taxa and 77 separate cell characteristics (excluding molecular sequence data), red algae are positioned as an eukaryotic outgroup defining the root of the universal eukaryotic tree. Molecular sequence trees based on host-cell-specific gene markers are incongruent. Hori and Osawa (1987), using 5S rRNA sequences, suggest that red algae comprise the first diverging eukaryotic branch. However, phylogenies inferred from small (18S)- and large (28S)-subunit rRNAs (Perasso et al. 1989; Bhattacharya et al. 1990; Hendriks et al. 1991) seem to indicate that rhodophytes separated at about the same time as all other major eukaryotic groups but as a lineage distinct from those of green and brown algae, supporting the polyphyletic origin of algal host cells.

Similar discrepancies have been reported concerning the origin of rhodoplasts based on the plastid-specific gene markers for 16S rRNA and *rbc*S/L encoding the small and large subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase (Martin et al. 1992, Ueda and Shibuya 1992). While 16S rRNA phylogenies (Douglas and Turner 1991) seem to suggest that rhodoplasts, like chloroplasts, are derived from cyanobacteria, *rbc*S/L sequences from rhodoplasts (Douglas et al. 1990; Valentin and Zetsche 1990; Morden and Golden 1991) show surprisingly close affinities to the homologous sequences of alpha and beta purple bacteria (sensu Woese, Woese 1987) which are distantly related to cyanobacteria in terms of 16S rRNA sequences.

Here we report the first phylogenetic analysis based on nuclear protein genes from rhodophytes. We present full-length cDNA sequences for cytosolic and chloroplast glyceraldehyde-3-phosphate dehydrogenases (GAPDH, phosphorylating) from *Chondrus crispus*. The nuclear GAPDH gene system is particularly suited to investigate the origin of algae, since it contains two independent evolutionary markers, each comprising over 700 nonsynonymous nucleotide positions which are homologous, highly conserved, and colinear in all organisms: A universal gene marker encoding glycolytic GAPDH (EC 1.2.1.12, gene *GapC*) present in the nuclei of *all* eukaryotes, and a *plant*-specific gene marker encoding Calvin-cycle GAPDH (EC 1.2.1.13, related genes *GapA* and *GapB;* for GAPDH nomenclature see Materials and Methods), restricted to the nuclei of photosynthetic eukaryotes (Cerff 1982; Martin and Cerff 1986; Shih et al. 1986; Brinkmann et al. 1987, 1989; Martin et al. 1989; Smith 1989; Doolittle et al. 1990; Michels et al. 1991; Fothergill-Gillmore and Michels 1992; Martin et al. 1992).

In a recent publication (Martin et al. 1993b) we demonstrated that both types of genes, *GapA/GapB (GapA/B)* and *GapC,* also occur in eubacteria, indicating that they were gained by lateral transfer from the ancestors of chloroplasts (cyanobacteria, *GapA/B*) and mitochondria (purple bacteria, *GapC*), respectively. The present findings suggest that rhodoplasts and chloroplasts have a common origin and that red algae occupy a relatively late branch on the eukaryotic tree which separated from green plants at about the same time as fungi and animals and after the gene transfer events leading to the fixation of *GapA/B* and *GapC* in the nuclei of these organisms.

Materials and Methods

Plant Material. Gametophytes of *Chondrus crispus* (Stackh.) were provided by Sanofi-Bio-Industrie (Carentan, France). The algae were grown in running seawater under controlled culture conditions (light intensity: 60 $\mu E \times m^{-2} \times s^{-1}$; 12-h-light: 12-h-dark photoperiod; temperature ranging from 10°C in winter to 17°C in summer).

Isolation of Poly (A)$^+$ mRNA and Construction of cDNA Library. Total nucleic acids were extracted from protoplasts of *C. crispus* apical tips. Protoplasts were prepared as previously described (Le Gall et al. 1990). For total RNA, the protoplast pellet was resuspended in a lysis buffer containing 4 M guanidinium thiocyanate, 25 mM sodium citrate, 0.5% sodium lauryl sarcosinate, and 0.1 M mercaptoethanol, and the extract was centrifuged in a cesium chloride gradient (C. Passaquet, personal communication). The RNA pellet was resuspended in TE buffer (10 mM Tris/HCl, 1 mM EDTA, pH 8.0) and the total RNA was further purified by a phenol extraction followed by a precipitation with LiCl (2 M final concentration). The poly (A)$^+$-containing mRNA fraction was purified from total RNA by adsorption to oligo(dT) cellulose. cDNA was synthesized using a cDNA synthesis kit (Pharmacia, standard protocol) and cloned into the *Eco*RI site of the expression vector λgt11. The resulting phage DNA was packaged in vitro into phage particles using packaging extracts prepared according to Hohn (Hohn 1979) and plated with *Escherichia coli* strain Y1090 (Young and Davis 1983).

Isolation of cDNA Clones. The cDNA library was screened by hybridization with the random-prime-labeled (Feinberg and Vogelstein 1984) cDNAs coding for GAPA and GAPC of maize (inserts of plasmids pZm9 and pZm57, respectively; Brinkmann et al. 1987). The hybridization was performed at 55°C overnight in 6 × SSPE, 0.2% PVP, 0.2% Ficoll, 0.1% SDS, 0.5 μg/ml denatured salmon sperm DNA, and the labeled probe. Filters were washed twice in 2 × SSPE, 0.1% SDS at 55°C for 20 min. All cDNA insertions obtained from the purified lambda phage after digestion were cloned into the *Eco*RI site of the plasmid vector BlueScript (Stratagene).

DNA Sequence Analysis. An ordered set of deletion clones was prepared for each clone using exonuclease III and following the Strat-

agene protocol. The deletion clones were sequenced by the dideoxy chain termination method as double-stranded DNA with T7 polymerase (Pharmacia protocol). When required, oligonucleotides (17-mers) were synthesized according to sequence information and used directly as primers for further sequencing.

Phylogenetic Data Analysis. Deduced amino acid sequences were aligned with the GCG package (Devereux et al. 1984). Phylogenetic trees were constructed by the neighbor-joining method (Saitou and Nei 1987), which has been shown to be very efficient in recovery of the correct topology under a variety of sequence parameters (Nei 1991). Distances were measured as nucleotide divergence at nonsynonymous sites (d_N, Nei and Gojobori 1986) calculated on the basis of an average 990 homologous nucleotide positions (average 760 nonsynonomous sites). Reliability of the distance matrix tree was estimated through comparison to bootstrap parsimony trees (DNABOOT of PHYLIP, Felsenstein 1985) performed on a data set consisting of first and second codon positions for the 21 sequences (734 sites each). This procedure approximates the removal of synonymous sites and improves the parsimony result when divergence between sequences is large (Martin et al. 1992).

Nomenclature of the GAPDH Genes from Plants and Eubacteria. Proteins corresponding to subunits A and B of chloroplast GAPDH (Calvin cycle) and subunit C of cytosolic GAPDH (glycolysis) are specified as GAPA, GAPB, and GAPC, respectively; products from different members of the same gene family are numbered consecutively, e.g., GAPA1, GAPA2, . . . , GAPB1, GAPB2, . . . , GAPC1, GAPC2, . . . , etc. The corresponding genes are designated *GapA1, GapA2, . . . , GapB1, GapB2, . . . , GapC1, GapC2, . . . , etc. GAPDH genes from eubacteria are numbered consecutively, e.g., gap1, gap2, gap3,* etc. The plant nomenclature corresponds to that previously proposed (Liaud et al. 1990) except that plant genes are now designated by the same four letters as their corresponding products and that the last letter designating the subunit type has been capitalized (i.e., *GapA1* instead of *Gpa1*).

Results

Isolation and Characterization of cDNAs Encoding GAPC and GAPA from Chondrus crispus

An amplified cDNA library (based on 2×10^6 original recombinants cloned into the vector lambda gt11, see Materials and Methods) was screened at low stringency with heterologous cDNAs encoding subunits GAPC and GAPA of maize, respectively (Brinkmann et al. 1987). For each probe the longest positive clone was isolated, subcloned into the *Eco*RI site of the vector BlueScript, and submitted to sequence analysis. The cDNA of *Chondrus crispus GapC* (recombinant plasmid pCc*GapC1*) is 1,165 bp long and contains the entire coding region (1,008 bp), a 5′ leader of 99 bp, a 3′ trailer of 58 bp, and no poly(A) tail. The cDNA of *Chondrus crispus GapA* (recombinant plasmid pCc*GapA1*) is 1,655 bp long, comprising the entire coding region (228 bp for the transit peptide and 1,017 bp for the mature subunit), a long 5′ leader of 307 bp, and a 3′ trailer of 103 bp. The frequencies of the two cDNA clones encoding GAPC and GAPA were about 0.1 and 0.01%, respectively, suggesting that *GapC* was more strongly expressed than *GapA* in *Chondrus crispus* protoplasts.

GapC and GapA cDNA Sequences of Chondrus crispus and the Evolutionary Position of Red Algae

The deduced polypeptide sequences of *Chondrus crispus GapC* and *GapA* are aligned in Fig. 1 together with 15 other GAPDH sequences from plants, animals, fungi, and bacteria. As shown at the bottom of the figure, *Chondrus crispus* GAPC has between 67 and 69% amino acid identity with the other eukaryotic glycolytic GAPDH sequences. The similarities to the eubacterial genes *gap1* of *E. coli* and *Anabaena variabilis* are 67% and 60%, respectively. *Chondrus crispus* GAPA is 71% identical to angiosperm chloroplast GAPDH (GAPA and GAPB), 67% identical to *Anabaena* GAPDH (gene *gap2,* Martin et al. 1993b), and only 43% identical to its glycolytic counterpart in the same cell. The GAPDH gene characterized for *Rhodobacter sphaeroides* (a photosynthetic α purple bacterium) (Gibson and Tabita 1988) has 46 and 48% amino acid similarity with *Chondrus crispus* GAPC and GAPA, respectively.

In Fig. 2 a phylogenetic tree has been constructed by means of the neighbor-joining method (Saitou and Nei 1987) on the basis of 21 GAPDH nucleotide sequences from 10 eukaryotes and two eubacteria (*Anabaena variabilis, E. coli*). Bootstrap parsimony analysis (100 replicates, DNABOOT of the PHYLIP package; Felsenstein 1985) was performed with the same sequences after removal of all third codon positions leaving 736 first and second codon positions. The topology of the *GapA* subtree suggests a relatively close relationship of *Chondrus crispus GapA* to genes *GapA* and *GapB* from higher plants. This common branch is separated from *Anabaena variabilis gap2* by a bootstrap value of 95. On the *GapC* subtree the branching order of the major eukaryotic lineages cannot be definitely resolved, although the common branching of *Chondrus crispus* with green plants relative to animals and fungi is supported by a bootstrap value of 58.

The Transit Peptide of Chondrus crispus GapA

In view of the well-documented evolutionary history of *GapA/GapB* transit peptides (see Discussion), it is interesting to note that the transit peptide of *Chondrus crispus GapA,* a gene considerably older than the *GapA/GapB* gene duplication (see Fig. 2), has retained up to 15% amino acid sequence similarity relative to *GapA* and *GapB* of angiosperms (Fig. 3). However, the proteolytic cleavage site AlalLys (Brinkmann et al. 1989) is not conserved in *Chondrus crispus.* The sequence -Pro-Thr-MetlLys- would seem to be the most likely candidate with the closest affinity to the motif

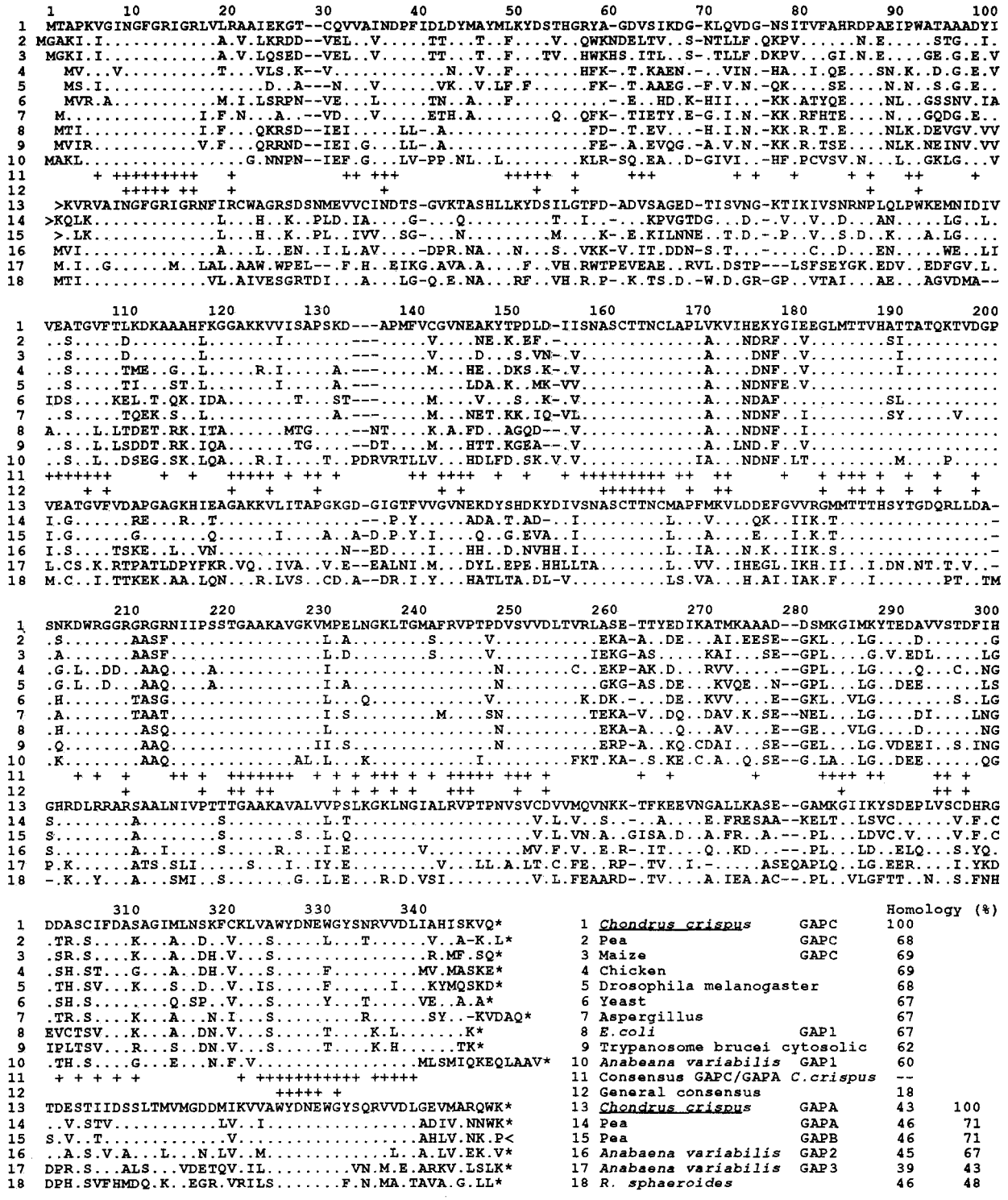

Fig. 1. Amino acid sequence alignment of 16 GAPDH sequences, comprising eight eukaryotic GAPC, three eukaryotic GAPA/GAPB, and five eubacterial GAPDH polypeptides from 11 different species as specified at the bottom of the figure. The sequences are presented in two blocks, a GAPC block (lines 1–10) and a GAPA block (lines 13–18). The sequences in each block are compared to *Chondrus crispus* GAPC and GAPA, respectively, the only sequences written in full. Only amino acids not identical to these reference sequences are shown in either group. The first and the last residues of each sequence are shown irrespective of homology. Stop codons are symbolized by *asterisks*. Amino-terminal (transit peptides of GAPA and GAPB) and carboxy-terminal (GAPB) extensions are not shown but their presence is indicated by *arrowheads*. Lines 11 and 12: *Plus signs* designate conserved residues between GAPC/GAPA of *Chondrus crispus* and all species (general consensus), respectively. Sources of sequences: *Chondrus crispus,* sequences 1 and 13; this paper; pea, sequences 2, 14, 15 (Brinkmann et al. 1989; Martin et al. 1993a); maize, sequence 3 (Brinkmann et al. 1987); *E. coli,* sequence 8 (Branlant and Branlant 1985); *Anabaena variabilis,* sequences 10, 16, 17 (Martin et al. 1993b); *Rhodobacter sphaeroides,* sequence 18 (Gibson and Tabita 1988). For all other sources see refs. given in Michels et al. (1991).

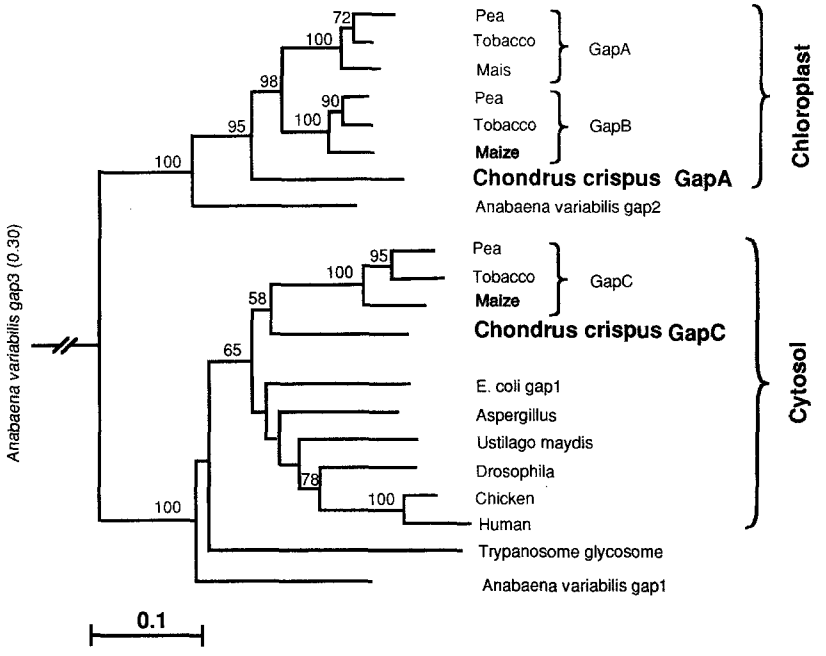

Fig. 2. Phylogenetic GAPDH tree constructed by the neighbor-joining method (Saitou and Nei 1987) from a matrix of d_N values for nucleotide divergence at nonsynonymous sites (Nei and Gojobori 1986). The distance matrix is based on an average of 990 homologous nucleotide positions with an average of 760 nonsynonymous sites. (See Materials and Methods.) The *scale bar* indicates a length of $d_N = 0.10$; branch lengths (horizontal) are drawn to scale. The *gap3* sequence of *Anabaena variabilis* (branch length = 0.30) was used as the outgroup. Sources of sequences are: *Chondrus crispus GapC* and *GapA*, this paper; pea *GapA*, *GapB*, and *GapC* (Brinkmann et al. 1989; Martin et al. 1993a); tobacco *GapA*, *GapB*, and *GapC* (Shih et al. 1986; Brinkmann et al. 1989); maize *GapA* and *GapC* (Brinkmann et al. 1987); maize *GapB* (Liaud et al. unpublished); *Ustilago maydis* (Smith and Leong 1990); *E. coli gap1* (Branlant and Branlant 1985); *Rhodobacter sphaeroides* (Gibson and Tabita 1988); *Anabaena variabilis gap1*, *gap2*, and *gap3* (Martin et al. 1993b). For all other sources see refs. given in Michels et al. (1991).

-Thr-Val(Glu)-AlalLys- in angiosperm *GapA/GapB* (see Fig. 3) and to cleavage sites of chloroplast transit peptides in general (Gavel and von Heijne 1990).

Discussion

Molecular Evidence Suggesting that Red Algae and Green Plants form a Monophyletic Group with Respect to Plastids and Mitochondria

The red alga *Chonus crispus* resembles higher plants in that it contains nuclear genes for two different phosphorylating GAPDH enzymes functioning in glycolysis (*GapC*) and photosynthesis (*GapA*), respectively. The two types of genes show an average of 45% amino acid sequence similarity and specify two separate major branches of the universal GAPDH tree (subtrees *GapA* and *GapC* in Fig. 2). The branching order and bootstrap values of the *GapA* subtree demonstrate that *Chondrus crispus GapA* is more closely related to *GapA* or *GapB* of higher plants than any of these genes is to *gap2* of *Anabaena variabilis*, suggesting that rhodophyte *GapA*, as chlorophyte *GapA/B* (Martin et al. 1993b), is of cyanobacterial descent. Under the reasonable assumption that the transfer of *GapA* to the nucleus occurred in the context of intracellular symbiosis, this implies that rhodoplasts are closely related to chloroplasts and cyanobacteria, respectively. If they were related to purple bacteria, as suggested by several recent contributions based on *rbc*L phylogenies (see below), we would expect *Chondrus GapA* to root more deeply than *Anabaena gap2*. These arguments do not exclude the possibility that chloroplasts and rhodoplasts derived from two different cyanobacterial ancestors.

The *GapC* subtree is particularly interesting, since it carries two eubacterial sequences, *E. coli gap1* and *Anabaena gap1*, highly homologous to eukaryotic *GapC* genes (over 60% amino acid similarity, see Fig. 1; 100% bootstrap significance, see Fig. 2). In striking analogy to the cyanobacterial origin of photosynthetic *GapA*, this strongly suggests that eukaryotic *GapC* is also of eubacterial descent (Martin et al. 1993b). The earliest branch of the *GapC* subtree in Fig. 2 carries *Anabaena gap1* followed by glycosomal *GapC* of trypanosomes (Michels et al. 1991). Gene *gap1* of *E. coli* (a γ-purple bacterium *sensu* Woese; Woese 1987) and the *GapC* genes of animals, fungi, green plants, and red algae represent, approximately, a star phylogeny which is not resolved by our bootstrap analysis. This is compatible with the idea that all eukaryotic *GapC* genes, except for glycosomal *GapC* of *Trypanosoma*, diverged from a common *gap1* gene laterally transferred to the nucleus from ancient purple bacteria, probably the precursors of present-day mitochondria (Smith 1989; Martin et al. 1993b). Whether glycosomal *GapC* of trypanosomes was derived from a different transfer event as *GapC* of other eukaryotes cannot presently be determined. Gly-

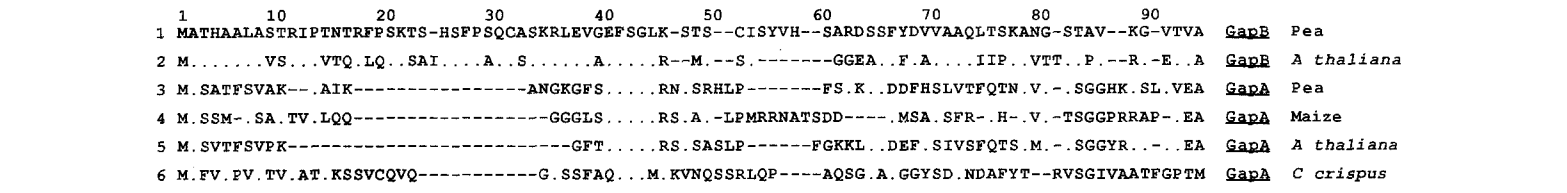

Fig. 3. Alignment of *GapA* and *GapB* transit peptides from higher plants and *Chondrus crispus*. All transit sequences are compared to that of pea *GapB*. Identical amino acids are marked by *dots* and indels by *dashes*. The proteolytic cleavage site of the *Chondrus crispus* transit peptide is probably Met/Lys rather than Ala/Lys as in higher plants. (See text and Gavel and von Heijne 1990.) Sources of sequences: Pea, sequences 1 and 3 (Brinkmann et al. 1989); maize, sequence 4 (Brinkmann et al. 1987); *Arabidopsis thaliana*, sequences 2 and 5 (Shih et al. 1992); *Chondrus crispus*, this paper.

cosomal GAPC possesses several indels found in no other published GAPDH sequences (Michels et al. 1991) which may affect its behavior in phylogenetic inference.

As detailed in the introduction and demonstrated by the GAPDH tree in Fig. 2, the phylogeny of algae must be considered in the context of eubacterial evolution and endosymbiotic gene transfer. Figure 4 illustrates our current view of the evolutionary scenario leading to the GAPDH gene distribution found in present-day eubacteria, rhodophytes, green plants, animals, and fungi. The key to the understanding of this scenario is our recent finding (Martin et al. 1993b) that eubacteria contain ancient GAPDH gene families, which arose by gene duplication in a common eubacterial ancestor before the separation of distinct eubacterial lineages. At present up to six separate types of eubacterial GAPDH genes can be distinguished, which show between 50 and 60% amino acid sequence divergence in pairwise comparisons. As depicted in Fig. 4, cyanobacteria contain GAPDH types 1, 2, and 3 and share type 1 with *E. coli* (Figs. 1 and 2, Martin et al. 1993b). Three additional GAPDH types are found in purple bacteria which may be termed X, Y, and Z, as in Fig. 4. Amino acid sequence alignments (not shown except for GAPDH type X of *Rhodobacter sphaeroides*, lane 18 in Fig. 1) suggest that these types of sequences may be as deeply rooted as GAPDH types 1, 2, and 3. Pairwise comparisons between orthologous members (e.g., type 1 in *E. coli* and cyanobacteria, Figs. 1 and 2; type X in α and β purple bacteria, not shown) show less than 40% amino acid sequence divergence and the bootstrap values of the corresponding branches are virtually 100% significant. (See above and Figs. 1 and 2.) Systematic screens of eubacterial genomes will possibly reveal further orthologous and paralogous members of this ancient gene family. However, it already seems clear that GAPDH genes of eubacteria and eukaryotes are deeply rooted in the *eubacterial* kingdom and that *GapA* and *GapC* became fixed in eukaryotic cells, probably via independent endosymbioses and lateral gene transfer leading to mitochondria and chloroplasts as depicted in Fig. 4. Since rhodophytes and chlorophytes seem to form a monophyletic group for the marker genes *GapA* and *GapC*, this scenario suggests that they have the same origin with respect to plastids and mitochondria in general. Figure 4 assumes that genes were transferred to the nucleus shortly after the fixation of endosymbionts. However, it may well be that fixation of the organellar ancestors and transfer of genes were separated by long evolutionary intervals, so that pastid-nuclear transfer of genes *GapC*, *GapA*, and *GapB* occurred several times independently in independent eukaryotic lineages (Liaud et al. 1990).

Our finding of a relatively close relationship between rhodoplasts and chloroplasts for the *GapA* gene agrees with a recent analysis based on 16S-rRNA sequences (Douglas and Turner 1991) but is at complete variance with tree topologies inferred from *rbc*L/S gene sequences (Douglas et al. 1990; Valentin and Zetsche 1990; Morden and Golden 1991). Trees constructed from *rbc*L genes show a deep dichotomy of two major branches, one of which carries green plastids, cyanobacteria, and γ-purple bacteria; the other branch carries red plastids and α- and β-purple bacteria. While some authors believe (Valentin and Zetsche 1990; Morden and Golden 1991) that this unusual *rbc*L topology reflects polyphylesis of red and green plastids; others suggest (Assali et al. 1990, 1991) that the ancestral endosymbionts may have possessed chimeric genomes due to early plasmid-mediated gene exchanges between purple bacteria and cyanobacteria. In a recent review Martin et al. (1992) suggest another alternative scenario in which the ancestors of plastids, cyanobacteria, and purple bacteria possessed two *rbc*L genes, one of which was secondarily lost in each of the lineages, leading to red plastids and α/β-purple bacteria on the one hand and to green plastids, cyanobacteria, and γ-purple bacteria on the other. This hypothesis proposing early duplications and subsequent differential loss of *rbc*L genes precisely reflects the situation found for GAPDH genes in eubacteria (see above) and also best explains the deep split between *rbc*L genes of γ-purple/cyanobacteria and α/β-purple bacteria, respectively. (For details see Martin et al. 1992; see also Ueda and Shibuya 1992.) In other words, we suggest that the two major *rbc*L branches represent independent trees of two ancient duplicated genes, respectively, as is the case for branches *GapC*

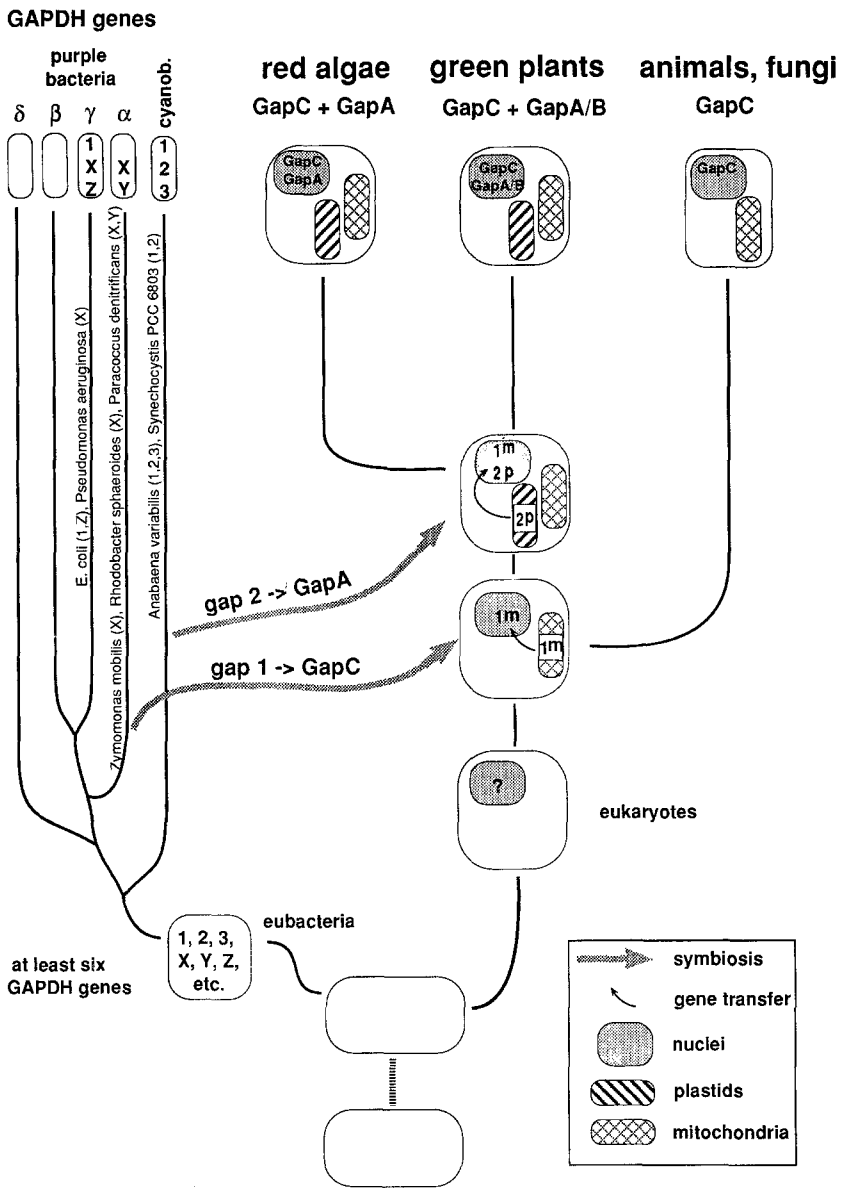

Fig. 4. Evolutionary scenario explaining the distribution of GAPDH genes in present-day eubacteria and eukaryotes. The schematic eubacterial tree on the left showing the branching order of purple bacteria (α, β, γ) and cyanobacteria is based on the 16S rRNA phylogeny of Woese (Woese 1987). There are up to six different types of GAPDH genes in present-day eubacteria showing between 50 and 60% amino acid sequence divergence in pairwise comparisons. Cyanobacteria contain types 1, 2, and 3 and share type 1 with *E. coli* (Figs. 1 and 2). Types X and Y occur in α-purple bacteria and types 1, X, and Z in γ-purple bacteria. (See text.) Eukaryotic genes *GapC* and *GapA* are closely related to genes *gap1* and *gap2* of purple bacteria and cyanobacteria, respectively (see Figs. 1 and 2), suggesting that they were acquired by lateral transfer from the ancestors of present-day mitochondria (*gap1* → *GapC*) and chloroplasts (*gap2* → *GapA*). 1^m and 2^p symbolize genes *gap1* (=*GapC*) and *gap2* (=*GapA/B*) of mitochondrial and plastid origin, respectively. Sources of eubacterial sequences: *Anabaena variabilis gap1, gap2,* and *gap3* (Martin et al. 1993b); *Synechocystis* PCC6803 *gap1* and *gap2* (Brinkmann et al. unpublished); *Zymomonas mobilis* (GAPDH type X; Conway et al. 1987); *Rhodobacter sphaeroides* (GAPDH type X; Gibson and Tabita 1988); *Paracoccus denitrificans* (GAPDH types X and Y, Brinkmann et al. unpublished); *Pseudomonas aeruginosa* (GAPDH type X; Temple et al. unpublished, EMBL data base ID: PAHEXC); *E. coli* (GAPDH types 1 and Z: Branlant and Branlant 1985, and Alefounder and Perham 1989, respectively).

and *GapA* in composite GAPDH topologies (Fig. 2). The major difference between the two gene systems seems to be that in the case of *rbc*L no eubacterial species have (yet) been found which harbor both paralogous gene copies within the same cell.

The Origin of the GapA Transit Peptide Probably Antedates the Rhodophyte/Chlorophyte Separation

Transit peptides of genes *GapA* and *GapB* from higher plants have been shown (Brinkmann et al. 1989; Liaud et al. 1990) to be similar in amino acid sequence both between *GapA* and *GapB* and for a given gene across the monocot/dicot separation. In addition, all higher-plant *GapA* and *GapB* transit peptides have the same cleavage-site AlalLys (Brinkmann et al. 1989), and the corresponding gene sequences are interrupted by two introns in conserved positions (Liaud et al. 1990; Shih et al. 1992). These findings clearly indicate that the fusion of the functional GAPDH subunit with its transit peptide occurred before the *GapA/GapB* separation more than 500 MY ago and long before the angiosperm radiation (Fig. 2 and Liaud et al. 1990). Given the significant conservation of the *Chondrus crispus* transit peptide with respect to primary structure (Fig. 3), the evolutionary history of the *GapA* transit peptide may now be traced back even further, up to the rhodophyte/chlorophyte separation, which may have occurred as early as one billion years ago. This is the oldest chloroplast transit peptide so far reported and there is some indirect evidence that the transit peptide of *GapA,* as the mature subunit to which it is fused, may be of cyanobacterial origin (Liaud et al. 1990).

Note Added in Proof

The nucleotide sequence data reported will appear in the EMBL, GenBank, and DDBJ Nucleotide Sequence Databases under the accession numbers X73035 (*GapA* cDNA of *Chondrus crispus,* 1655 bp) and X73034 (*GapC* cDNA of *Chondrus crispus,* 1165 bp).

Acknowledgments This work was financially supported by Deutsche Forschungsgemeinschaft (Ce 1/15-1) and Deutscher Akademischer Austauschdienst (PROCOPE, 311-pro-91-ca). We thank Ulrike Brandt (Braunschweig) for excellent technical assistance and the Gesellschaft für Biotechnologische Forschung Braunschweig for the use of their computer facilities. The authors are grateful to Chantal Passaquet (Ecole Normale Supérieure, Paris) and to C. Boyen (Roscoff) for their help in preparing algal RNA.

References

Alefounder PR, Perham RN (1989) Identification, molecular cloning and sequence analysis of a gene cluster encoding the class II fructose 1,6-bisphosphate aldolase, 3-phosphoglycerate kinase and a putative second glyceraldehyde 3-phosphate dehydrogenase of *Escherichia coli.* Mol Microbiol 3:723–732

Assali N-E, Mache R, Loiseaux-de Goër S (1990) Evidence for a composite phylogenetic origin of the plastid genome of the brown alga *Pylaiella littoralis* (L.) Kjellm. Plant Mol Biol 15:307–315

Assali N-E, Martin WF, Sommerville CC, Loiseaux-De Goer S (1991) Evolution of the Rubisco operon from prokaryotes to algae: structure and analysis of the *rbc*S gene of the brown algae *Pylaiella littoralis.* Plant Mol Biol 17:853–863

Bhattacharya D, Elwood HJ, Goff LJ, Sogin ML (1990) Phylogeny of *Gracilaria lemaneiformis* (rhodophyta) based on sequence analysis of its small subunit ribosomal RNA coding region. J Phycol 26:181–186

Branlant G, Branlant C (1985) Nucleotide sequence of the *Escherichia coli* gap gene. Different evolutionary behavior of the NAD^+-binding domain and the catalytic domain of D-glyceraldehyde-3-phosphate dehydrogenase. Eur J Biochem 150:61–66

Brinkmann H, Cerff R, Salomon M, Soll (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

Brinkmann H, Martinez P, Martin W, Quigley F, 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

Cavalier-Smith T (1991) The evolution of cells. In: Osawa S, Honjo T (eds) Evolution of life. Fossils, molecules, and culture, pp 271–304. Springer-Verlag, Tokyo

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

Conway T, Sewell GW, Ingram LO (1987) Glyceraldehyde-3-phosphate dehydrogenase gene from *Zymomonas mobilis:* cloning, sequencing, and identification of promotor region. J Bacteriol 169:5653–5662

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Doolittle RF, Feng DF, Anderson KL, Alberro MR (1990) A naturally occurring horizontal gene transfer from a eukaryote to a prokaryote. J Mol Evol 31:383–388

Doolittle WF (1988) Bacterial evolution. Can J Microbiol 34:547–551

Douglas SE, Durnford DG, Morden CW (1990) Nucleotide sequence of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from *Cryptomonas* F: evidence supporting the polyphyletic origin of plastids. J Phycol 26:508–522

Douglas SE, Turner S (1991) Molecular evidence for the origin of plastids from a cyanobacterium-like ancestor. J Mol Evol 33: 267–273

Feinberg AP, Vogelstein B (1984) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137:266

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

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

Gavel Y, von Heijne G (1990) A conserved cleavage-site motif in chloroplast transit peptides. FEBS Lett 261:455–458

Gibson JL, Tabita FR (1988) J Bacteriol 170:2153–2158

Gray MW (1989) The evolutionary origins of organelles. Trends Genet 5:294–299

Hendriks L, De Baere R, Van de Peer Y, Neefs J, Goris A, De Wachter R (1991) The evolutionary position of the rhodophyte *Porphyra umbilicalis* and the basidiomycete *Leucosporidium scotii* among other eukaryotes as deduced from complete sequences of small ribosomal subunit RNA. J Mol Evol 32:167–177

Hohn B (1979) In vitro packaging of I and cosmid DNA. Methods Enzymol 68:299

Hori H, Osawa S (1987) Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences. Mol Biol Evol 4:445–472

Le Gall Y, Braud JP, Kloareg B (1990) Protoplast production in *Chondrus crispus* gametophytes (Gigartinales, Rhodophyta). Plant Cell Rep 8:582–585

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

Lipscomb DL (1989) Relationships among the eukaryotes. In: Fernholm B, Bremer K, Jörnvall H (eds) The hierarchy of life. Elsevier Science Publishers, Amsterdam, Netherlands, pp 161–178

Margulis L, Schwartz KV (1982) Five kingdoms: an illustrated guide to the phyla of life on earth, 338. WH Freeman, New York

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

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, Gierl A, Saedler H (1989) Molecular evidence for pre-Cretaceous angiosperm origins. Nature 339:46–48

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

Martin W, Sommerville CC, Loiseaux-de Goër (1992) Molecular phylogenies of plastid origins and algal evolution. J Mol Evol 35:385–403

Michels PAM, 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

Morden CW, Golden SS (1991) Sequence analysis and phylogenetic reconstruction of the genes encoding the large and small subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase from the chlorophyll b-containing prokaryote *Prochlorothrix hollandica.* J Mol Evol 32:379–395

Nei M (1991) Relative efficiencies of different tree making methods for molecular data. In: Miymoto MM, Cracraft JL (eds) Recent ad-

vances in phylogenetic studies of DNA sequences. Oxford University Press, Oxford, pp 90–128

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous substitutions. Mol Biol Evol 3:418–426

Perasso R, Baroin A, Qu LH, Bachellerie JP, Adoutte A (1989) Origin of the algae. Nature 339:142–144

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

Scagel RF, Bandoni RJ, Maze JR, Rouse GE, Schofield WB, Stein JR (1982) Nonvascular plants: an evolutionary survey. Wadsworth, Inc, Belmont, CA, p 570

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 dehydrogenases. Cell 47:73–80

Shih MC, Lazar G, 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

Smith TL (1989) Disparate evolution of yeasts and filamentous fungi indicated by phylogenetic analysis of glyceraldehyde-3-phosphate dehydrogenase genes. Proc Natl Acad Sci USA 86:7063–7066

Smith TL, Leong SA (1990) Isolation and characterization of a *Ustilago maydis* glyceraldehyde-3-phosphate dehydrogenase-encoding gene. Gene 93:111–117

Ueda K, Shibuya H (1993) Molecular phylogeny of *rbc*L and its bearing on the origin of plastids: bacteria or cyanophyta. In: Endocytobiology V. Tübingen University Press, pp. 369–376

Valentin K, Zetsche K (1990) Structure of the Rubisco operon from the unicellular red alga *Cyanidium caldarium:* evidence for a polyphyletic origin of the plastids. Mol Gen Genet 222:425–430

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

Young RA, Davis RW (1983) Efficient isolation of genes by using antibody probes. Proc Natl Acad Sci USA 80:1194